Crystal Structure of the Catalytic Domain of Drosophila β1,4-Galactosyltransferase-7*

Boopathy Ramakrishnan† and Pradman K. Qasba‡

From the †Structural Glycobiology Section and ‡Basic Research Program, SAIC-Frederick, Inc., Center for Cancer Research Nanobiology Program, Center for Cancer Research, NCI, National Institutes of Health, Frederick, Maryland 21702

The β1,4-galactosyltransferase-7 (β4Gal-T7) enzyme, one of seven members of the β4Gal-T family, transfers in the presence of manganese Gal from UDP-Gal to an acceptor sugar (xylose) that is attached to a side chain hydroxyl group of Ser/Thr residues of proteoglycan proteins. It exhibits the least protein sequence similarity with the other family members, including the well studied family member β4Gal-T1, which, in the presence of manganese, transfers Gal from UDP-Gal to GlcNAc. We report here the crystal structure of the catalytic domain of β4Gal-T7 from Drosophila in the presence of manganese and UDP at 1.81 Å resolution. In the crystal structure, a new manganese ion-binding motif (HXH) has been observed. Superposition of the crystal structures of β4Gal-T7 and β4Gal-T1 shows that the catalytic pocket and the substrate-binding sites in these proteins are similar. Compared with GlcNAc, xylose has a hydroxyl group (instead of an N-acetyl group) at C2 and lacks the CH₂OH group at C5; thus, these protein structures show significant differences in their acceptor-binding site. Modeling of xylose in the acceptor-binding site of the β4Gal-T7 crystal structure shows that the aromatic side chain of Tyr127 interacts strongly with the C5 atom of xylose, causing steric hindrance to any additional group at C5. Because Drosophila Cd7 has a 73% protein sequence similarity to human Cd7, the present crystal structure offers a structure-based explanation for the mutations in human Cd7 that have been linked to Ehlers-Danlos syndrome.

Proteoglycans such as heparin/heparan sulfate, chondroitin, and dermatan sulfates are found widely on the cell surface and in the extracellular matrix of various tissues and are known to play important roles in several cellular functions such as cell growth and differentiation (1, 2). These glycosaminoglycans are biosynthesized as extended oligosaccharides on a linker tetrasaccharide (GlcUAβ1,3Glcβ1,3Glcβ1,4Xylβ-O-2) that is attached to a Ser residue on a core protein (3). Of the glycosyltransferases that are involved in the linker saccharide synthesis, the crystal structure of β1,3-glucurononyltransferase I, which transfers βGlcUA to Galβ1,3Galβ1,4Xylβ-O, and α1,4-N-acetylhexasaminyltransferase (EXTL2), which transfers either αGlcNAc or αGalNAc to the terminal βGlcUA residue of the linker tetrasaccharide, are available (4, 5). The Galβ1,4Xyl disaccharide moiety in the linker saccharide is synthesized by the enzyme β1,4-galactosyltransferase-7 (β4Gal-T7), which transfers Gal to Xyl in the presence of manganese (6–9). In humans, β4Gal-T7 is one of seven members of the β4Gal-T family, β4Gal-T1 to β4Gal-T7. Its homolog is present in all vertebrates and invertebrate species (10). It has been shown that β4Gal-T7 is an essential enzyme for the species viability, and mutation in the β4Gal-T7 gene has been linked to Ehlers-Danlos syndrome (9, 11).

The β4Gal-T7 protein shows a 36% sequence similarity to its family member β4Gal-T1, whose structure and function are well known (12, 13). Briefly, in the presence of manganese, the β4Gal-T1 enzyme transfers Gal from UDP-Gal to GlcNAc present at the nonreducing end of an oligosaccharide acceptor (14). Also, in mammals, during lactation in the mammary gland, its acceptor specificity is altered from GlcNAc to Glc by the mammary gland-specific protein α-lactalbumin, thus synthesizing lactose present in milk (15, 16). The structure and function studies on β4Gal-T1 have shown that the apoenzyme exists in an open conformation with its catalytic pocket exposed to the solvent environment to facilitate the binding of manganese and UDP-Gal (13). Upon manganese and UDP-Gal binding, it undergoes conformational changes involving two flexible loops: 1) a short flexible loop where the side chain of a Trp residue moves from outside to inside the catalytic pocket binding to UDP-Gal, and 2) a long flexible loop that moves to cover the bound UDP-Gal by forming a manganese ion coordination bond with its N-terminal His residue while its C-terminal residues undergo loop-to-helix transition, creating the binding site for the acceptor sugar substrate and α-lactalbumin (12, 13, 17). Upon binding of the acceptor substrate GlcNAc, Asp318 in the bovine β4Gal-T1 enzyme (or Asp314 in human) acts as the catalytic base, enabling the O4 atom of the GlcNAc to initiate a nucleophilic attack on the C1 atom of the galactose moiety of the UDP-Gal following the S₂ catalytic mechanism (18, 19). After catalysis, the product disaccharide leaves the enzyme,

---

*This work was supported, in whole or in part, by National Institutes of Health Contract HHSN261200800001E from NCI and by the Intramural Research Program of the National Institutes of Health, NCI, Center for Cancer Research. Use of beam line 22-ID/BM of the Southeast Regional Collaborative Access Team (SER-CAT), located at the Advanced Photon Source, Argonne National Laboratory, was supported by Contract W-31-109-Eng-38 from the United States Department of Energy, Office of Science, Office of Basic Energy Sciences.

†The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1–S5.

The atomic coordinates and structure factors (code 3LW6) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

†To whom correspondence should be addressed: Structural Glycobiology Section, CCRNP, CCR, NCI-Frederick, Bldg. 469, Rm. 221, Frederick, MD 21702. Tel.: 301-846-1934; Fax: 301-846-7149; E-mail: qasba@helix.nih.gov.

‡The abbreviations used are: GlcUA, glucuronic acid; Xyl, xylose; β4Gal-T, β1,4-galactosyltransferase; MPD, 2-methyl-2,4-pentanediol.
allowing it to revert back to the open conformation to exchange the bound UDP for UDP-Gal to start a new catalytic cycle (12, 13).

Although the β4Gal-T7 molecule shares protein sequence similarity with the β4Gal-T1 molecule and is expected to have a similar overall three-dimensional structure, it exhibits distinct differences from β4Gal-T1, including its acceptor sugar specificity. For example, it has been shown that the binding of manganese and UDP alone can introduce conformational changes in the β4Gal-T7 molecule, whereas manganese and UDP-Gal are necessary in the β4Gal-T1 molecule (20–23). Also, the β4Gal-T1 enzyme exhibits high catalytic activity above pH 7.0, whereas the β4Gal-T7 enzyme exhibits maximum catalytic activity at pH 6.5 (7). Therefore, the three-dimensional structure of the β4Gal-T7 molecule is essential to understand its structure and function. We present here the crystal structure of the catalytic domain of β4Gal-T7 from Drosophila in the presence of manganese and UDP at 1.81 Å resolution.

**EXPERIMENTAL PROCEDURES**

The Drosophila melanogaster β4Gal-T7 cDNA (AY094665) was purchased from Open Biosystems. It was found to have a single amino acid mutation, Asp212 to Gly, which was corrected while cloning. The DNA fragments coding the catalytic domain (Cd7) sequence (residues 71–322) and the C-terminal 11-amino acid deletion (Cd7ΔC) sequence (residues 71–311) were cloned into a pET23a vector between the restriction enzyme sites BamHI and EcoRI (Fig. 1A). The N-terminal fusion proteins P-Cd7ΔC and P1-Cd7ΔC, containing a 46- and a 33-amino acid peptide (P and P1), respectively, from bovine β4Gal-T1 (residues 130–175 and 143–175, respectively), were constructed by inserting the DNA fragments corresponding to the fusion peptides P and P1 at the N terminus of the Cd7ΔC protein between the restriction enzyme sites BamHI and NcoI (see Fig. 1A). The fusion constructs P-Cd7ΔC and P1-Cd7ΔC have the DNA sequence corresponding to the C-terminal deleted catalytic domain of Drosophila β4Gal-T7 (Cd7ΔC) inserted between the restriction enzyme sites NcoI and EcoRI. The four plasmids containing cDNA sequences of Cd7, Cd7ΔC, P-Cd7ΔC, and P1-Cd7ΔC were first confirmed by DNA sequencing and then transfected into a Rosetta(DE3)pLysS cell for protein expression.

**Protein Expression and Refolding**—The four Rosetta(DE3)-pLysS cells, each containing one of the four forms of Drosophila Cd7 cDNA sequences, Cd7, Cd7ΔC, P-Cd7ΔC, and P1-Cd7ΔC, in the pET23a vector described above, were grown to an absorbance of 0.7–0.8 and then induced with 1 mM isopropyl β-D-thiogalactopyranoside. The inclusion bodies were purified from the bacterial pellet as described (24, 25). From 1 liter of bacterial culture, 60–70 mg of protein were obtained as inclusion bodies. The *in vitro* folding of Cd7 was carried out in a way similar to that of β4Gal-T1 (25). Typically, 100 mg of sulfonated protein were folded for 48 h in 1 liter of folding solution containing oxidifying agents and 500 mM arginine HCl. After refolding the protein, the folding solution was extensively dialyzed in water. During dialysis, the misfolded protein precipitated out, whereas the folded protein remained soluble. The soluble active protein was concentrated on an Amicon stirred cell using a YM-10 membrane, and no further purification was necessary. Nearly 20 mg of folded and active *Drosophila* Cd7 protein were obtained from 1 liter of folding solution. The C-terminal deletion (Cd7ΔC) or the addition of N-terminal fusion peptide P (P-Cd7ΔC) or P1 (P1-Cd7ΔC) did not affect the folding efficiency.

**Crystallization and Structure Determination**—The crystals of P1-Cd7ΔC were grown by the hanging drop vapor diffusion method by mixing equal amounts of P1-Cd7ΔC protein solution containing 10–20 mg/ml protein, 33 mM UDP, and 66 mM MnCl2 with a solution of precipitating agent containing 100 mM Tris·HCl (pH 8.0), 1 M NaCl, 15% (v/v) MPD, and 5% (w/v) polyethylene glycol 6000. The tetragonal bipyramidal crystals grew in 1–2 days to a size of 0.2–0.4 mm³. A 1.81 Å resolution native data set was collected on the 22-BM beam line with an x-ray beam at a 1.0-Å wavelength. The data collection statistics are given in Table 1. Crystals grown in the presence of NaBr instead of NaCl were used for solving the crystal structure using anomalous dispersion arising from bromine atoms (26). Single-wavelength anomalous dispersion data up to 2 Å resolution were collected on the crystals grown with 1 M NaBr at a 0.919-Å wavelength on the 22-ID beam line. All frames were processed with HKL3000 (27). The protein structure was solved by single-wavelength anomalous dispersion methods using the program HKL2MAP (28). In the crystal structure, there is one P1-Cd7ΔC molecule in the asymmetric unit. Of the total 240 residues in the Cd7ΔC protein, 230 residues were automatically fitted into the solvent-flattened electron density maps generated by HKL2MAP by the web-based ARP/wARP program (29). At this stage, the model was refined using the native data up to 1.81 Å resolution. The missing 10 residues were better visible and were built based on the difference electron density maps using the program Coot (30). In addition to one manganese ion, UDP, and one MPD molecule, the solvent water molecules located using the program Coot were also included in the refinement. All refinements were carried out using Refmac5.5, which is part of the CCP4i package (31, 32). The final refinement statistics are given in Table 1. The final protein model contains all 240 residues from the Cd7ΔC protein. The N-terminal fusion peptide P1 could not be located in the electron density maps. Although the backbone conformation for Leu263 is in a generously allowed region of the Ramachandran plot, it has a well defined electron density. All figures were drawn using the PyMOL molecular graphics program. The structure factors and coordinates have been deposited in the Protein Data Bank (code 3LW6).

**RESULTS AND DISCUSSION**

**Catalytic Domain of Drosophila β4Gal-T7 and Its Crystallization**—A protein sequence comparison of the human and *Drosophila* β4Gal-T7 enzymes shows that a strong similarity exists, beginning with His93 of the human sequence and His76 of the *Drosophila* sequence, suggesting that the catalytic domain may start from this residue (Fig. 1B). This was further confirmed by comparing more known β4Gal-T7 protein sequences from other species (supplemental Fig. S1). Thus, the *Drosophila* Cd7 expression construct contained Gly71–Thr322. The *in vitro* refolded active Cd7 protein exhibits characteristic catalytic
activity shown for Drosophila β4Gal-T7 (supplemental Fig. S2) (7). Crystallization of this refolded Drosophila Cd7 protein has not been successful. Because only the β4Gal-T7 enzymes from flying insects have the C-terminal extended peptide (Fig. 1B and supplemental Fig. S1), it was hypothesized that this extended peptide might interfere with crystal packing. Therefore, the C-terminal 11-amino acid extension in Drosophila Cd7 (residues 312–322) was deleted. The refolding of the C-terminal deletion Drosophila protein Cd7ΔC also generated soluble and active protein without any loss of catalytic activity (supplemental Fig. S2) or folding efficiency. However, so far, crystallization of this protein has also failed. Because β4Gal-T7 is a member of the β4Gal-T family, we compared the protein sequences of β4Gal-T7 from human and Drosophila with the catalytic domain of bovine β4Gal-T1 (Cd1) (Fig. 1B). Such a comparison shows that the Cd7 protein exhibits its sequence similarity to the bovine β4Gal-T1 catalytic domain (Cd1), beginning with residue 176. The Cd7 protein lacks a similar N-terminal sequence corresponding to bovine Cd1 residues 129–175. Because bovine Cd1 crystallizes readily, it was hypothesized that the absence of the similar N-terminal residues in Drosophila Cd7 may be responsible for the difficulties in crystallizing it. Therefore, the N-terminal peptides from bovine Cd1 (residues 129–175 (P) or residues 143–175 (P1)) were fused as an N-terminal fusion peptide with the Cd7ΔC protein (Fig. 1A). The presence of the N-terminal fusion peptide P or P1 affected neither the catalytic activity nor the folding efficiency. The fusion protein P-Cd7ΔC could not be crystallized. In contrast, the P1-Cd7ΔC protein readily crystallized in the presence of MnCl2 and UDP or UDP-Gal. Inclusion of its own stem region (residues 34–70) with the catalytic domain decreased the folding efficiency to ~1% only, and this soluble domain protein has not been crystallized so far.

Overall Crystal Structure of Drosophila P1-Cd7ΔC—In the crystal structure of P1-Cd7ΔC, all 240 residues in the Cd7ΔC molecule have been located, although the N-terminal fusion peptide P1 from bovine Cd1 could not be clearly located from electron density mapping (Fig. 2A). However, SDS-PAGE analysis of the protein crystals showed the presence of the intact fusion protein (P1-Cd7ΔC) in the crystals. Therefore, the fusion peptide in the crystal is considered to be disordered. In the crystal structure of Cd7ΔC, there are two disulfide bonds found at the C terminus between Cys255 and Cys310 and between Cys300 and Cys308 (Fig. 2A). The cysteine residues of the former disulfide bond are present only in the β4Gal-T7 protein from flying insects, whereas the latter are conserved in the β4Gal-T7 protein from all of the species (supplemental Fig. S1). This is in contrast to the crystal structure of the bovine β4Gal-T1 catalytic domain Cd1 (Fig. 2B) (22), in which the two disulfide bonds found are in the N-terminal region and are conserved in all six family members, β4Gal-T1 to β4Gal-T6, from all of the species.

The crystal structure of the Cd7ΔC protein (Fig. 2A) forms the conventional GT-A fold similar to bovine Cd1 (Fig. 2B). Superposition of these two crystal structures (Fig. 2C) using combinatorial extension methods (33) shows a root mean square deviation of 1.8 Å between the Drosophila Cd7 structure (His75–Cys300) and the bovine Cd1 structure (His180–Thr400) (22, 23). A total of 209 residues were used in the alignment; the Z-score is 6.8. In this superposition, there are five regions showing significant differences (labeled 1–5 in Fig. 2C). The N-terminal residues Cys134–Cys176 of bovine β4Gal-T1 and their interactions with its C terminus are absent in the Cd7ΔC crystal structure (labeled 1 and 2, respectively, in Fig. 2C). The C-terminal region of the Cd7ΔC molecule has two disulfide bonds (Fig. 2A) and exhibits a significant difference from the β4Gal-T1 structure (labeled 3 in Fig. 2C). The acceptor substrate-binding site in these structures, particularly the N2-acetyl and C6 H2OH groups of the GlcNAc-binding regions, shows significant differences (labeled 4 and 5, respectively, in Fig. 2C). It has been shown that the catalytic domain of bovine β4Gal-T1 begins with Cys134 and that its catalytic activity is lost either upon mutation of Cys134 to Ser or upon deletion of 9 amino acids at the N terminus, suggesting that these residues in the N-terminal region of bovine Cd1 play an important role in catalytic activity (24). In the crystal structure of bovine β4Gal-T1, Cys134 forms a disulfide bond with Cys176 (Fig. 2B), and part of this N-terminal region interacts with the C-terminal residues (17, 22, 23). The lack of this region in β4Gal-T7 (labeled 1 in Fig. 2C) may have been compensated by the longer C-terminal extension with a disulfide bond (labeled 3 in Fig. 2C) and extended C-terminal sheet structure (labeled 2 in Fig. 2C).

Manganese and UDP Binding to Drosophila Cd7ΔC—In the Drosophila Cd7ΔC crystal structure, one manganese ion and one UDP molecule have been located from electron density mapping (Fig. 3A). The manganese ion exhibits six coordination bonds: one with the side chain carboxylate oxygen atom of Asp147; two with the side chain nitrogen atoms Ne2 and Nδ1 of His241 and His243, respectively; two with the oxygen atoms of each of the two phosphate groups of the UDP molecule; and one
with a water molecule (Fig. 3A). Asp147 is part of the well known metal-binding DxD motif that is found in most glycosyltransferases (34). However, in the present structure, a new metal-binding motif, H241HXH243, has been observed for the first time in the crystal structure of a glycosyltransferase. The HXH motif is conserved in the H92524Gal-T7 enzymes from all species. Superposition of the crystal structure of the Drosophila Cd7 enzyme (22, 23) with the structure of the wild-type bovine Cd1 enzyme shows that the binding of the manganese ion and UDP molecule in these proteins is very similar (22, 23). However, the $K_d$ value for manganese binding to the Drosophila enzyme ($K_d = 1.5 \pm 0.2 \mu M$) (supplemental Fig. S3) is 20 times lower than the $K_d$ value for its binding to the wild-type bovine Cd1 enzyme ($K_d = 30 \mu M$), whereas it is comparable with the $K_d$ value for the M344H-4Gal-T1 mutant enzyme ($K_d = 1.4 \mu M$) (35). The higher $K_d$ value for manganese binding to wild-type bovine Cd1 may be due to the coordination of the side chain S atom of Met344 with the manganese ion (22, 23). When Met344 in bovine Cd1 is mutated to His, as in the M344H-4Gal-T1 enzyme, the N2

FIGURE 1. A, schematic diagram showing the bovine catalytic domain of 4Gal-T1 (Cd1) and various protein constructs of the catalytic domain of the Drosophila 4Gal-T7 molecule (Cd7, Cd7ΔC, P-Cd7ΔC, and P1-Cd7ΔC). B, protein sequence comparison of human and Drosophila 4Gal-T7 proteins with the catalytic domain of the bovine 4Gal-T1 (Cd1) protein. The protein sequence similarity and identity between the human and Drosophila 4Gal-T7 enzymes are 58 and 43%, respectively, and the similarity between bovine Cd1 and the human and Drosophila 4Gal-T7 enzymes is 33 and 43% respectively. The 4Gal-T7 proteins show high similarity, starting with His93 in human and His75 in Drosophila, suggesting that their catalytic domains might start with these residues. The catalytic domains of 4Gal-T7 proteins lack the N-terminal region corresponding residues 130–175 of the bovine Cd1 protein. However, the functional residues (underlined) in 4Gal-T1, such as the metal-binding motif DVD and the catalytic pocket residues WGGEDDD, are conserved in the 4Gal-T7 proteins. The disulfide bond-forming residues in the human and Drosophila 4Gal-T7 proteins are shown by arrows. The three individual amino acid mutations (A186D, L206P, and R270C) that have been linked to Ehlers-Danlos syndrome in human 4Gal-T7 are indicated (#) (11, 42, 43).

Crystal Structure of 4Gal-T7

Bovine Cd1

Drosophila Cd7

Drosophila Cd7ΔC

Drosophila P-Cd7ΔC

Drosophila P1-Cd7ΔC

Bovine 4Gal-T1–CD

Human 4Gal-T7

Drosophila 4Gal-T7

Human 4Gal-T7

Drosophila 4Gal-T7

Human 4Gal-T7

Drosophila 4Gal-T7

Human 4Gal-T7

Drosophila 4Gal-T7

Bovine 4Gal-T1–CD

Human 4Gal-T7

Drosophila 4Gal-T7

Human 4Gal-T7

Drosophila 4Gal-T7

Human 4Gal-T7

Drosophila 4Gal-T7

Human 4Gal-T7

Drosophila 4Gal-T7

with a water molecule (Fig. 3A). Asp147 is part of the well known metal-binding DxD motif that is found in most glycosyltransferases (34). However, in the present structure, a new metal-binding motif, H241HXH243, has been observed for the first time in the crystal structure of a glycosyltransferase. The HXH motif is conserved in the 4Gal-T7 proteins from all species. Superposition of the crystal structure of the Drosophila Cd7ΔC-Mn$^{2+}$–UDP complex with the crystal structure of the wild-type bovine Cd1-Mn$^{2+}$–UDP–Gal complex shows that the binding of the manganese ion and UDP molecule in these proteins is very similar (22, 23). However, the $K_d$ value for manganese binding to the Drosophila Cd7ΔC enzyme ($K_d = 1.5 \pm 0.2 \mu M$) (supplemental Fig. S3) is 20 times lower than the $K_d$ value for its binding to the wild-type bovine Cd1 enzyme ($K_d = 30 \mu M$), whereas it is comparable with the $K_d$ value for the M344H-4Gal-T1 mutant enzyme ($K_d = 1.4 \mu M$) (35). The higher $K_d$ value for manganese binding to wild-type bovine Cd1 may be due to the coordination of the side chain S atom of Met344 with the manganese ion (22, 23). When Met344 in bovine Cd1 is mutated to His, as in the M344H-4Gal-T1 enzyme, the Nε2...
Although we grew crystals of the *Drosophila* P1-Cd7ΔC protein in the presence of UDP-Gal and manganese, in the crystal structure, the galactose moiety was not observed (Fig. 3A), suggesting that the UDP-Gal might have hydrolyzed. The *Drosophila* Cd7 protein naturally exhibits hydrolysis activity with UDP-Gal in the absence of the acceptor substrate, and this may be responsible for the absence of the Gal moiety in the crystals of the P1-Cd7ΔC protein grown with UDP-Gal and manganese. This is similar to the bovine α3Gal-T enzyme, which also exhibits UDP-Gal hydrolysis activity and whose crystal structure with bound UDP-Gal could be determined only with a mutant enzyme that had very low catalytic activity (37). In the present crystal structure (Fig. 3A), the binding of UDP is similar to the binding of UDP-Gal to the bovine Cd1 protein (Fig. 3B). The uridine base stacks on the side chain phenyl group of Phe122, similar to the bovine β4Gal-T1 enzyme, in which the uridine base stacks on the side chain of Phe226 (22). The β-phosphate oxygen atoms form two hydrogen bonds with the Cd7ΔC protein, where the side chain aromatic nitrogen atom of Trp207 and the side chain hydroxyl group of Tyr177 form hydrogen bonds with the β-phosphate oxygen atom (Fig. 3A). Only the former hydrogen bond is observed in the crystal structure of bovine Cd1, where the side chain nitrogen atom of Trp213 forms a hydrogen bond with the oxygen atom of the β-phosphate group (17). The binding of UDP in the crystal structure of Cd7ΔC is very similar to the binding of UDP-Gal to the bovine β4Gal-T1 molecule, where the residues that bind to galactose in bovine β4Gal-T1, such as Asp252, Glu317, and Asp318, are structurally conserved in the present *Drosophila* Cd7ΔC protein. Thus, the binding of Gal to Cd7 is expected to bind Asp145, Glu210, and Asp211, which will be quite similar to Gal binding to β4Gal-T1.

*Drosophila* Cd7ΔC Protein Is in the Closed Conformation—In the present crystal structure, residues 242–251 cover the bound Mn$^{2+}$ and the UDP molecule (Fig. 3A). This suggests that the observed *Drosophila* Cd7ΔC crystal structure has undergone conformational changes upon Mn$^{2+}$ and UDP binding, involving at least 9 amino acids, residues 242–251 or up to Cys255, which forms a disulfide bond with Cys310. This is substantiated by earlier calorimetric studies on human β4Gal-T7 that showed that the protein molecule undergoes conformational change upon Mn$^{2+}$ and UDP binding (20). In the bovine Cd1 crystal

atom of His$^{344}$ coordinates with the manganese ion (35), similar to the situation in the *Drosophila* Cd7ΔC crystal structure (Fig. 3A). Interestingly, the bovine M344H-β4Gal-T1 mutant enzyme exhibits better catalytic activity with magnesium than with manganese (35), whereas the *Drosophila* Cd7 enzyme exhibits only very low catalytic activity with magnesium. In the M344H-β4Gal-T1-Mn$^{2+}$-UDP-Gal complex crystal structure, the Ne2 atoms of both His$^{344}$ and His$^{347}$ form a coordination bond with the manganese ion (35); this is in contrast to the present crystal structure of *Drosophila* Cd7ΔC, in which the Ne2 and Nδ1 atoms of His$^{241}$ and His$^{243}$, respectively, form a coordination bond with the manganese ion. In solution, the tautomers of the His residue (Ne-H and Nδ1-H) exist in equilibrium, and the ratio of these two tautomers depends on the pH of the solution. It has been suggested that, due to the steric effect, the Nδ1-H tautomer is often found to coordinate with the metal ion rather than with the Ne2-H tautomer (36). This may explain why these two factors might be responsible for the observed high catalytic activity at pH 6.5 and lower $K_d$ value for the manganese ion for *Drosophila* Cd7ΔC compared with β4Gal-T1, where a maximum catalytic activity has been observed above the neutral pH values.

The structure of the Cd7 molecule (Fig. 2C) progressively changes from red to yellow as one moves from the N-terminal residues to the C-terminal residues; a similar region in the bovine α3Gal-T enzyme (38) may have been structurally compensated by the difference in the C-terminal region of Cd7ΔC (labeled 4 and 5 in Fig. 2C).

![Crystal Structure of β4Gal-T7](image)

**Crystal Structure of β4Gal-T7**

Although we grew crystals of the *Drosophila* P1-Cd7ΔC protein in the presence of UDP-Gal and manganese, in the crystal structure, the galactose moiety was not observed (Fig. 3A), suggesting that the UDP-Gal might have hydrolyzed. The *Drosophila* Cd7 protein naturally exhibits hydrolysis activity with UDP-Gal in the absence of the acceptor substrate, and this may be responsible for the absence of the Gal moiety in the crystals of the P1-Cd7ΔC protein grown with UDP-Gal and manganese. This is similar to the bovine α3Gal-T enzyme, which also exhibits UDP-Gal hydrolysis activity and whose crystal structure with bound UDP-Gal could be determined only with a mutant enzyme that had very low catalytic activity (37). In the present crystal structure (Fig. 3A), the binding of UDP is similar to the binding of UDP-Gal to the bovine Cd1 protein (Fig. 3B). The uridine base stacks on the side chain phenyl group of Phe122, similar to the bovine β4Gal-T1 enzyme, in which the uridine base stacks on the side chain of Phe226 (22). The β-phosphate oxygen atoms form two hydrogen bonds with the Cd7ΔC protein, where the side chain aromatic nitrogen atom of Trp207 and the side chain hydroxyl group of Tyr177 form hydrogen bonds with the β-phosphate oxygen atom (Fig. 3A). Only the former hydrogen bond is observed in the crystal structure of bovine Cd1, where the side chain nitrogen atom of Trp213 forms a hydrogen bond with the oxygen atom of the β-phosphate group (17). The binding of UDP in the crystal structure of Cd7ΔC is very similar to the binding of UDP-Gal to the bovine β4Gal-T1 molecule, where the residues that bind to galactose in bovine β4Gal-T1, such as Asp252, Glu317, and Asp318, are structurally conserved in the present *Drosophila* Cd7ΔC protein. Thus, the binding of Gal to Cd7 is expected to bind Asp145, Glu210, and Asp211, which will be quite similar to Gal binding to β4Gal-T1.

*Drosophila* Cd7ΔC Protein Is in the Closed Conformation—In the present crystal structure, residues 242–251 cover the bound Mn$^{2+}$ and the UDP molecule (Fig. 3A). This suggests that the observed *Drosophila* Cd7ΔC crystal structure has undergone conformational changes upon Mn$^{2+}$ and UDP binding, involving at least 9 amino acids, residues 242–251 or up to Cys255, which forms a disulfide bond with Cys310. This is substantiated by earlier calorimetric studies on human β4Gal-T7 that showed that the protein molecule undergoes conformational change upon Mn$^{2+}$ and UDP binding (20). In the bovine Cd1 crystal

atom of His$^{344}$ coordinates with the manganese ion (35), similar to the situation in the *Drosophila* Cd7ΔC crystal structure (Fig. 3A). Interestingly, the bovine M344H-β4Gal-T1 mutant enzyme exhibits better catalytic activity with magnesium than with manganese (35), whereas the *Drosophila* Cd7 enzyme exhibits only very low catalytic activity with magnesium. In the M344H-β4Gal-T1-Mn$^{2+}$-UDP-Gal complex crystal structure, the Ne2 atoms of both His$^{344}$ and His$^{347}$ form a coordination bond with the manganese ion (35); this is in contrast to the present crystal structure of *Drosophila* Cd7ΔC, in which the Ne2 and Nδ1 atoms of His$^{241}$ and His$^{243}$, respectively, form a coordination bond with the manganese ion. In solution, the tautomers of the His residue (Ne-H and Nδ1-H) exist in equilibrium, and the ratio of these two tautomers depends on the pH of the solution. It has been suggested that, due to the steric effect, the Nδ1-H tautomer is often found to coordinate with the metal ion rather than with the Ne2-H tautomer (36). This may explain why these two factors might be responsible for the observed high catalytic activity at pH 6.5 and lower $K_d$ value for the manganese ion for *Drosophila* Cd7ΔC compared with β4Gal-T1, where a maximum catalytic activity has been observed above the neutral pH values.

**FIGURE 2.** Shown are the crystal structures of the catalytic domain of the *Drosophila* Cd7ΔC molecule with a bound manganese ion and UDP molecule (A) and the catalytic domain of the bovine β4Gal-T1 molecule in the closed conformation (B), shown as a composite picture with the bound manganese, UDP-Gal, and GlcNAc molecules (generated from individual structures with bound manganese and UDP-Gal and GlcNAc molecules, Protein Data Bank codes 1O0R and 1NQI, respectively). The coloring of the ribbon diagram of the β4Gal-T1 molecule (B) progressively changes from blue for the N-terminal residues to red for the C-terminal residues; a similar region in the *Drosophila* Cd7ΔC molecule (B) is colored likewise. In the ribbon diagram, the disulfide bond-forming Cys residues in both protein molecules are shown (yellow). Superposition of Cα atoms of Cd7ΔC (in red) and bovine β4Gal-T1 (in blue) in stereo is shown in C, as predicted from the sequence comparison, the N-terminal polypeptide region (residues 134–176; shown in blue) in bovine β4Gal-T1 is absent in the crystal structure of the *Drosophila* Cd7ΔC molecule. Superposition of the Cα atoms of these two molecules (C) shows five surface regions with significant differences (labeled 1–5 in Fig. 2C). The differences around the acceptor substrate-binding site (labeled 2 and 3 in Fig. 2C) may be possibly due to the different sugar acceptor specificity of the Cd7ΔC molecule, whereas the lack of the region corresponding to the N-terminal region in bovine β4Gal-T1 (residues 134–176) (labeled 1 in Fig. 2C) may have been structurally compensated by the difference in the C-terminal region of Cd7ΔC (labeled 4 and 5 in Fig. 2C).
structure, a similar region (residues 346–365, with 19 amino acids) undergoes conformational changes upon Mn$^{2+}$ and UDP-Gal binding (17). In addition to this long flexible loop, a short flexible loop (residues 313–315, Gly$^{313}$–Trp$^{314}$–Gly$^{315}$) also undergoes conformational changes, where the Trp$^{314}$ side chain moves from outside to inside the catalytic pocket, forming a hydrogen bond with the β-phosphate oxygen atom of the bound UDP-Gal (17). The same tripeptide sequence is present in Drosophila Cd7ΔC (Gly$^{206}$–Trp$^{207}$–Gly$^{208}$), and the Trp$^{207}$ side chain nitrogen atom forms a hydrogen bond with the β-phosphate oxygen atom of the bound UDP molecule (Fig. 3A). The short loops of both the Drosophila Cd7ΔC and bovine Cd1 molecules exactly superimpose in the superposition of the overall crystal structures of these molecules (Fig. 3B). Therefore, in the present Drosophila Cd7ΔC crystal structure, it is likely that this short loop also has undergone conformational changes from the open to closed state. However, the crystal structures of both the open and closed conformations of the enzyme are required to accurately describe the conformational changes. The conformational changes, from the open to closed state, have been observed in several other glycosyltransferases (38). The important difference between β4Gal-T1 and β4Gal-T7 in their conformational changes is that the binding of manganese and the UDP molecule alone induces conformational change in β4Gal-T7, as observed in the present crystal structure, whereas only manganese and the UDP-Gal molecule, not the UDP molecule, induce conformational change in β4Gal-T1 (21, 22). The ability of β4Gal-T7 to undergo conformational change with the UDP molecule may have been due to the additional hydrogen-bonding interactions between the binding site and the UDP-Gal molecule involving Tyr$^{277}$ and Arg$^{250}$ and the UDP molecule, and they are absent in β4Gal-T1 (Fig. 3A).

Acceptor Xylose Binding—All our attempts to crystallize the Drosophila P1-Cd7ΔC protein with the bound acceptor substrate xylose by co-crystallizing it with manganese, UDP, and xylose molecules or by soaking the crystals that were grown with manganese and UDP in xylose solution failed. It has been shown that the acceptor GlcNAC- or Glc-binding site in β4Gal-T1 is created upon the conformational changes induced by the binding of manganese and UDP-Gal (12, 13, 17). Because the present Drosophila Cd7ΔC crystal structure is in the closed conformation, the acceptor-binding site is expected to be present in the crystal structure. Because the UDP-Gal-binding sites in the Cd7ΔC and β4Gal-T1 crystal structures are similar, the acceptor sugar-binding site in these proteins is also expected to be similar. However, there have to be some binding differences because Cd7ΔC has to accommodate the xylose sugar instead of the GlcNAC or Glc residue, as is the case with β4Gal-T1. Superposition of the crystal structures of the bovine Cd1-Glc complex and Drosophila Cd7ΔC shows that the glucose molecule fits nicely into a cavity in the Cd7ΔC crystal structure, similar to the Glc-binding cavity in the bovine Cd1 crystal structure. The O2, O3, and O4 hydroxyl groups of this modeled glucose molecule form a hydrogen bond with the side chain carboxylate oxygen atoms of Asp$^{212}$ and Asp$^{311}$ of the Cd7ΔC molecule, similar to the interactions with Asp$^{319}$ and Asp$^{318}$ in the bovine Cd1 molecule (17). However, the side chain aromatic group of Tyr$^{277}$ causes steric hindrance with the C6-O6 atoms of the modeled glucose molecule, suggesting that the acceptor-binding site may accommodate only a glucose molecule in the CDH opioid at C5, i.e., a xylose molecule as an acceptor substrate (Fig. 3B). Tyr$^{277}$ is conserved in all of the β4Gal-T7 proteins. Similarly, modeling of the GlcNAC molecule in the Drosophila Cd7ΔC crystal structure indicates that, in addition to the Tyr$^{277}$ steric hindrance with the C6-O6 atoms, the side chain of Leu$^{209}$ causes steric hindrance with the N-acetyl moiety of the GlcNAC sugar. Either Leu or Arg is found at this position in the β4Gal-T7 proteins. Thus, xylose binding to β4Gal-T7 is expected to be similar to Glc/GlcNAC binding to...
Crystal Structure of β4Gal-T7

Therefore, it is possible to offer a structure-based explanation of the effect of these mutations based on the present Cd7ΔC crystal structure. Ala169, Val189, and Lys254 of Drosophila Cd7ΔC correspond to Ala186, Leu206, and Arg270 of human β4Gal-T7, respectively (Fig. 1). In the present crystal structure, Ala169 and Val189 are located in the hydrophobic core of the protein, whereas Lys254 is on the long flexible loop (supplemental Fig. S5). The side chain methyl group of the conserved residue Ala169 is surrounded by bulky aromatic residues; thus, a bulky substitution such as Asp may not be readily accommodated. Furthermore, a polar residue in the hydrophobic environment may cause additional instability to the mutant enzyme. On the other hand, a Pro substitution for Val189 may be possible. However, a substituted Pro residue for Val189 cannot adopt the same backbone conformation because it would be in the disallowed region of the Ramachandran plot for a Pro residue. Thus, a Pro mutation is expected to disrupt the secondary or even the tertiary structure of the protein. Lys254 is located at the C-terminal end of the long flexible loop (residues 242–255), and only a basic amino acid substitution has been observed in the β4Gal-T7 proteins; therefore, a Cys substitution is expected to affect the conformational flexibility of the long flexible loop.

References

1. Bishop, J. R., Schuksz, M., and Esko, J. D. (2007) Nature 446, 1030–1037
2. Conrad, H. E. (1998) Heparin-Binding Proteins, Academic Press, San Diego, CA
3. Gandhi, N. S., and Mancera, R. L. (2008) Chem. Biol. Drug Des. 72, 455–482
4. Pedersen, L. C., Tsuchida, K., Kitagawa, H., Sugahara, K., Darden, T. A., and Negishi, M. (2000) J. Biol. Chem. 275, 34580–34585
5. Pedersen, L. C., Dong, J., Taniguchi, F., Kitagawa, H., Kranh, J. M., Pedersen, L. G., Sugahara, K., and Negishi, M. (2003) J. Biol. Chem. 278, 14420–14428
6. Almeida, R., Levery, S. B., Mandel, U., Kresse, H., Schwientek, T., Bennett, E. P., and Clausen, H. (1999) J. Biol. Chem. 274, 26165–26171
7. Vadaie, N., Hulinsky, R. S., and Jarvis, D. L. (2002) Glycobiology 12, 589–597
8. Nakamura, Y., Haines, N., Chen, J., Okajima, T., Furukawa, K., Urano, T., Stanley, P., Irvine, K. D., and Furukawa, K. (2002) J. Biol. Chem. 277, 46280–46288
9. Takemae, H., Ueda, R., Okubo, R., Nakato, H., Izumi, S., Saigo, K., and Nishihara, S. (2003) J. Biol. Chem. 278, 15571–15578
10. Hennet T., T. (2002) Cell. Mol. Life Sci. 59, 1081–1095
11. Götte, M., Spillmann, D., Yip, G. W., Versteeg, E., Eckermeyer, F. G., van Kuppevelt, T. H., and Kiesel, L. (2008) Hum. Mol. Genet. 17, 996–1009
12. Qasba, P. K., Ramakrishnan, B., and Boegeman, E. (2008) Curr. Drug Targets 9, 292–309
13. Ramakrishnan, B., Boegeman, E., Ramasamy, V., and Qasba, P. K. (2004) Curr. Opin. Struct. Biol. 14, 593–600
14. Brew, K., Vanaman, T. C., and Hill, R. L. (1968) Proc. Natl. Acad. Sci. U.S.A. 59, 491–497
15. Brodbeck, U., Donnet, W. L., Tanahashi, N., and Ebner, K. E. (1967) J. Biol. Chem. 242, 1391–1397
16. Qasba, P. K., and Kumar, S. (1997) Crit. Rev. Biochem. Mol. Biol. 32, 255–306
17. Ramakrishnan, B., and Qasba, P. K. (2001) J. Mol. Biol. 310, 205–218

FIGURE 4. Docked glucose molecule in the acceptor-binding site of the Drosophila Cd7ΔC molecule. Superposition of the Drosophila Cd7ΔC crystal structure with the bovine β4Gal-T1/Glc complex crystal structure places the Glc molecule in the acceptor-binding pocket in the Drosophila Cd7ΔC molecule. In this docked structure, the hydrogen-bonding interactions observed between the O2, O3, and O4 hydroxyl groups of Glc and the Drosophila Cd7ΔC molecule are similar to those found in the crystal structure of the β4Gal-T1/Glc complex molecule. However, the side chain aromatic group of Tyr177 causes severe steric hindrance to the C6-O6 atoms of the docked glucose molecule, indicating that the acceptor substrate-binding pocket in the present Drosophila Cd7ΔC molecule can accommodate only the xylose sugar instead of glucose.

β4Gal-T1, with its O4 hydroxyl group strongly hydrogen-bonded with the side chain carboxylate oxygen atom of Asp211, and during catalysis, Asp211 is expected to act as a catalytic base, similar to Asp318 in bovine β4Gal-T1 (18, 19).

Although the xylose-binding site is present in the Cd7ΔC crystal structure (Fig. 4), crystals with a bound xylose molecule could not be obtained. This could be due to the presence of an MPD molecule in the catalytic pocket. Modeling of a xylose molecule in the acceptor-binding site in the present crystal structure shows that the bound MPD molecule might cause steric hindrance to the binding of the xylose sugar (supplemental Fig. S4). Because diffraction quality crystals have been grown only in the presence of MPD, it is always present in all of the crystal structures. In place of xylose binding, well-organized water molecules are found in the Drosophila Cd7ΔC crystal structure.

Structural Explanation of the Effect of Mutations in Human β4Gal-T7 That Are Linked to Ehlers-Danlos Syndrome—Mutations in glycosyltransferases and glycosidases have been linked to several human diseases, and structure-based explanations of the effect of such mutations on some of these enzymes are known (39–41). In human β4Gal-T7, three individual amino acid mutations, A186D, L206P, and R270C, have been linked to Ehlers-Danlos syndrome (Fig. 1) (11, 42, 43). It has been shown that the L206P mutation abrogates the catalytic activity, whereas the other two mutations reduce the catalytic activity of the enzyme. In the β4Gal-T7 enzymes from different species, Ala186 is conserved, whereas Val or Ile is found for Leu206, and only a basic residue such as Lys or Arg is present for Arg270 (supplemental Fig. S1). Because the Drosophila Cd7ΔC sequence has a 73% protein sequence similarity to human Cd7, their three-dimensional structures are expected to be similar.
Crystal Structure of β4Gal-T7

Ramakrishnan, B., Ramasamy, V., and Qasba, P. K. (2006) J. Mol. Biol. 357, 1619–1633

Krupicka, M., and Tvaroska, I. (2009) J. Phys. Chem. B 113, 11314–11319

Daligault, F., Rahuel-Clermont, S., Gulberti, S., Cung, M. T., Branlant, G., Netter, P., Magdalou, J., and Lattard, V. (2009) Biochem. J. 418, 605–614

Geren, C. R., Magee, S. C., and Ebner, K. E. (1975) Biochemistry 14, 1461–1463

Ramakrishnan, B., Balaji, P. V., and Qasba, P. K. (2002) J. Mol. Biol. 318, 491–502

Ramakrishnan, B., and Qasba, P. K. (2003) J. Biomol. Struct. Dyn. 21, 1–8

Boeggeman, E. E., Balaji, P. V., Sethi, N., Masibay, A. S., and Qasba, P. K. (1993) Protein Eng. 6, 779–785

Boeggeman, E. E., Ramakrishnan, B., and Qasba, P. K. (2003) Protein Expr. Purif. 30, 219–229

Dauter, Z. (2002) Curr. Opin. Struct. Biol. 12, 674–678

Otwinowski, Z., and Minor, W. (1997) Methods Enzymol. 276, 307–326

Pape, T., and Schneider, T. R. (2004) J. Appl. Crystallogr. 37, 843–844

Morris, R. I., Perrakis, A., and Lamzin, V. S. (2002) Acta Crystallogr. Sect. D Biol. Crystallogr. 58, 968–975

Emsley, P., and Cowtan, K. (2004) Acta Crystallogr. Sect. D Biol. Crystallogr. 60, 2126–2132

Potterton, E., Briggs, P., Turkenburg, M., and Dodson E. (2003) Acta Crystallogr. Sect. D Biol. Crystallogr. 59, 1131–1137

Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Acta Crystallogr. Sect. D Biol. Crystallogr. 53, 240–255

Shindyalov, I. N., and Bourne, P. E. (1998) Protein Eng. 11, 739–747

Wiggins, C. A., and Munro, S. (1998) Proc. Natl. Acad. Sci. U.S.A. 95, 7945–7950

Ramakrishnan, B., Boeggeman, E., and Qasba, P. K. (2004) Biochemistry 43, 12513–12522

Chakrabarti, P. (1990) Protein Eng. 4, 57–63

Tumbale, P., Jamaluddin, H., Thiyagarajan, N., Brew, K., and Acharya, K. R. (2008) Biochemistry 47, 8711–8718

Qasba, P. K., Ramakrishnan, B., and Boeggeman, E. (2005) Trends Biochem. Sci. 30, 53–62

Zhao, H., and Grabowski, G. A. (2002) Cell. Mol. Life Sci. 59, 694–707

Garman, S. C., and Garboczi, D. N. (2004) J. Mol. Biol. 337, 319–335

Ju, T., and Cummings, R. D. (2005) Nature 437, 1252

Okajima, T., Fukumoto, S., Furukawa, K., and Urano, T. (1999) J. Biol. Chem. 274, 28841–28844

Seidler, D. G., Faiyaz-Ul-Haque, M., Hansen, U., Yip, G. W., Zaidi, S. H., Teebi, A. S., Kiesel, L., and Götte, M. (2006) J. Mol. Med. 84, 583–594