Dynamic Association between the Catalytic and Lectin Domains of Human UDP-GalNAc:Polypeptide α-N-Acetylgalactosaminytransferase-2*

Timothy A. Fritz, Jayalakshmi Raman, and Lawrence A. Tabak†

From the Section on Biological Chemistry, NIDDK, National Institutes of Health, Bethesda, Maryland 20892

The family of UDP-GalNAc:polypeptide α-N-acetylgalactosaminytransferases (ppGalNAcTs) is unique among glycosyltransferases, containing both catalytic and lectin domains that we have previously shown to be closely associated. Here we describe the x-ray crystal structures of human ppGalNAcT-2 (hT2) bound to the product UDP at 2.75 Å resolution and to UDP and an acceptor peptide substrate EA2 (PTTDSTPAPTTK) at 1.64 Å resolution. The conformations of both UDP and residues Arg362–Ser373 vary greatly between the two structures. In the hT2-UDP-EA2 complex, residues Arg362–Ser373 comprise a loop that forms a lid over UDP, sealing it in the active site, whereas in the hT2-UDP complex this loop is folded back, exposing UDP to bulk solvent. EA2 binds in a shallow groove with threonine 7 positioned consistent with in vitro data showing it to be the preferred site of glycosylation. The relative orientations of the hT2 catalytic and lectin domains differ dramatically from that of murine ppGalNAcT-1 and also vary considerably between the two hT2 complexes. Indeed, in the hT2-UDP-EA2 complex essentially no contact is made between the catalytic and lectin domains except for the peptide bridge between them. Thus, the hT2 structures reveal an unexpected flexibility between the catalytic and lectin domains and suggest a new mechanism used by hT2 to capture glycosylated substrates. Kinetic analysis of hT2 lacking the lectin domain confirmed the importance of this domain in acting on glycopeptide but not peptide substrates. The structure of the hT2-UDP-EA2 complex also resolves long standing questions regarding ppGalNAcT acceptor substrate specificity.

The first committed step of carbohydrate addition to mucin-type glycoproteins is catalyzed by a family of UDP-GalNAc:polypeptide α-N-acetylgalactosaminytransferases (ppGalNAcTs), yielding the Tn antigen (GalNAc-α-1-O-Ser/Thr). This family is large (with ~24 mammalian isoforms) and phylogenetically conserved with Drosophila and an acceptor peptide substrate EA2 (PTTDSTPAPTTK). These details of substrate binding, we have now solved the x-ray crystal structures of human ppGalNAcT-2 (hT2) bound to both UDP and to UDP and an acceptor peptide substrate EA2 (PTTDSTPAPTTK). These structures suggest that the association of ppGalNAcTs catalytic and lectin domains can be dynamic and also reveal the molecular basis of substrate recognition by the ppGalNAcTs.

EXPERIMENTAL PROCEDURES

The annealed primers 5'-AATTCGATGCCTCATCATCATCATCATCATCATCATCATCT-3' and 5'-CGGTCAGTTGAAAATCTTTTACTTCAACAG3' were used to amplify the cDNA from HL-60 cells (ATCC) using a NucleoSpin (BD Biosciences) kit. PCR amplification was done using a Superscript III one-step reverse transcription-PCR kit (Stratagene) and the primers 5'-GCCGTTCA-3' to create the plasmid pKNN5-6His-TEV. Residues 75–571 of hT2 encoding 6 histidine residues followed by a tobacco etch virus (TEV) protease cleavage site were cloned into the EcoRI/MluI sites of the plasmid pKK223-3 to create the plasmid pKK223-3. These structures suggest that the association of ppGalNAcTs catalytic and lectin domains can be dynamic and also reveal the molecular basis of substrate recognition by the ppGalNAcTs.

The first committed step of carbohydrate addition to mucin-type glycoproteins is catalyzed by a family of UDP-GalNAc:polypeptide α-N-acetylgalactosaminytransferases (ppGalNAcTs), yielding the Tn antigen (GalNAc-α-1-O-Ser/Thr). This family is large (with ~24 mammalian isoforms) and phylogenetically conserved with Drosophila and an acceptor peptide substrate EA2 (PTTDSTPAPTTK). These details of substrate binding, we have now solved the x-ray crystal structures of human ppGalNAcT-2 (hT2) bound to both UDP and to UDP and an acceptor peptide substrate EA2 (PTTDSTPAPTTK). These structures suggest that the association of ppGalNAcTs catalytic and lectin domains can be dynamic and also reveal the molecular basis of substrate recognition by the ppGalNAcTs.
cloned between the MluI/AgeI sites of pKN55-N6His-TEV. The catalytic domain of hT2 (residues 75–440) was PCR-amplified using the primers 5′-ACCACCGCTTGAAATCGGTCGCCACTT and 5′-ACCACCGCTCTAGAAAGTCGTTGATCGGAGTACGC and cloned between the MluI/AgeI sites of pKN55-N6His-TEV. The plasmids were linearized and electroporated into Pichia pastoris strain SMD1168 to create stable transformants as previously described (16).

Pichia transformants were grown at 30 °C in rich medium (2% peptone, 1% yeast extract, 1% casamino acids, 1% yeast nitrogen base, 1% glycerol) to an A600 = 2–4. The cells were centrifuged, resuspended in 1/10 the volume of the same medium in which 2% methanol was substituted for glycerol, and induced for 16 h at 20 °C. The cells were removed by centrifugation, and the supernatant was adjusted to 10 mM β-mercaptoethanol (β-ME) and 5 mM EDTA. The supernatant was concentrated and dialyzed against 20 mM NaPO4 (pH 7.5–8) and 0.1–0.2 M NaCl (dialfiltration buffer) using a Millipore tangential flow membrane with a 10-kDa molecular mass cut-off. The sample was concentrated and applied to a 5-ml HiTrap chelate column (GE Bio-Sciences) and eluted using a 5-column volume gradient of 0–500 mM imidazole in dialfiltration buffer. For some purifications, the column was washed with dialfiltration buffer containing 25 mM imidazole and eluted with a linear gradient of 25–500 mM imidazole in dialfiltration buffer.

The product fractions were pooled and incubated with an equimolar amount of TEV protease at 4 °C overnight in 50 mM NaPO4 (pH 8.0), 25 mM imidazole, 0.2 M NaCl, and 10 mg/ml β-ME (cleavage buffer). The sample was centrifuged and passed over a nickel-nitrilotriacetic acid resin (New England Biolabs) in cleavage buffer to remove the six-histidine peptide and TEV protease, and hT2 was diazylized against 2 mM Tris (pH 8.0), 0.5 mM EDTA, and 10 mM β-ME at 4 °C.

Crystals were grown by hanging drop vapor diffusion at room temperature. Termary complex (hT2-UDP-EA2-Mn2⁺) crystal growth was initiated by mixing 0.5–1 μl of protein solution containing 5.8 mg/ml hT2, 2 mM Tris (pH 8.0), 0.5 mM EDTA, 10 mM β-ME, 10 mM UDP, 10 mM MnCl2, and 5 mM EA2 with an equal volume of precipitant solution containing 23–25% polyethylene glycol 1000, 100 mM Heps (pH 7.0). Binary complex (hT2-UDP-Mn2⁺) crystal growth was initiated by mixing 0.5–1 μl of protein solution containing 5.8 mg/ml hT2, 2 mM Tris (pH 8.0), 0.5 mM EDTA, 10 mM β-ME, 10 mM UDP, 10 mM MnCl2, and 5 mM EA2 with an equal volume of precipitant solution containing 7–10% polyethylene glycol 6000, 100 mM Heps (pH 7.0). Although EA2 was included in the crystallization solution, no electron density for the peptide was observed in the crystal structure. The crystals were grown over 0.3 ml of precipitant solution in 48-well plates, appeared in 3–4 days, and were transferred briefly (30–60 s) to a mother liqueur solution lacking protein but containing 10% glycerol before flash cooling in a nitrogen stream or liquid propane.

Diffraction intensities from single binary complex crystals were collected using 1° oscillations on an in-house Raxis-IV detector and a rotating anode generator (Rigaku/MSC) or at SER-CAT beamline 22ID at the Advanced Photon Source. Diffraction intensities from single ternary complex crystals were collected using 0.5° oscillations on the in-house Raxis-IV detector. Intensities from 560 (ternaire complex in-house) or 110 (binary complex Advanced Photon Source) or 90 (binary complex in-house) frames were integrated and scaled using the programs DENZO/SCALEPACK (17). The hT2 binary complex crystal structure was solved by molecular replacement using the program Phaser (18) and a search model prepared from separate catalytic (residues 95–427) and lectin (residues 428–548) domains of the mT1 crystal structure (Protein Data Bank code 1XHB) in which nonconserved residues were changed to alanine. Model building was done using XtalView (19). A partial model (74% complete) of the hT2 binary complex was built and refined against a 3.2 Å data set using several rounds of torsional simulated annealing in CNS (20) before changing to a higher resolution (2.75 Å) data set. The two noncrystallographic symmetry-related monomers of the hT2 binary complex were kept identical until the final rounds of energy minimization and B-factor refinement. The hT2 ternary complex structure was solved by molecular replacement using the program Phaser and a search model prepared from separate catalytic and lectin domains of the hT2 binary complex structure without UDP. Domain contact areas were calculated, and Figs. 1–3 were created using the program CCP4MG (21). Protein sequence alignments were created using ClustalX (22) and edited using Seaview (23). The structures were aligned using LSQMAN (24) and optimized using the “improve” option. Fig. 4 was created using PyMol.

Glycopeptides were synthesized by Anaspec, and enzyme activity was measured as previously described (25). The reactions were initiated by adding 0.05 pmol of enzyme, and incubation times were such that not more than 10% of the limiting substrate was converted to product. EA2 and Muc5Ac-3,13 were varied from 46.8 μM to 3 mM with UDP-GalNAc at 157.3 μM (0.06 μCi/mmole). Muc5Ac and Muc5Ac-3 were varied from 3 μM to 200 μM, and Muc5Ac-13 was between 7.8 μM and 1.0 mM with UDP-GalNAc at 165 μM (0.12 μCi/mmole). For UDP-GalNAc Km determinations, concentrations were varied from 10.4 μM to 207.3 μM with Muc5Ac at 300 μM. Pseudo first order kinetic constants were determined by nonlinear regression fitting to the Michaelis-Menten equation using the program GraphPad, and the initial velocities were determined from duplicate measurements.

**RESULTS**

**Overall Protein Fold**—Binary complex crystals (hT2-UDP-Mn2⁺) contained two molecules in the asymmetric unit, and electron density was observed for all residues except Thr90–Asn102/Lys103, Ala476, Gly571, and Gin571. Because the monomers are structurally similar (root mean square deviation = 0.39 Å) and each contains UDP and Mn2⁺, only details for the A monomer are described. Ternary complex crystals (hT2-UDP-EA2-Mn2⁺) contained a single monomer in the asymmetric unit, and electron density was observed for all residues except Lys192A, Gln571. Phi/psi angles of three residues in each complex (Lys192A, Gln571, and Lys571) were determined using the program LSQMAN (24) and optimized using the “improve” option. Fig. 4 was created using PyMol.

**Table 1** Crystallographic data and refinement statistics

|                   | hT2-UDP    | hT2-UDP-EA2 |
|-------------------|------------|-------------|
| **Data collection** |            |             |
| Space group       | P4 2 2     | P61         |
| a = b = c = 153.23 | a = b = 69.34 |
| c = 110.14        | c = 169.12 |
| Resolution (Å)    | 2.75 (2.92–2.75) | 1.64 (1.74–1.64) |
| Unique reflections | 34265 (5192) | 55904 (8825) |
| Completeness (%)  | 98.8 (96.1) | 99.4 (95.5) |
| Rmerge (%)        | 11.3 (67.5) | 5.5 (28.4) |
| **Molecules/asymmetric unit** | 2 | 1 |

**Refinement statistics**

|                          | hT2-UDP    | hT2-UDP-EA2 |
|--------------------------|------------|-------------|
| Resolution range         | 49.1–2.75  | 25.4–1.64  |
| R (%)                    | 22.5       | 17.8       |
| Rmerge (%)               | 28.3       | 20.6       |
| Root mean square deviations |          |            |
| Bond length (Å)          | 0.008      | 0.004      |
| Bond angle (°)           | 1.4        | 1.3        |
| Average B factor (Å²)    | 42.6       | 17.9       |
| No. of protein atoms     | 7720       | 3959       |
| No. of solvent atoms     | 0          | 495        |

*Statistics shown in parentheses are for the highest resolution shell.

**Overall**

- **Polypeptide α-N-Acetylgalactosaminyltransferase Structures**
Lys\textsuperscript{323A} and Lys\textsuperscript{323B} for the binary complex and Lys\textsuperscript{323}, Val\textsuperscript{330}, and Met\textsuperscript{493} for the ternary complex) were in disallowed regions of the Ramachandran plot, but because electron density for each residue was well defined, these angles were unchanged. The crystallographic data are shown in Table 1.

The catalytic domains of mT1 and hT2 are structurally similar (Fig. 1). The average root mean square deviation between corresponding C\textsubscript{α} carbons of the catalytic domains varies from 0.94 (hT2 binary/ternary complexes) to 1.13 Å (hT2 binary complex/mT1). Electron density for an additional 39 (ternary complex) or 42 (binary complex) residues compared with the mT1 structure was observed at the N termini of the hT2 structures. N-terminal residues in the hT2 complexes not seen in the mT1 structure are shown in red, and UDP and EA2 are shown in green. The putative carbohydrate-binding sites of the mT1 lectin domain are labeled as α, β, and γ.

FIGURE 1. The hT2 lectin domain is flexibly tethered to the catalytic domain. Catalytic domains of the hT2 binary and ternary structures (residues Asp\textsuperscript{115}–Phe\textsuperscript{361} and Gly\textsuperscript{374}–Asn\textsuperscript{432}) were aligned to the corresponding residues of the mT1 structure (Asn\textsuperscript{95}–Phe\textsuperscript{346} and Gly\textsuperscript{359}–Asn\textsuperscript{417}). Residues Arg\textsuperscript{347}–Thr\textsuperscript{356} of mT1 and Arg\textsuperscript{362}–Ser\textsuperscript{377} of hT2 were not used in the alignment because no density was observed for them in the mT1 structure. These residues are shown in yellow in the hT2 structures. N-terminal residues in the hT2 complexes not seen in the mT1 structure are shown in red, and UDP and EA2 are shown in green. The putative carbohydrate-binding sites of the mT1 lectin domain are labeled as α, β, and γ.

As expected from the mT1 structure, the lectin domain of each of the hT2 structures forms a β-trefoil fold, but the orientation of this domain relative to the catalytic domain in the two hT2 structures differs from that of mT1 and from each other (Fig. 1). The catalytic and lectin domains of mT1 form a close association in which \( \sim 645 \) Å\textsuperscript{2} of each domain is buried. This interaction is substantially reduced to \( \sim 325 \) Å\textsuperscript{2} domain in the hT2 binary complex and the two domains of the hT2 ternary complex do not associate except for the amino acids connecting them. In fact, residues Gln\textsuperscript{341}–Ala\textsuperscript{446}, which form the first strand of a β-sheet in the lectin domain of the hT2-UDP complex, unfold from this

FIGURE 2. Binding of UDP in the binary (A) and ternary (B) complexes. The main chain of residues of the flexible loop (Arg\textsuperscript{362}–Ser\textsuperscript{373}) is shown by the yellow ribbon, whereas the gray ribbon represents the main chain residues that have unwound from an adjacent α-helix (orange). Hydrogen bonds are shown by the blue dashed lines. The side chains of residues Thr\textsuperscript{6} and Thr\textsuperscript{7} of EA2 are shown in green in B. Simulated annealing omit electron density maps (salmon mesh) surrounding UDP and the Mn\textsuperscript{2+} ion were calculated with ligands omitted and are shown contoured at 2.5 σ.
Polypeptide α-N-Acetylgalactosaminyltransferase Structures

sheet in the hT2-UDP-EA2 structure and extend the peptide tether linking the catalytic and lectin domains.

**UDP Binding**—The binding of UDP differs dramatically between the binary and ternary complexes (Fig. 2). Compared with the ternary complex, UDP is inverted in the binary complex with the ribose group shifted out of its ternary complex pocket to face bulk solvent. Indeed, UDP is a product of the reaction catalyzed by hT2, and its observed conformation in the binary complex is consistent with UDP leaving the active site following catalysis. Residues Arg362–Ser373 of a flexible loop fold out of the way to accommodate this orientation, and the C terminus corresponding lengths in Ångstroms. Hydrophobic interactions are shown by the blue dashed lines (Tg).

**TABLE 2**

The UDP-binding residues of active ppGalNAcT isoforms are highly conserved

| ppGalNAcT Isoform | Residue |
|--------------------|---------|
| T2, pGANT2, Gly4   | H D R L D S H W H R H Y |
| T14                | H D R L D S H W H R H Y |
| G1y5a,b,c          | C D R L D S H I W H R S Y |
| T11, pGANT-35a     | C D R L D S H I W H R S Y |
| T3                 | V H D R L D A H I W H R S H |
| T4, T12            | A Y D R L C H V W H P A Y |
| T15                | C H D R A D A H L R H Q D H |
| T1, T13            | V H D R L D A H I W H R T Y |
| pGANT5             | V H D R L D A H I W H R S Y |
| pGANT3, Gly3       | V H D R L D A H I W H R T Y |
| T5                 | C V D R L D S H W H R N Y |
| pGANT1             | I F D R L A H Y Y Y Y Y Y Y |
| T7                 | V Y D R L D S H I Y H R Q Y |
| TgT1               | V Y D R L D S H F W H R S Y |
| TgT2               | V Y D R L D S H F W H R S Y |
| T7                 | V H D R L D A H I W H R N P |
| pGANT7             | V H D R L D A H V W H R M Y |
| pGANT6             | I F D R L D S H I W H R Q - |
| pGANT4             | I F D R L D S H I W H R Q - |
| T16                | P H D R L D S H I W H R V Y |
| pGANT8             | T H D Q L D S H I W H R I Y |

 disregard the differences in the loop residue positions and UDP orientation, the conformation of several residues that hydrogen bond to the uridine ring and the phosphate moieties of UDP is substantially unaltered between the binary and ternary complexes (Fig. 2). These residues include Thr143, Asp176, and Arg201, which bind the uridine ring, and Asp224 and His226 of the DXH motif and His229, which bind the phosphate groups via coordination with the Mn2+ ion. Several residues of mT1 corresponding to those of hT2 mediating UDP binding have been mutated, and the activity of enzymes carrying these mutations have been described (27). Adjacent to UDP is a cavity occupied by five water molecules (see Fig. 4) presumed to be the GalNAc-binding pocket based upon similarly located pockets shown to be the sites of sugar binding for other retaining glycosyltransferases (28–30). This pocket is lined by invariant (Arg200 and Glu314) and highly conserved (Trp331 and Asn335) residues.

**EA2 Binding**—A schematic diagram of EA2 binding indicating the hydrogen bonds and hydrophobic interactions it forms with hT2 is shown in Fig. 3. Electron density for the first 4 residues of EA2 was absent so only residues Ser5–Lys13 are shown. EA2 binds in an extended conformation with each amino acid except Lys13, assuming phi/psi angles favored by β-strands. The binding of acceptor substrates in an extended conformation was previously hypothesized based upon secondary structure predictions of residues flanking potential glycosylation sites (31). The side chain hydroxyl of Thr7, shown to be the preferred residue of initial glycosylation of EA2 by hT2 and several other isoforms (32), forms a strong hydrogen bond with a β-phosphate oxygen of UDP and is ideally located to be the GalNAc acceptor. Analysis of EA2 binding shows that the majority of hydrogen bonds between hT2 and EA2 occur between EA2 residues Ser5–Pro8, whereas hydrophobic interactions dominate the binding of residues Ala9–Lys13.

**FIGURE 3.** Hydrogen bonds and hydrophobic interactions mediating binding between hT2 and EA2. EA2 is shown as a stylized drawing with yellow carbon atoms, and the individual residues Ser5–Lys13 are labeled in red. Hydrogen bonds are shown by the blue dashed lines along with their corresponding lengths in Ångstroms. Hydrophobic interactions are shown by the red “eyelashes.” Water molecules are shown as red spheres. The diagram was created by editing the output from the program Ligplot (47).
EA2 binds in a shallow cleft on the surface of hT2 that broadens toward the C-terminal end of EA2 and narrows toward the N terminus of EA2 (Fig. 4). Residues Pro^δ–Lys^δ of EA2 bind against one side of the cleft with the cyclic side chain of Pro^δ inserted into a pocket formed by Val^255, Leu^270, Trp^282, and Phe^361 and stacked against the side chain of Trp^282. The opposite side of the cleft contains additional, shallower pockets, two of which are occupied by water molecules. Threonine 7 points into the UDP-binding pocket with the hydroxyl group hydrogen bonded to a β-phosphate oxygen atom and the methyl group directed into a hydrophobic cavity lined by the side chains of Phe^360, Ala^307, and Phe^361. Threonine 6 packs against a ridge in the surface of the enzyme formed by residues Arg^362–His^365 of the flexible loop with the hydroxyl group of T6 hydrogen bonding to the main chain carbonyl of Arg^362. Approximately 540 Å² of surface is covered by EA2, and only 22% of this area is contributed by UDP and residues within the mobile loop of hT2. Thus, 78% of the EA2-binding site is preformed.

**Catalytic Domain Activity**—The lack of interaction between the hT2 catalytic and lectin domains observed in the ternary complex suggests that the catalytic domain may not require the lectin domain for activity. Based on the crystal structure we designed and expressed the hT2 catalytic domain (residues 74–440) lacking the entire lectin domain and compared its activity to the full-length enzyme against peptide and glycopeptide acceptors. As shown in Table 3, both $k_{cat}$ and $K_m$ values (and thus the $k_{cat}/K_m$ ratio) for peptides EA2 and Muc5Ac are similar for full-length hT2 and the hT2 catalytic domain. However, removal of the hT2 lectin domain reduced glycopeptidase $k_{cat}/K_m$ ratios of the catalytic domain 4–18-fold compared with full-length hT2 (Table 3). Thus, the absence of the hT2 lectin domain affected the transfer of GalNAc to the glycopeptidases but not to the peptide substrates tested. For the Muc5Ac-3 glycopeptidase, the smaller $k_{cat}/K_m$ value was dominated by a reduced $k_{cat}$ whereas for the Muc5Ac-13 glycopeptidase it was dominated by an increase in $K_m$. For the Muc5Ac-3/13 glycopeptidase, the diminished $k_{cat}/K_m$ value was caused by both a smaller $k_{cat}$ and larger $K_m$. Because both the $k_{cat}$ and $K_m$ values represent a collection of individual rate constants (and thus are apparent catalytic constants) that have not been determined, the specific step(s) of the catalytic mechanism most affected by the absence of the lectin domain remains unknown. $K_m$ values for UDP-GalNAc were similar for full-length hT2 (11.5 ± 2.4 μM) and the hT2 catalytic domain (7.9 ± 2.3 μM).

**DISCUSSION**

The properties of ppGalNAcT acceptor substrates have been studied for nearly 15 years through data base analyses of known O-glycosylation sites (33, 34), in vitro studies using defined peptide acceptors (31, 35), sequencing of tissue-extracted mucins (36, 37), and more recently systematic variation of the amino acids flanking acceptor threonine residues (38). Common findings among these investigations are a strong bias for proline 3 residues C-terminal to the site of glycosylation (the “+3 site”) and a preference of threonine over serine for glycosylation. The molecular basis for the proline preference is explained by the hT2-UDP-EA2 structure, which shows that the +3 proline of EA2 (Pro^δ) inserts into a cavity surrounded by hydrophobic residues (Val^255, Leu^270, Trp^282, and Phe^361) and stacks against the side chain of Trp^282 (Fig. 4). Each of these “proline pocket” residues except Leu^270 is conserved in the T1 isoform (Table 4), which also prefers proline at the +3 site, although not to the same degree as hT2 (39). There is also considerable variability in the residues lining this pocket, and the extent to which they and others control acceptor specificity will be revealed as peptide preferences for individual, purified isoforms are determined. In turn, this information should lead to more reliable predictions of isoform-specific substrates and/or sites of mucin-type glycosylation than are currently available (40). The preference for glycosylating threonine versus serine residues can be explained by the fact that loss of the methyl group would lead to a loss of interaction energy provided by the hT2 residues that bind the methyl group.

The previously determined structure of mT1 revealed that the catalytic and lectin domains form a close association in which each of the three carbohydrate-binding sites of the lectin domain face the same side of the enzyme as the catalytic site (Fig. 1). This suggested a model in which these carbohydrate-binding sites permit the capture of an array of
Polypeptide α-N-Acetylgalactosaminytransferase Structures

TABLE 3
Kinetic parameters of full-length hT2 (FL) and the hT2 catalytic domain (CD) for peptide (EA2 and Muc5Ac) and glycopeptide (Muc5Ac-3, Muc5Ac-13 and Muc5Ac-3,13) substrates

| Peptide | Sequence | Full-length hT2 | hT2 Catalytic Domain | FL | CD | Ratio |
|---------|----------|----------------|----------------------|----|----|-------|
| EA2     | PTTSDDTSFAPPTKK | 3.70 ± 0.22 | 943 ± 120 | 3.29 ± 0.14 | 1157 ± 130 | 1.28 ± 0.09 | 1.1 ± 0.3 |
| Muc5Ac  | GTSPPSVSVTSSGTAP | 0.29 ± 0.02 | 20.2 ± 4.6 | 0.22 ± 0.01 | 17.5 ± 4.3 | 0.8 ± 0.1 | 1.1 ± 0.4 |
| Muc5Ac-3 | GTSPPSVSVTSSGTAP | 1.56 ± 0.07 | 280 ± 4.8 | 0.31 ± 0.02 | 32.9 ± 5.8 | 0.16 ± 0.1 | 5.9 ± 1.5 |
| Muc5Ac-13 | GTSPPSVSVTSSGTAP | 0.85 ± 0.04 | 16.5 ± 4.0 | 0.57 ± 0.02 | 200 ± 31 | 0.26 ± 0.01 | 18.0 ± 5.4 |
| Muc5Ac-3,13 | GTSPPSVSVTSSGTAP | 0.82 ± 0.06 | 406 ± 79 | 0.47 ± 0.04 | 933 ± 204 | 0.54 ± 0.01 | 1.5 ± 4.3 |

The extant sites of previously glycosylated substrates for subsequent GalNAc transfer and that was consistent with the mT1-catalyzed, in vitro pattern of glycosylation of a Muc1-based peptide (16). The structures of hT2 presented in this paper indicate that, at least for this isoform, a new variable should be considered in a model of substrate capture that incorporates conformational changes within the residues linking the two domains. This “flexible tether” model suggests that lectin domain mobility endows certain ppGalNAcTs with an even greater capacity to adapt to and capture glycosylated substrates, thus ensuring the high density of glycosylation characteristic of mucin domains. Whether this model applies to other isoforms (including T1) will require solving the binary and/or ternary complex structures of additional ppGalNAcTs. Determining whether substrate binding triggers lectin domain mobility will require solving the corresponding apo enzyme structures. It should be noted that the relative orientations of the catalytic and lectin domains seen in the hT2 binary and ternary complex structures may have been influenced in part by crystal packing forces and thus may not accurately reflect in vivo conformations. However, our conclusion that the two conformations observed imply an inherent flexibility between the two domains is not dependent on either of the specific orientations.

Our result showing that the lectin domain plays a role in the transfer of GalNAc to glycopeptide but not peptide substrates is consistent with studies showing that point mutations in the putative carbohydrate-binding sites of the lectin domain diminish activity toward glycopeptide but not peptide substrates (14, 15). However, it was noted that a lectin-dependent model of substrate capture is not the only mechanism used by ppGalNAcTs, at least with some glycopeptides used to monitor transferase activity (16). This conclusion was based on results showing that ppGalNAcT-7 and T-10 transfer GalNAc to threonine 6 when presented with EA2 containing GalNAc on threonine 7 (41). The structure of EA2 bound to hT2 shows how GalNAc could be attached to the residue C-terminal to the current site of glycosylation because the side chain of this residue points away from the enzyme surface (Fig. 4).

Our finding that the hT2 lectin domain is dispensable for catalytic activity contrasts with a prior study examining the effects of lectin domain truncations on the function of rat ppGalNAcT-1 (rT1). It was shown that the removal of 12 or more residues from the C terminus of the rT1 lectin domain eliminated enzymatic activity for both peptide (PPDAATAAPLR) and apomucin acceptors, even though enzyme expression was only moderately reduced (14). Our results show that ppGalNAcTs lacking a lectin domain may still be active. One such transferase, Gly8, has been identified in *C. elegans*, but no activity has been demonstrated for this isoform, perhaps because a suitable acceptor substrate has not been found (3).

The pattern of substrate binding for several glycosyltransferases is ordered sequential with the sugar nucleotide donor binding first followed by the acceptor substrate (28 – 30, 42). In contrast, a kinetic investigation of the mechanism of bovine ppGalNAcT-1 using erythropoietin-derived peptide EPO-T (PPDAATAAPLR) indicated that substrate binding follows a random sequential pattern (43). The current structure of EA2 bound to hT2 helps rationalize this finding because the EA2-binding site and, by extension, the EPO-T binding site is largely independent of interactions with both UDP and residues Arg162–Ser373 of the flexible loop. However, with full-length protein substrates that make more extensive interactions with the loop residues, binding may well become ordered sequential with UDP-GalNAc binding first followed by acceptor protein.

The catalytic mechanism of retaining glycosyltransferases remains undetermined, but recently an aspartic acid has been identified as a potential nucleophile for the lipopolysaccharide transferase LgtC (44). Because this aspartate is 8.9 Å away from the donor substrate, the authors noted that a conformational change would be required during catalysis for it to function as the nucleophile. Inspection of hT2 for potential nucleophiles surrounding the putative GalNAc pocket identified residues Arg192–Ser335 of the flexible loop. However, with full-length protein substrates that make more extensive interactions with the loop residues, binding may well become ordered sequential with UDP-GalNAc binding first followed by acceptor protein.

The catalytic mechanism of retaining glycosyltransferases remains undetermined, but recently an aspartic acid has been identified as a potential nucleophile for the lipopolysaccharide transferase LgtC (44). Because this aspartate is 8.9 Å away from the donor substrate, the authors noted that a conformational change would be required during catalysis for it to function as the nucleophile. Inspection of hT2 for potential nucleophiles surrounding the putative GalNAc pocket identified residues Arg192–Ser335 of the flexible loop. However, with full-length protein substrates that make more extensive interactions with the loop residues, binding may well become ordered sequential with UDP-GalNAc binding first followed by acceptor protein.
Polypeptide α-N-Acetylgalactosaminyltransferase Structures

ppGalNAcT catalytic mechanism and perhaps that of other retaining glycosyltransferases.

A key challenge for investigating ppGalNAcT function is the identification of unknown, isoform-specific ppGalNAcT protein substrates. The large size of this transference family coupled with the fact that each shares a common donor precludes using radiolabeled UDP-GalNAc for identifying such protein substrates in cells or tissues expressing multiple ppGalNAcTs because it would not be possible to know which isoform was responsible for labeling a given substrate. A similar situation applies to the Src family of protein kinases, each of which uses ATP as the donor substrate. A structure-based approach has been used to create ATP analogs that serve as substrates for mutant but not wild-type Src kinases (45, 46). Labeling of cell lysates with the ATP analog and mutant kinases has been successful in identifying direct substrates of a given Src kinase (45). The hT2 structures determined in this study will aid the design of UDP-GalNAc analog/mutant ppGalNAcT pairs that can be used to identify isoform-specific ppGalNAcT acceptor substrates and thus will help to define the biological functions of this transference family.

Acknowledgments—We thank Dr. Susan Buchanan (NIDDK, National Institutes of Health) for helpful discussions and x-ray equipment for critical reading of the manuscript. We also thank Dr. Fred Dyda (NIDDK, National Institutes of Health) for helpful discussions and x-ray equipment maintenance.

REFERENCES
1. Ten Hagen, K. G., and Tran, D. T. (2002) J. Biol. Chem. 277, 22616–22622
2. Schwientek, T., Bennett, E. P., Flores, C., Thacker, J., Hollmann, M., Reis, C. A., Behrens, J., Mandel, U., Keck, B., Schafer, M. A., Haselmann, K., Zubarev, R., Roepstorff, P., Burchell, J. M., Taylor-Papadimitriou, J., Hollingsworth, M. A., and Clausen, P. (2002) J. Biol. Chem. 277, 22623–22638
3. Hagen, F. K., and Nehrke, K. (1998) J. Biol. Chem. 273, 8268–8277
4. Xia, L., Ju, T., Westmuckett, A., An, G., Ivanciu, L., McDaniel, J. M., Lupu, F., Cummings, R. D., and McEver, R. P. (2004) J. Biol. Chem. 279, 44–59
5. Kawakubo, M., Ito, Y., Oikimura, Y., Kobayashi, M., Sakura, K., Kasama, S., Fukuda, M. N., Fukuda, M., Katsuyama, T., and Nakayama, J. (2004) Science 305, 1003–1006
6. Rosen, S. D., Hwang, S. T., Giblin, P. A., and Singer, M. S. (1997) Biochem. Soc. Trans. 25, 428–433
7. Borsig, L., Wong, R., Fereissaco, J., Nadeau, D. R., Varki, N. M., and Varki, A. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 3352–3357
8. Topaz, O., Shurman, D. L., Bergman, R., Indelman, M., Ratajczak, P., Mizrachi, M., Ratajczak, A., Richard, G., and Sprecher, E. (2004) J. Biol. Chem. 279, 15–21
9. Ten Hagen, K. G., Bedi, G. S., Tetaert, D., Kingsley, P. D., Hagen, F. K., Balys, M. M., Shah, K., Liu, Y., Deirmengian, C., and Shokat, K. M. (1997) J. Biol. Chem. 272, 1097–1106
10. Hagen, F. K., Hazes, B., Raffo, R., deSa, D., and Tabak, L. A. (1999) J. Biol. Chem. 274, 6797–6803
11. Boix, E., Zhang, Y., Swaminathan, G. J., Brew, K., and Acharya, K. R. (2002) J. Biol. Chem. 277, 28310–28318
12. Persson, K., Ly, H. D., Dieckelmman, M., Wakarchuk, W., Withers, S. G., and Strynadka, N. C. (2001) Nat. Struct. Biol. 8, 166–175
13. Pedersen, L. C., Dang, J., Taniguchi, F., Kita, H., Krahn, J. M., Pedersen, L. C., Sugahara, K., and Negishi, M. (2003) J. Biol. Chem. 278, 14420–14428
14. Elhammer, A. P., Poorman, R. A., Brown, E., Maggiora, L. L., Hoogerheide, J. G., and Kezdy, F. J. (1993) J. Biol. Chem. 268, 10029–10038
15. Pratt, M. R., Hang, H. C., Ten Hagen, K. G., Rarick, J., Gerken, T. A., Tabak, L. A., and Bertozzi, C. R. (2004) Chem. Biol. 11, 1009–1016
16. O’Connell, B., Tabak, L. A., and Ramasubbu, N. (1991) Biochem. Biophys. Res. Commun. 180, 1024–1030
17. Wilson, I. B., Gavel, Y., and von Heijne, G. (1991) Biochem. J. 275, 529–534
18. O’Connell, B. C., Hagen, F. K., and Tabak, L. A. (1992) J. Biol. Chem. 267, 25010–25018
19. Gerken, T. A., Gilmore, M., and Zhang, I. (2002) J. Biol. Chem. 277, 7736–7751
20. Gerken, T. A., Owens, C. L., and Pausmarty, M. (1997) J. Biol. Chem. 272, 9709–9719
21. Gerken, T. A. and Rarick, J. (2004) Glycobiology 14, 1190
22. Gerken, T. A., Tek, C., and Rarick, J. (2004) Biochemistry 43, 9888–9900
23. Julenius, K., Molgaard, A., Gupta, R., and Brunak, S. (2005) Glycobiology 15, 153–164
24. Ten Hagen, K. G., Bedi, G. S., Tetaert, D., Kingsley, P. D., Hagen, F. K., Bals, M. M., Beres, T. M., Degand, P., and Tabak, L. A. (2001) J. Biol. Chem. 276, 17495–17404
25. Morrison, J. F., and Ebner, K. E. (1971) J. Biol. Chem. 246, 3977–3984
26. Wragg, S., Hagen, F. K., and Tabak, L. A. (1995) J. Biol. Chem. 270, 16947–16954
27. Lahm, L., Chiu, C., Li, S. H., Ho, S., Wakarchuk, W., Strynadka, N. C., and Withers, S. G. (2004) J. Biol. Chem. 279, 28339–28344
28. Ulrich, S. M., Persson, K., Ly, H. D., Strynadka, N. C., and Withers, S. G. (2000) J. Biol. Chem. 275, 28339–28344
29. Shah, K. M., Ulrich, S. M., and Shokat, K. M. (2003) Glycobiology 13, 9988–9900
30. Hasuo, K., Liu, Y., Deimengian, C., and Shokat, K. M. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 5356–5357
31. Wallace, A., C., Laskowski, R. A., and Thornton, J. M. (1995) Protein Eng. 8, 127–134
32. Ten Hagen, K. G., Tran, D. T., Gerken, T. A., Stein, D. S., and Zhang, Z. (2003) J. Biol. Chem. 278, 35039–35048
33. Tetaert, D., Ten Hagen, K. G., Richet, C., Boersma, A., Gagnon, J., and Degand, P. (2001) Biochem. J. 357, 313–320