Identification of Common and Unique Peptide Substrate Preferences for the UDP-GalNAc:Polypeptide α-N-acetylgalactosaminyltransferases T1 and T2 Derived from Oriented Random Peptide Substrates*

Thomas A. Gerken, Jayalakshmi Raman, Timothy A. Fritz, and Oliver Jamison

From the W. A. Bernbaum Center for Cystic Fibrosis Research, Departments of Pediatrics and Biochemistry, Case Western Reserve University School of Medicine, Cleveland, Ohio 44106 and the Section on Biological Chemistry, NIDDK, National Institutes of Health, Bethesda, Maryland 20892

A large family of UDP-GalNAc:polypeptide α-N-acetylgalactosaminyltransferases (ppGalNAc Ts) catalyzes the first step of mucin-type protein O-glycosylation by transferring GalNAc to serine and threonine residues of acceptor polypeptides. The acceptor peptide substrate specificity and specific protein targets of the individual ppGalNAc T family members remain poorly characterized and poorly understood, despite the fact that mutations in two individual isoforms are deleterious to man and the fly. In this work a series of oriented random peptide substrate libraries, based on the GAGXXXTXXAGAGK sequence motif (where X = randomized positions), have been used to obtain the first comprehensive determination of the peptide substrate specificities of the mammalian ppGalNAc T1 and T2 isoforms. ppGalNAc T-glycosylated random peptides were isolated by lectin affinity chromatography, and transferase amino acid preferences were determined by Edman amino acid sequencing. The results reveal common and unique position-sensitive features for both transferases, consistent with previous reports of the preferences of ppGalNAc T1 and T2. The random peptide substrates also reveal additional specific features that have never been described before that are consistent with the x-ray crystal structures of the two transferases and furthermore are reflected in a data base analysis of in vivo O-glycosylation sites. By using the transferase-specific preferences, optimum and selective acceptor peptide substrates have been generated for each transferase. This approach represents a relatively complete, facile, and reproducible method for obtaining ppGalNAc T peptide substrate specificity. Such information will be invaluable for identifying isoform-specific peptide acceptors, creating isoform-specific substrates, and predicting O-glycosylation sites.
thought to serve a “filling in” activity, by completing the glyco-
sylation of densely glycosylated regions of the polypeptide (28).

Obtaining the detailed peptide/glycopeptide specificity of individual ppGalNAc transferase isoforms is an enormous and
important task that has not been adequately addressed (15). To
begin to address this deficiency, we have used as substrates, for
ppGalNAc T1 and T2, the large apomucin tandem repeat
domains containing multiple Ser and Thr residues from the pig
dog salivary gland mucins (PSM and CSM) (29, 30). Com-
bined, these substrates contain over 50 unique glycosylation
sites. By characterizing the rates of site-specific glycosylation,
in terms of neighboring glycosylation and peptide sequence, we
have developed a kinetic model capable of approximating the
observed ppGalNAc T1 and T2 glycosylation patterns (29, 30).
This approach has identified a ppGalNAc T2 motif and is capa-
bility of reproducing the ppGalNAc T2 glycosylation patterns for
the IgA1 hinge domain (30). Although significant specificity
information can be obtained from the study of these substrates,
we are still limited by the relatively small number of unique
glycosylation sites available in these substrates compared with
the vast number of possible peptide glycosylation sites found in
nature.

We now report on an approach for obtaining a more com-
plete determination of the peptide substrate specificities of
ppGalNAc T1 and T2 by utilizing a series of oriented random
peptide substrates, each containing a central Thr or Ser residue
flanked by a series of randomized residues. Similar oriented
random peptide approaches have been successfully used for the
characterization of the peptide substrate preferences of phos-
phopeptide-binding proteins and protein kinases from which
optimal and specific peptide substrate sequences have been
obtained (31–33). Our results with these substrates reveal com-
mon and unique position-sensitive features for both trans-
f erases. Many of the preferences displayed in our random pep-
tide work are consistent with previous reports on the substrate
preferences of ppGalNAc T1 and T2 (30) and are reflected in
data base analyses of O-glycosylation sites. However, our
random peptide work reveals numerous specific features that
have never been detected before for either transferase. The
characterization of ppGalNAc transferase isoform specific-
ity by this approach will be invaluable for the identifica-
tion of isoform-specific substrates, the creation of isoform-
specific inhibitors, and the prediction of O-glycosylation
sites. These substrates will also be important for character-
izing the conservation of specificity across divergent species
of orthologous ppGalNAc Ts.

**EXPERIMENTAL PROCEDURES**

Soluble recombinant bovine ppGalNAc T1 and initial ali-
quots of human ppGalNAc T2 (34, 35) were the gifts of Dr. Ake
Elhammer (Kalamazoo, MI) as described previously (36, 37)
and as used in our earlier work (29, 30). Additional human
ppGalNAc T1 and T2 were gifts of Dr. Lawrence Tabak (Office
of Director, NIDCR, National Institutes of Health, Bethesda) as
described (38). All transferases were highly purified revealing
single bands on SDS-PAGE (data not shown) and were suit-
able for crystallization studies. Random peptide substrates I,
IIs, II, IIs, III, IV, and V (see Table 1), having the sequence
GAGAXXT( or S),XXXAGAGK, where X = randomized posi-
tions, were custom-synthesized by QCB Inc. (formally Bio-
Source Inc.) (Hopkinton, MA). The quality of the synthesis
was assessed by pulse liquid phase Edman amino acid
sequencing on a Prociex 494 protein sequencer (Applied Bio-
systems, Foster City CA). Typically, the mole fraction of the
residues at each random cycle ranged between 5 and 20%. It
was also found that the mole fractions varied with different
synthetic batches of peptide. Prior to use, synthetic peptides
were pH-adjusted to ~7.5 with dilute NaOH and/or HCl and
lyophilized several times from water. Stock random peptide
solutions (100 mg/ml) were aliquoted and frozen for subse-
quent use. UDP-GalNAc was purchased from Sigma, and
UDP-[3H]GalNAc was purchased from ARC Inc. (St. Louis,
MO). [3H]GalNAc content was determined by liquid scintil-
cation counting on a Beckman model LS5801 liquid scintil-
cation counter.

**Glycosylation of Random Peptide by Bovine ppGalNAc T1**

and Human ppGalNAc T2—Typical random peptide glyco-
sylation conditions consisted of 0.05–0.10 ml of the following:
10 mg/ml random peptide, 10 mM MnCl2, ~22 μg/ml ppGalNAc
T1, or ~16 or ~25 μg/ml ppGalNAc T2, 2 mM UDP-GalNAc
(3H-labeled, 0.1 μCi/0.10 ml), 1:100 dilution protease inhibit-
or P8340 and P88499 (Sigma), 1 mM EDTA, and 100 mM
HEPES, final pH 7.5, giving approximately a 3:1 mole ratio of
peptide acceptor to UDP-GalNAc. Incubation times ranged from
0.5 to ~24 h (typically overnight for both transferases) at
37 °C and were stopped by addition of a 2X volume of 250 mM
EDTA. For random peptides III, IV, and V, usually less than 5%
of the peptide was glycosylated by ppGalNAc T1 and less than
3% by ppGalNAc T2 in overnight incubations based on radio-
elabel incorporation. Incorporations were typically 2-fold higher
for peptide I compared with the other peptides. Random Ser
peptide Is, when glycosylated overnight under similar conditions
as the Thr peptide, gave ~0.5 and ~1% incorporations for ppGal-
NAc T1 and T2, respectively. Nonhydrolyzed UDP-GalNAc and
UDP were removed on small ~3-ml Dowex 1-X8 anion exchange
columns (Bio-Rad) after dilution to 1–2 ml in water. After lyoph-
ilization of the eluate, the random (glyco)peptide mixture was sep-
ated by gel chromatography on Sephadex G-10 (0.7 × 113-cm column) (Amersham Biosciences) in 50 mM acetic
acid buffer, pH 4.5 (with NH4OH). Fractions were monitored
in arbitrary units on a Beckman 121S. Fractions were monitored
for absorbance at 220 nm and for 3H-label by scintillation coun-
ing of 30–50 μl of every other fraction. High molecular weight radioactive random glycopeptide
fractions were pooled and lyophilized. In some instances the
gel filtration chromatography was performed using Bio-Rad
P2 (0.7 × 113-cm column) in acetic acid buffer. Reported molecular weight exclusion limits for Sephadex G-10 and
Bio-Rad P2 are 700 and 1,800 daltons respectively.

**Isolation of Glycosylated Random Peptide on Mixed Bed Lec-
tin Column Chromatography**—Glycosylated random peptides
from Sephadex G-10 chromatography were isolated from non-

---

3 Combined, protease inhibitors P8340 and P8849 contain phenylmethylsul-
fonyl fluoride, phosphoramidon, trans-epoxysuccinyl-l-leucylamido(4-
guanidino)-butane, 4-(2-aminoethyl) benzenesulfonyl fluoride, pepstatin
A, bestatin, leupeptin, and aprotinin.
glycosylated random peptides by passage across a 10-ml (22 × 0.8 cm) mixed bed immobilized lectin column. The column bed contained 2 ml each of the agarose-bound lectins SJA (Sophora japonica), SBA (Glycine max), and HPA (Helix pomatia) from EY Laboratories (San Mateo, CA) in addition to 4 ml of the VVA (Vicia villosa, predominantly the B4 isofrom) lectin from Vector Laboratories ( Burlingame, CA). After equilibration with buffer (10 mM NaCl, 2.5 mM Tris, pH 8.0, and 0.002% NaN₃) at 4 °C, 1 ml of partially glycosylated random peptide was loaded on the column in the presence of freshly added protease inhibitors P8340 and P8849 and washed with ~30 ml of buffer at a flow rate of ~8 ml/h. A portion of this sample was removed for Edman sequencing prior to addition of inhibitors or buffer. Bound glycopeptide was released from the lectin column with 5 ml of 5 or 10 mM GalNAc in buffer, and the column was re-equilibrated with an overnight wash with buffer. Fractions were monitored for peptide (and added GalNAc) by absorbance at 220 nm and for ³H-labeled glycopeptide by scintillation counting. No glycopeptide was found to be irreversibly bound to the column based on the integration of the ³H radioactivity released from the column compared with that which was loaded. GalNAc released ³H-glycopeptide fractions were pooled, lyophilized, and separated from salt, buffer, and GalNAc by gel filtration chromatography on Sephadex G-10 (or Bio-Rad P2) as described above.

Random Glycopeptide Amino Acid Sequencing and Data Analysis—Pulsed liquid Edman amino acid sequencing of the random (glyco)peptides was performed on an Applied Biosystems Procise 494 Edman protein sequencer as described previously using a C18 PTH column temperature of 45 °C (peptides I and V) or 55 °C (peptides II, III, and IV) (30). Peak areas were provided by the ABI Procise software and if necessary corrected for base-line errors by cutting and weighing of the peaks. Data analysis was performed using an average minimum value baseline correction approach as described previously (39). However, to prevent over-correction of the base line, residues flanking (and including) the central Thr or Ser residue (i.e., cycles 7–9) were assigned minimum values of 0 for inclusion in the base-line running average. Response values and factors for the determination of the extent of glycosylation of Thr and Ser were based on earlier work on authentic synthetic glycopeptide and standards (39). The mole fraction of each residue type (see Table 1) was calculated from the amino acid composition obtained from the Edman sequencing analysis at each randomized position (i.e., cycles 5–11). Mole fractions were obtained for both the random (glyco)peptide mixture immediately prior to loading on to the lectin column and from the Sephadex G-10 purified random glycopeptide that was released from the lectin column by GalNAc. Transferase-specific enhancement factors were obtained by dividing the bound glycopeptide mole fractions by the control pre-lectin random (glyco)peptide mole fractions for each individual amino acid residue at each random cycle position. For each peptide and transferase at least two independent experiments were performed and the data averaged.

In the course of the peptide sequence determinations of peptides III and V, it was found that the Lys residue content diminished relative to other residues when low amounts of (glyco-peptide were sequenced. To compensate for this behavior both the pre-lectin and post-lectin sequence determinations were performed using similar amounts of peptide where identical losses of Lys would occur in both sequence determinations. In addition, sequence determinations of peptide V revealed the near-complete loss of Trp in the pre- and post-lectin sequence determinations. It was subsequently determined that the Trp residue was lost principally after the initial Sephadex G-10 gel filtration step. Bio-Rad P2 gel filtration chromatography also significantly reduced the Trp content of peptide V. Multiple efforts to inhibit the presumed Trp oxidation at this step were not successful. Therefore, enhancement values for Trp could not be determined.

Peptide Substrate Kinetic Studies—Human ppGalNAc T1 and T2 activities against the selected peptides in Table 2 were measured as described previously (40). The human ppGalNAc T1 utilized in these studies gave a Kₘ of 0.102 ± 0.018 mM and a Vₘₐₓ of 7.1 ± 0.3 nmol/min/μg with the EA2 peptide PTTD-STTPAPTTK. An identical batch of human ppGalNAc T2 gave a Kₘ of 0.943 ± 0.119 mM and Vₘₐₓ of 3.7 ± 0.2 nmol/min/μg with the EA2 peptide (38). The full-length 16-residue peptides containing the flanking GAGA- and -AGAGK sequences and having a central Thr residue were used as substrates. Reactions were initiated by adding 0.5 to 0.015 pmol of enzyme, adjusted to ensure that not more than 10% of the limiting substrate was converted to product in the reaction time of 1 h. Peptide concentrations varied between 47 μM and 6.0 mM except in reactions of ppGalNAc T2 with the optimal peptides where the peptide concentrations were varied between 31 μM and 2 mM. UDP-GalNAc was held constant at 107 μM (0.06 μCi/mmol). 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. Data in Table 2 represent an average of two independent measurements. Peptides were synthesized and high pressure liquid chromatography-purified by the Facility for Biotechnology Resources, Center for Biologics Evaluation and Research, National Institutes of Health.

It should be noted that the kinetic studies on the individual peptides were performed using the human ppGalNAc T1 transferase, whereas the random peptide experiments were performed using the bovine ppGalNAc T1 transferase. The human and bovine ppGalNAc T1 sequences are 98.9% identical with only 6 conservative residue differences out of 559 residues. Because none of the differing residues are located near the peptide or UDP-GalNAc substrate-binding sites, the peptide substrate specificities of the bovine and human transferases are expected to be identical.

RESULTS

We report a novel approach for the systematic characterization of the peptide substrate specificity of ppGalNAc transferases utilizing random peptide substrates containing a single central Thr (or Ser) residue.

Key to the approach is the use of lectin affinity chromatography to isolate glycosylated random peptide from nonglycosylated peptide and the subsequent use of Edman amino acid sequencing to quantify the amino acid composition at each ran-
TABLE 1

ppGalNAc transferase random peptide substrates

| Peptide: GAGXXX(T/S)XXXAGAGK | Random residues, X = | No. of unique sequences | No. of determinations* |
|-------------------------------|----------------------|-------------------------|------------------------|
| Thr peptides                  |                      |                         |                        |
| I                             | G, A, P, R, E, N, Y, V| 262,144                 | 8                      |
| II                            | G, A, P, K, D, Q, H, F, I, M, W| 1,771,561              | 4                      |
| III                           | G, A, P, R, D, Q, K, F, I| 531,441                | 2                      |
| IV                            | G, A, P, R, E, Q, H, L | 262,144                 | 2                      |
| V                             | G, A, P, R, D, N, K, M (W)* | 531,441            | 3                      |
| Ser peptides                  |                      |                         |                        |
| Is                            | G, A, P, R, E, N, Y, V| 262,144                 | 2                      |
| IIs                           | G, A, P, K, D, Q, H, F, I, M, W| 1,771,561         | 2*                     |

* Number of individual glycosylation experiments performed to obtain enhancement values plotted in Figs. 2–5. Peptides II and IIs were not used for reasons described under "Results."

* Asp and Lys could not be quantified in one sequence determination.

**The random peptide substrates utilized in these studies, GAGXXX(T/S)XXXAGAGK, where X = random residues (Table 1), address several technical issues. The first was the ability to readily separate random (glyco)peptide from free GalNAc that arises from the UDP-GalNAc hydrolyase activity found with most ppGalNAc transfers (29). GalNAc must be separated from glycopeptide because glycopeptide binding to the lectin affinity column is inhibited by free GalNAc. By using Sephadex G-10 gel filtration chromatography (reported exclusion limit of 700 daltons), we found that peptides on the order of 16 or more residues gave relatively clean separations from free GalNAc, whereas shorter length peptides (8–10 residues) were not adequately separated from GalNAc (data not shown). Because our previous modeling studies suggest that three residues N- and C-terminal to the glycosylation site (i.e. Thr or Ser) are sufficient for modeling mucin tandem repeat O-glycosylation (29, 30), three randomized residues were placed N- and C-terminal to the acceptor Thr or Ser. Symmetrical GAGA flanking sequences were added to increase peptide size while providing sequence identification markers, thereby allowing an assessment of (glyco)peptide purity and the extent of (glyco)peptide degradation. Moreover, the flanking GAGA sequences were chosen to represent a relatively extended peptide conformation and to further serve to eliminate end effects likely to be present with shorter peptide substrates. A C-terminal Lys residue was added as a unique functional group in the event that additional selection was required.

In initial studies with peptide I, the flanking randomized residues were limited to include those amino acid residues common to mucin glycoproteins while including representative residues for each of the different classes of amino acid side chains. Once the feasibility of the approach with random peptide I was demonstrated, random peptides II and IIs were synthesized with the amino acid composition given in Table 1. However, these peptides with their additional hydrophobic residues were partially insoluble and not good substrates of ppGalNAc T1 or T2. Therefore, a series of less hydrophobic random peptides (III–V) were synthesized, each containing the common residues Gly, Ala, Pro, and Arg along with 4–5 additional residues, of which only one was highly hydrophobic. These peptides displayed significantly improved solubility and were shown to be good substrates (see below). With peptides I and III–V, nearly all amino acid residue space can be examined except for Thr, Ser, Cys, and Trp (Trp is lost in the process of peptide isolation, and its enrichment values not determined).

Isolation of Random Glycopeptides on Sephadex G-10 and Lectin Affinity Chromatography—Representative Sephadex G-10 and mixed bed lectin column chromatograms are given in Fig. 1 demonstrating the isolation of random 3H-labeled glycopeptide I obtained after overnight incubation with ppGalNAc T1 and T2 (left and right panels). Fig. 1A demonstrates the initial separation of random (glyco)peptide from free GalNAc. Fig. 1B shows the separation of nonglycosylated random glycopeptide from 3H-labeled glycosylated random peptide eluted by GalNAc. Fig. 1C shows the final separation of the 3H-labeled random glycopeptide from GalNAc after release from the lectin column by GalNAc. Note that ppGalNAc T1 produces a greater amount of free GalNAc than ppGalNAc T2 (compare Fig. 1A, left and right panels) consistent with the greater UDP-GalNAc hydrolyase activity for ppGalNAc T1 (29). To improve glycopeptide binding, a mixed bed lectin column composed of multiple immobilized lectins with affinities to the peptide O-linked GalNAc residue (Tn antigen) or free GalNAc was utilized (see “Experimental Procedures” for composition). As shown in Fig. 1C, essentially all 3H-labeled glycopeptide binds to the lectin (i.e. no 3H label is associated with the peptide flow-through), and all of the 3H label is released from the column in the GalNAc wash. Based on total column integrations, none of the 3H radiolabel is irreversibly bound to the lectin. Identical gel filtration and lectin column behavior were observed for peptides III–V after glycosylation by ppGalNAc T1 or T2 (data not shown). (Note that in preliminary studies, random peptide I glycosylated by ppGalNAc T2 did not tightly bind to the lectin column in buffer containing 10% ethanol used as an antimicrobial agent. However, in the absence of ethanol, tight binding of glycopeptide to the lectin was restored.)

Edman Sequencing of Random (Glyco)Peptides and Determination of Enhancement Values—Edman amino acid sequencing of the isolated random (glyco)peptides was used to obtain the...
amino acid residue mole fractions at each position for peptide I with ppGalNAc T1 and T2, as shown in supplemental Fig. S1. In supplemental Fig. S1, the upper plots (A) represent data for the pre-lectin random (glyco)peptide sequence, whereas the lower plots (B) represent data for the lectin-bound Sephadex G-10-purified random glycopeptide. The plots clearly show the flanking GAGA sequences and the conversion of residue 8 from a Thr in the pre-lectin sequence to a glycosylated Thr residue in the post-lectin sequence. In addition changes in the composition of the random region, residues 5–7 and 9–11, are evident between the pre- and post-lectin column sequence determinations. To simply the analysis, residue-specific enhancement factors were obtained by dividing the post- and pre-lectin column mole fractions for each random residue. These enhancement factors will be taken as specific indicators of transferase preference, with values greater than 1 indicating an increased preference and values below 1 a decreased preference for the given residue (31).

A summary of the residue-specific enhancement factors obtained from using peptides I and III–V are given in Figs. 2–4 for both ppGalNAc T1 and T2. The results clearly show many differences and similarities between the site-specific enhancement factors for these transferases. Before these results are further discussed, an assessment of the reproducibility and validity of the method is given below.

Assessment of Reproducibility and Data Quality—We first examined to what extent the degree of peptide glycosylation could alter the obtained enrichment factors. As shown in the supplemental Fig. S2, no significant differences in enrichment factors were observed as a function of random peptide I glycosylation up to 10% conversion by ppGalNAc T1. We believe this tolerance to relatively high conversion, compared with random peptide studies of others (31), arises from our limiting of the randomized residues to only 8 or 9 individual residues, thus reducing the number of unique peptide sequences (Table 1) while also increasing the residue mole fraction at each random cycle. As we discuss below, the ability to obtain consistent enrichment values at high conversion may also suggest the transferase recognizes the residues of the peptide as independent sites with little cooperativity. Furthermore, supplemental Fig. S2 demonstrates that the enhancement factors are highly reproducible between different experiments using the same random peptide substrate. In supplemental Fig. S3, two different constructs of ppGalNAc T2 (expressed from baculovirus (36, 37) and yeast (38)) were compared against two different synthetic batches of random peptide I containing slightly different initial random residue compositions. Except for some variability in the Pro enhancement factors, neither the enzyme source nor peptide synthesis appears to have significant effects on the enrichment factors observed. Next we compared the ppGalNAc T1 and T2 enrichment factors obtained from Ser peptide I to the factors obtained from Thr peptide I. As shown in supplemental Figs. S4 and S5, few differences were found between the Ser and Thr random peptide enrichment factors indicating that the transferases recognize neighboring residues identically, regardless of the nature of the acceptor residue, and despite the fact that Ser residues are more slowly glycosylated than Thr residues (29, 30). Therefore, Ser analogues of peptides III–V were not obtained for study. Next, the consistency of the value of the
enancements obtained from the different peptide substrates (i.e. peptides I and III–V) was assessed. As shown in supplemental Figs. S6 and S7, for both ppGalNAc T1 and T2 the enrichment factors for the common residues found in all four random peptides (i.e. Gly, Ala, Pro, and Arg, top panels) or found in two of the random peptides (i.e. Glu, Asp, Gln, Asn, and Lys, center panels) are nearly identical and do not signif-

cantly vary with peptide substrate. Therefore, the obtained enhancement values do not appear to be sen-
tive to the overall composition of the peptide. Taken together we con-
clude from the results pre-
sented in supplemental Figs. S2–S7 that the obtained enrich-
ment factors are highly reproduc-
ible, do not significant vary with
the nature of the random peptide
substrate and extent of glycosyla-
tion (up to ~10%), and are essen-
tially identical for both Thr and Ser acceptor peptides.

Hydrophobic Residue Enhance-
ment Factors—Enhancement values
for the hydrophobic residues Gly,
Ala, Pro, Val, Ile, Leu, Met, Phe, and
Tyr are plotted relative to the site of
glycosylation (i.e. −3, −2, −1, +1,
+2, and +3) for both ppGalNAc T1
and T2 in Fig. 2, A and B. For further
comparison, the data are separately
plotted by residue type in Fig. 3. An
inspection of the plots in Figs. 2 and
3 reveals that the enhancement
patterns for both transferases are
nearly identical at each of the
terminal positions +1, +2, and
+3. At the +1 position both trans-
ferases show large, ~2-fold, Pro
enhancements, somewhat smaller
Val enhancements, and moderate
inhibitions with Phe, Tyr, Leu, and
Gly. At the +2 position, moderate
enhancements are observed for
Gly and Ala, and none to weak
inhibitions for Ile, Leu, Phe, and
Tyr. Both ppGalNAc T1 and T2, at
the +3 position, exhibit large
~3-fold enhancements for Pro and
eutral to slightly inhibitory
enhancement values for the remain-
ing residues except for Tyr. Interest-
ingly, Tyr at the +3 position shows a
large ~2-fold enhancement for
ppGalNAc T1 and slight inhibition
for ppGalNAc T2 Tyr. This makes
Tyr at the +3 position the only
hydrophobic residue, C-terminal
of the glycosylation site, capable of discriminating between
the two transferases.

In contrast, the N-terminal residues to the site of glycosyla-
tion, positions −3 to −1, reveal significant differences between
transferases (Fig. 2, A and B); only Ala, Leu, and Met appear to
have identical preferences for both transferase at all three
N-terminal positions (Fig. 3). At the −3 position, for ppGalNAc
FIGURE 3. Comparison of the ppGalNAc T1 and T2 positional enhancement factors for individual hydrophobic residues (Gly, Ala, Pro, Val, Ile, Leu, Met, Phe, and Tyr). The dark and light bars represent ppGalNAc T1 and ppGalNAc T2 enhancement values, respectively.
ppGalNAc Transferase Peptide Substrate Preferences

T1, all of the hydrophobic residue enhancements cluster around 1, indicating no strong hydrophobic residue preference at this site, whereas ppGalNAc T2 displays nearly 2-fold enhancements for Pro and the β-branching amino acid residues Val and Ile, while showing slight inhibitory effects for Gly and Tyr. At the −2 position, except for a nearly 2-fold enhancement for Gly for ppGalNAc T2, both transferases display similar near neutral enhancement patterns, with Ile being the least favored residue. The largest differences in hydrophobic residue preferences between the two transferases are observed at the −1 position, the residue directly N-terminal to the glycosylation site. With ppGalNAc T1, Pro, Val, Ala, Ile, and Phe show moderate, ~1.5- to ~2-fold, enhancements, whereas Met and Leu appear moderate to highly inhibitory at the −1 position. In contrast, with ppGalNAc T2, Pro gives a very large ~4.5-fold enhancement (the largest enhancement observed); Gly and Ala give relatively neutral effects, whereas the remaining hydrophobic residues are moderately to highly inhibitory. The only common features between transferases at this site are the values of the enhancements at Gly, Ala, Leu, and perhaps Met. Therefore, the hydrophobic residues at the −1 position most uniquely differentiate the preferences of the two transferases, whereas additional discrimination between transferases is possible at the −3, −2, and +3 sites.

Further analysis of the individual residue plots in Fig. 3 is warranted. From the plots it is clear that the Ala residue shows no strong positional or transferase-specific preferences and therefore could perhaps be considered the only completely neutral residue. Similarly, the Pro residue is the only other hydrophobic residue that displays enhancements of ~1 or more for all positions for both transferases. Although the N-terminal enhancements for Pro differ between transferases, this residue shows an approximate high-low-high pattern of preferences for both the C- and N-terminal positions. Interestingly, Gly residues are moderately inhibitory at the +1 site for both transferases and show the inverse of Pro, having an approximate alternating low-high-low pattern at both the N- and C-terminal positions. The more hydrophobic β-branched residues, Val and Ile, display similar enhancement patterns but differ from the other hydrophobic residues Leu, Phe, Tyr, and Met. The Val and Ile C-terminal enhancements appear to decline from moderately enhancing at the +1 position to moderately inhibitory at the +3 position for both transferases. At the N-terminal, Val and Ile are favored by ppGalNAc T1 at the −1 position, whereas these residues are favored by ppGalNAc T2 at the −3 position. In contrast to Val and Ile, the enhancements for Leu are largely identical for both transferases at all N- and C-terminal positions, with Leu having neutral effects at the −3 and −2 positions and moderately inhibitory effects at the other positions. Phe, Tyr, and Met also show similar enhancement patterns as Leu, except that for ppGalNAc T1, Phe and Tyr are not inhibitory at the −1 position and Tyr shows an enhancement not observed for Phe at the +3 position. These results clearly demonstrate that these transferases exhibit both similar and unique hydrophobic residue preferences that must reflect differences in the amino acid sequence and structure of the two transferases.

Hydrophilic Residue Enhancement Factors—Enhancement values for the hydrophilic residues, Glu, Asp, Gln, Asn, Arg, Lys, and His, are plotted together in Fig. 2, C and D, and individually in Fig. 4 for both ppGalNAc T1 and T2. As observed for the hydrophobic residues, the C-terminal hydrophilic residues display nearly identical patterns of enhancement for both ppGalNAc T1 and T2. The most significant differences between transferases are again observed at the −1 position, although the differences at this position are not as unique as observed for the hydrophobic residues (Fig. 2, A and B). Interestingly, no hydrophilic residues are favored at this position for either transferase.

A comparison of the different classes of hydrophilic residues reveals several interesting features. First, a comparison of the acidic residues, Glu and Asp (red bars in Fig. 2, C and B, top panels in Fig. 4), reveals that these residues behave similarly for both transferases and that Glu and Asp could be easily exchanged, except for the −1 and +3 positions, where Glu is more favored than Asp for both transferases. Except for this Glu/Asp difference, the acidic residues are moderately favored at the +1 position, are highly unfavorable at −1, and approximately neutral at the other positions (Fig. 4).

The uncharged Gln and Asn residues also show similarities with each other and between transferases, although in contrast with the acidic residues the largest differences between Gln and Asn are observed at the +1 and +2 positions for both transferases, where Gln is mostly enhancing and Asn inhibitory (Fig. 2, C and D, green bars). It is interesting to note that the Gln and Glu enhancements appear to be comparable for both transferases at nearly all positions, whereas for the Asn-Asp pair, the Asn enhancements are always less, particularly for the C-terminal positions, than Asp for both transferases (Fig. 4).

The basic residues, Arg, Lys, and His, exhibit as a group generally similar behavior being predominantly inhibitory for both transferases (Fig. 2, C and D, blue bars, and Fig. 4, bottom panels). Lys is typically the most inhibitory residue, and His and Arg are typically the least inhibitory. The largest difference between transferases is found in the N-terminal His enhancements where His is nearly neutral in ppGalNAc T1 and inhibitory for ppGalNAc T2.

In summary the hydrophilic residues are largely neutral to inhibitory for both transferases, with only weak enhancements observed at the +1 position and perhaps at the −3 and +3 positions. For both transferases all hydrophilic residues at position −1 are inhibitory, although for ppGalNAc T2 the neutral and basic residues are more inhibitory than the acidic residues compared with ppGalNAc T1. The most significant differences between the enhancements of the acidic and basic residues are found at the +1 site, for both transferases, where the acidic residues are slightly enhancing to neutral and the basic residues moderately inhibitory. Generally, however, the hydrophilic residues show less significant differences between transferases compared with the hydrophobic residues, and hence the hydrophilic residues will less likely be major discriminants between the two transferases (Fig. 4).
Positional Ranking of Enhancements and Estimation of Optimal Peptide Substrates—In Fig. 5 we have plotted the residue enhancements in descending order at each random residue position for both ppGalNAc T1 and T2. These plots are useful for ranking positional residue preferences and for deriving optimal peptide substrate sequences for each transferase. As shown in Fig. 5 and summarized in Table 2, both transferases have the same optimal residues at position −1 and at all three C-terminal positions (i.e. Pro, Pro, Gly, and Pro, respectively). Only at the −3 and −2 positions do the two transferases differ (i.e. Phe/Asp and Phe/Ala for ppGalNAc T1 and Pro/Ile and Gly for ppGalNAc T2, respectively), although for ppGalNAc T1 the optimal residues are not overwhelmingly better than the others (see Fig. 5). If we take the products of the enhancement values of the optimal peptide sequences as a rough measure of the relative rates of glycosylation of each peptide (31), then the best ppGalNAc T1 and T2 substrate peptides might be predicted to be on the order of ~50- and ~200-fold better substrates compared with a hypothetical peptide composed of enhancement neutral residues (see Table 2, Net Enhancement columns). For comparison, a nearly neutral all Ala peptide could be expected to have relative enhancements of 2.5 and 1.8 for ppGalNAc T1 and T2, respectively (see Table 2).

To test these predictions both the ppGalNAc T1 and T2 optimal peptides and the all Ala peptide were synthesized, and their kinetic properties against ppGalNAc T1 and T2 characterized are given in Table 2. As expected, each transferase was most active against its respective optimal peptide on the basis of their $k_{cat}/K_m$ values. Furthermore, ppGalNAc T2 showed the predicted (~40%) reduced activity against the optimal T1 peptide, although ppGalNAc T1 showed ~5-fold less activity than predicted (13 versus 62%) against the optimal T2 peptide. Interestingly, for both transferases their optimal peptides had the lowest $K_m$ values, although they did not necessarily have the highest $k_{cat}$ values. Also note that ppGalNAc T2 appears to be nearly ~10-fold more active against its optimal peptide substrate than ppGalNAc T1 is against its optimal substrate. Finally, the all Ala peptide, which was predicted to have only a few percent relative activity against both transferases, was found to be inactive under the conditions studied with both ppGalNAc T1 and T2. Generally, the obtained kinetic values for the optimal peptides are in the range of values previously reported for these transferases (36, 41, 42).

To derive peptide substrate sequences that in principle would optimally discriminate between the activities of the two transferases, we utilized the ratios of the ppGalNAc T1 to ppGalNAc T2 enhancement values (listed in supplemental Table S1) and the plots in Figs. 3–5. To maintain reasonable potential activities, residues whose specific transferase enhancements were significantly lower than 1 in the numerator were excluded. The transferase discriminating sequences

![Comparison of the ppGalNAc T1 and T2 positional enhancement factors for individual hydrophilic residues (Glu, Asp, Gln, Asn, Arg, Lys, and His). The dark and light bars represent ppGalNAc T1 and ppGalNAc T2 enhancement values, respectively.](image-url)
obtained in this manner are given in Table 2, where the most discriminating positions (ratios of $\sim 2$ or more) have been labeled for each transferase with an asterisk. To summarize, the $-3$ position ppGalNAc T1 largely prefers the charged residues His and Glu by factors of $\sim 1.5$ to $2$, whereas ppGalNAc T2 has an $\sim 2$-fold preference for the hydrophobic residues Pro and Ile. At the $-2$ position Val and Gly are favored by only factors of $\sim 1.5$ for ppGalNAc T1 and T2, respectively. At the $-1$ position ppGalNAc T1 is selected over ppGalNAc T2 by the hydrophobic residues Phe, Ile, and Val by factors between $3$ and $\sim 7$, whereas for ppGalNAc T2 Pro is $\sim 2$-fold preferred over ppGalNAc T1, the only residue that ppGalNAc T2 prefers over ppGalNAc T1 at this position. The differences between transferases at the $+1$ position are relatively small with ppGalNAc T1 favoring Glu and Ile by factors less than 1.5 and with Ala very slightly favored by ppGalNAc T2. Likewise, at the $+2$ position ppGalNAc T1 favors Arg by a factor of $\sim 1.5$, and ppGalNAc T2 slightly favors Ile and Glu by factors less than 1.5. At the $+3$ site, Tyr discriminates ppGalNAc T1 from ppGalNAc T2 by a factor over 2, whereas Gly and Asp are only weakly favored by ppGalNAc T2 by factors of less than 1.5. The net product of the ppGalNAc T1/T2 (or ppGalNAc T2/T1) enhancement ratios, for the proposed transferase-selective peptides, would therefore be expected to reflect the net selectivity of that peptide for the two transferases; these values are given in Table 2 (Selectivity column). From Table 2 the discriminating ppGalNAc T1 peptide would be expected to have a net rate enhancement of $\sim 4$ and would be expected to be $\sim 100$-fold more selective for ppGalNAc T1 than for ppGalNAc T2 (assuming equivalent intrinsic rates for both transferases). For ppGalNAc T2, its selective peptide would be expected to have a nearly 20-fold net rate enhancement and would be expected to be $\sim 16$-fold (1:0.06) more selective for ppGalNAc T2 than for ppGalNAc T1.

The above selective peptide predictions were also tested experimentally with both ppGalNAc T1 and T2 as shown in Table 2. Again the predictions were relatively correct, with each transferase showing activity only against its selective peptide indicating high selectivity for each transferase. The enhancement value predictions suggest these peptides should be about 10%
less active than the respective best peptide substrates. This is indeed the case for the ppGalNAc T1 peptide, whereas the selective ppGalNAc T2 peptide is found to be about 20-fold less active than predicted. It is also interesting that the most selective ppGalNAc T1 peptide has the lowest $K_m$ value for ppGalNAc T1 among the peptides in Table 2, whereas the most selective ppGalNAc T2 peptide has the highest $K_m$ value for ppGalNAc T2.

**DISCUSSION**

By utilizing a series of oriented semi-random peptide substrates, we have performed the most comprehensive analysis to date of the peptide substrate preferences of ppGalNAc T1 and T2. Key to our success was the design of the random peptide substrates, the use of mixed bed lectin affinity chromatography, and the use of Edman amino acid sequencing. The peptide substrates were optimally designed to be fully soluble, to be large enough to be readily separated from free GalNAc, and to eliminate end effects and contained flanking control sequences. The obtained residue enrichment factors are highly reproducible, do not vary significantly with random peptide composition or extent of glycosylation, and are essentially identical for both Thr and Ser acceptor peptides.

The relatively high reproducibility of the enrichment factors obtained from peptides with different amino acid compositions suggests that ppGalNAc T1 and T2 may recognize each of the three N- and C-terminal flanking positions independent of the others. With this assumption, the product of the individual preferences can then be taken as comparative measures of the relative rates of glycosylation between different peptide sequences as was done for the protein kinase random peptide libraries (31). In this manner we have proposed the optimal and most selective peptide substrate sequences for both transferases (Table 2) and have confirmed their high activity and selectivity experimentally (Table 2). In most cases we found the relative rate predictions to be reflected in the experimental $k_{cat}/K_m$ values (Table 2) rather than any one single kinetic value. This is despite the fact that in the random peptide substrates the concentration of any given peptide sequence would be far below the $K_m$ values for the enzymes (31, 32). On this basis, the obtained enrichment values should be suitable for incorporation into an expanded version of our previously reported kinetic model (29, 30, 43) for characterizing the ppGalNAc T1 and T2 glycosylation patterns of the porcine and canine salivary gland mucin tandem repeats (PSM and CSM).

A comparison of our results with prior report of ppGalNAc T1 and T2 peptide specificity is instructive for both further validating our results and for identifying ppGalNAc T1 or T2 motifs that may be found in the existing O-glycosylation data bases. Prior to this work the most systematic peptide substrate studies of ppGalNAc T1 were of O’Connell et al. (44) and Yoshida et al. (41). The results of our random peptide studies are largely in agreement with these previous studies and our rankings very similar to those reported by Yoshida et al. (41). For example, similar to our findings, these workers observed that at the −1 site, hydrophobic residues, except for Leu, were favored and charged residues were disfavored. Also similar to our observations, both workers found Pro and Ala to be most favored, and the acidic, neutral, and hydrophobic residues to be least favored at the +2 site. Our random peptide studies also

---

**Note** that in this work we have limited our study to the +1−3 residues adjacent to the site of glycosylation. We do not presently know the effects of the flanking GAGA/AGAGK residues, whether residues more distant from the site of glycosylation may be significantly recognized by the transferases, nor how these may affect the reported enrichment values. This uncertainty is presently being addressed by new studies on random peptide substrates containing longer regions of randomized sequences.

---

**TABLE 2**

**Optimal and selective ppGalNAc transferase substrate sequences (GAGAXXXTXXXAGAXGK), prediction and experiment**

| Position, X Residue | Predicted Rates | Experiment Rates |
|---------------------|-----------------|-----------------|
|                     | Enhancement $^a$ Selectivity $^b$ | $k_{cat}$/mM $^c$ | $k_{cat}$/mM $^c$ |
| Optimal Sequence    | T1 | T2 | T1/T2 | T1 | T2 | T1/T2 |
| ppGalNAc T1         | F/D | F/A | P/V | T | P | G/A | P |
|                     | 45(100) | 87(42) | 0.52 | 2.61±0.8 | 11.3±1.0 | 4.3(100) | 0.19±0.03 | 3.25±0.16 | 17(44) |
| ppGalNAc T2         | P/I | G | T | P | P | G | P |
|                     | 28(62) | 209(100) | 0.13 | 5.1±0.4 | 2.9±0.6 | 0.53(13) | 0.05±0.008 | 1.95±0.08 | 39(100) |

Discriminating Sequence

| Sequence | T1 vs T2 | T2 vs T1 |
|----------|----------|----------|
|          | H/E | V | F/I/V* | T | E/I | R | Y* |
| T1 vs T2 | P/I | G* | P* | T | A | I/Q | G/D |
|          | 4.3(10) | 0.045(0) | 96 | 1.5±0.25 | 0.63±0.05 | 0.4(10) |
|          | 2.8(10) | 19(9) | 0.06 | n.s. | n.s. | n.s. |
| Ala Peptide | A | A | A | T | A | A |
|          | 2.56(6) | 1.8(1) | 1.4 | n.d | n.d | n.d |

$^a$ For the optimal sequence, peptide sequence(s) were composed of residues with the highest enhancement values. For the discriminating sequence, peptide sequence(s) were composed of residues with the highest $T1/T2$ or $T2/T1$ enhancement ratios as described under "Results." $^b$ Indicates major discriminating positions.

Net enhancement value represents the product of the individual positional enhancement values for ppGalNAc T1 or T2, respectively. Values were calculated for the sequence represented by the first residue given at each position (in boldface). Values in parentheses were normalized to 100 for best sequence rate. Values are in boldface for emphasis.

$^c$ Net apparent selectivity values represent the product of the ratios of the residue-specific transferase enhancement values ppGalNAc T1 over ppGalNAc T2. Values were calculated for the sequence represented by the first residue listed at each position (in boldface). Note that the relative selectivity ratios do not take into account any intrinsic differences in the activities of the two transferases.

Kinetic values were obtained as described under "Experimental Procedures" for the human ppGalNAc T1 and T2 against the peptide sequences in boldface. Note that all substrate peptides contain the GAGA- and -AGAGK flanking sequences. $k_{cat}/K_m$ values were normalized to 100 for best sequence value. Units on $k_{cat}$ are (mols of product)/(mole of transferase) s$^{-1}$. Values in boldface are for emphasis. n.s. indicates no significant activity, activity too low for kinetic values to be determined; n.d. indicates no activity detected.

---

5 T. A. Gerken and O. Jamison, manuscript in preparation.
ppGalNAc Transferase Peptide Substrate Preferences

revealed the rate-enhancing effects of Pro at the −1, +1, and +3 sites as observed previously (41, 44). Finally, the (AA)AT-PAP sequence proposed by Yoshida et al. (41), as the shortest motif for efficient glycosylation, is consistent with our random peptide derived optimal ppGalNAc T1 peptide sequence, -(F/D)/(F/A)/(V/P)/(T/G)/A/- (Table 2).

There are fewer studies on the peptide substrate specificity of ppGalNAc T2, although recently we reported a ppGalNAc T2 motif based on our ppGalNAc T2 glycosylation and modeling studies of the PSM and CSM tandem repeats and the human IgA1 hinge domain (30). In these studies we detected moderate Pro enhancements at the −3 and +3 positions and a very large Pro enhancement at the −1 position. Very similar enhancements are detected in our random peptide substrate studies with the addition that Pro is also enhanced at the +1 position (Fig. 3). As we noted previously (30) most good substrates for ppGalNAc T2 contain Pro-Ser or Pro-Thr sequences; thus the random peptide substrates have uniquely confirmed the large −1 Pro preference of ppGalNAc T2. Interestingly, the random peptide data indicate few other residues are tolerated by ppGalNAc T2 at this site. Based on the ppGalNAc T2 enrichment factors, we found the optimal ppGalNAc T2 substrate peptide sequence to be -(P/I)/GPTPGP- (Table 2).

As discussed above and under “Results,” there are considerable similarities with the preferences of ppGalNAc T1 and T2, particularly C-terminal of the acceptor Thr or Ser; hence, the optimal substrates for both transferases share the common PTPGP motif. However, by incorporating residues that have large differences in enrichment factors between ppGalNAc T1 and T2, we were able to derive peptide substrate sequences experimentally capable of discriminating between the two transferases (Table 2). The most discriminating positions (showing ∼2-fold or larger differences) appear to be at −3 where Pro, Ile, and Val are favored by ppGalNAc T2, at −1 where Phe, Ile, and Val are favored by ppGalNAc T1 and Pro favored by ppGalNAc T2, and at +3 where Tyr is favored by ppGalNAc T1. Therefore, even though there are many common features between transferases, there are additional specific differences that can be used to discriminate the activity of one from the other. It is worth noting that unlike previous reports of ppGalNAc T2 isoform-selective peptide substrates (2, 4, 42), our selective substrates were developed a priori entirely from the predictions of the random peptide-derived preferences. Interestingly, our ppGalNAc T1- and T2-selective peptides do not share any common sequence features (except for the −1 Pro for the ppGalNAc T2 substrate) with previously reported ppGalNAc T1- and T2-selective peptide substrates (42).

The random peptide-derived enrichment factors are also helpful for understanding the order of glycosylation of multiple acceptor substrates. For example the 20-residue tandem repeat of human MUC1, 1PAPGSTTAPPHTDPSTPAPDT20, is glycosylated by ppGalNAc T1 first at Thr-14 (underlined) followed by Thr-6 (underlined) (42, 45, 46). On the basis of the products of the ppGalNAc T1 random peptide enrichment values Thr-14 indeed should be clearly favored over Thr-6, whereas Thr-19 (underlined) would be least favored (products of 11.3, 3.7, and 0.4 respectively, assuming enrichment values of 1 for neighbor-

Ser or Thr residues). In contrast, ppGalNAc T2 is found to glycosylate Thr-6 first, followed by Thr-14 (42, 45), again consistent with the random peptide ppGalNAc T2 enrichment value products of Thr-6, Thr-14, and Thr-19 (products of 15.2, 2.4, and 0.6, respectively). Similarly, the human IgA1 hinge domain peptide, 1VPSTPTPPSTPSTPSK20, is most rapidly glycosylated by ppGalNAc T2 at Thr-7 and -15 (underlined) both having neighboring enhancing Pro residues at the −1, +1, and +3 positions (30, 47). Ser-11 (underlined) is the next most rapidly glycosylated residue having neighboring enhancing Pro residues at −1 and +3.

A comparison of our random peptide enrichment factors with the published data base analysis of protein O-glycosylation sites is also instructive. The detection of any ppGalNAc T1 or T2 transferase-specific motifs in the data base would suggest that these motifs indeed may be used by nature to target sites for specific O-glycosylation. Indeed, in the work of Hansen et al. (48) there appears to be an indication of both the ppGalNAc T1 and T2 motifs in their analysis of the single O-glycosylated Thr residues (48). In their analysis Pro dominates at the −1, +1, and +3 sites, consistent with both the ppGalNAc T1 and T2 random peptide enrichment factors (Fig. 3). The next most common residues in the data base are Val and Glu at the −1 and +1 sites, respectively, residues that would be favorable for ppGalNAc T1 and relatively unfavorable for ppGalNAc T2 (Figs. 3 and 4). The analysis also shows Leu to be very low at the −1 site, indeed as is observed by us for both transferases (Fig. 3) and in the work of Yoshida et al. (41) for ppGalNAc T1. The analysis of Hansen et al. (48) also shows Val and Leu to be the most prevalent residues at the −3 site; this coupled with a very large −1 Pro probability could also suggest the presence of a residual ppGalNAc T2 motif in the data base. Consistent with the relatively ubiquitous expression of ppGalNAc T1 and T2 (15, 49), we conclude that a number of O-glycosylation sites in the data base may have evolved to be specifically glycosylated by ppGalNAc T1 or T2. However, because the specificities of the remaining members of the ppGalNAc transferase family are presently unknown, we cannot rule out that other transferase isoforms may also have been targeted to these motifs.

Recently the x-ray crystal structure of ppGalNAc T2 with bound UDP and substrate peptide EA2 (1PTTDSTTPAPsT13) was reported (38). In the complex EA2, residues Ser-5 through Lys-13 are bound in a shallow groove, with Thr-7 (underlined) the preferred (and predicted by our work) in vitro glycosylated residue, appropriately positioned as the acceptor residue adjacent the terminal PO4 of the bound UDP (38) (see supplemental Fig. S8). An inspection of the peptide-binding site cleft reveals features that are consistent with the obtained random peptide enrichment factors and that may help rationalize the differences and similarities between the two transferases. For example, the ppGalNAc T2 surface and cleft residues responsible for binding the C-terminal PAP sequence of the EA2 peptide are well defined and nearly completely conserved in ppGalNAc T1 (see Table 3 and supplemental Fig. S8), thus accounting for their nearly identical C-terminal enrichment patterns and rankings (Fig. 5). This region of the cleft clearly shows two pockets for binding the +1 and +3 Pro residues with the
+2 Ala side chain methyl pointing outward, interacting with the surface of the side chain ring of Phe-361 in ppGalNAc T2 (see supplemental Fig. S8). The small size and surface-exposed nature of the +2 binding site is in agreement with the random peptide data that indicate that residues with the shortest side chains, i.e. Gly and Ala, are most favorable at this position for both transferases. Of the ppGalNAc T2 residues interacting with the PAP portion of the substrate peptide, only Leu-270 and Ala-266 differ, in ppGalNAc T1 being Thr-255 and Gly-251 (Table 3). Because both of these residues form a portion of the +3 peptide-binding site (38), it is conceivable that the alteration of one or both of these residues may account for the elevated preference for Tyr with ppGalNAc T1 at the +3 position.

Less can be said about the interactions of the N-terminal peptide substrate residues (positions −3 to −1) with the transferase, because in the ppGalNAc T2 crystal structure the EA2 residues N-terminal of the −2 position (Ser-5) are not bound to what appears to be the likely peptide binding cleft (38) (supplemental Fig. S8). In addition, we do not have transferase preferences for Ser and Thr residues; therefore, it is unknown whether Ser-5 and Thr-6 (−2 and −1 positions) will be favored at these positions. Nevertheless, it is clear from supplemental Fig. S8 and summarized in Table 3 that the residues composing the N-terminal portion of the putative and partially bound peptide substrate binding cleft are only partially conserved between ppGalNAc T1 and T2, where at least five residues differ. This lack of conservation may account for the different preferences observed between transferases for the N-terminal peptide substrate residues. Molecular modeling and additional crystal structure determinations are in progress to further characterize the N-terminal peptide-binding site.

Nearly all ppGalNAc transferase isoforms contain a C-terminal lectin domain composed of three contiguous ricin-like sequences that have been variously shown to interact with free GalNAc or GalNAc-glycosylated peptides (5, 8, 25–27, 50, 51). On the basis of the crystal structures of ppGalNAc T1 and T2, it is clear that the ricin domains are tethered to the catalytic domain and do not necessarily directly partake in binding of the peptide substrate residues directly neighboring the site of glycosylation (38, 52). We therefore conclude that the residue enhancement values obtained from the short random peptide substrates are not influenced by the lectin domain and will reflect the intrinsic peptide substrate specificity of the transferases. This is further supported by our observations that the presence of 0.5 mM free GalNAc, which has been shown to alter the glycopeptide specificity of a number of ppGalNAc transferases, does not alter the enrichment values obtained for peptide I with either ppGalNAc T1 or T2 (data not shown).

In this work we have successfully demonstrated the use of oriented random peptide substrates for the characterization of the peptide substrate specificities of the large family of ppGalNAc transferases that initiate the first step in O-glycan biosynthesis in a wide range of organisms. Importantly, our results are consistent with previous experimental studies, data base analysis of O-glycosylation sites, and with the x-ray crystal structures of the two transferases. We have furthermore utilized the results to design peptide sequences shown to be the optimal or most selective substrates for each transferase. The ability to obtain peptide substrate specificity for each ppGalNAc transferase will be invaluable for the identification of isoform-specific substrates and inhibitors, the identification of specific O-glycosylation sites, and for the comparison of ppGalNAc transferase orthologues across species.

Acknowledgments—We acknowledge the technical assistance of Jason Marick and the generous gift of transferases by Drs. Ake Elhammer (Kalumazoo, MI) and Lawrence Tabak (Office of Director, NIDCR, National Institutes of Health, Bethesda). We acknowledge the Facility for Biotechnology Resources, Center for Biologics Evaluation and Research at the National Institutes of Health for providing services in peptide synthesis.

REFERENCES

1. Clausen, H., and Bennett, E. P. (1996) Glycobiology 6, 635–646
2. Bennett, E. P., Hassan, H., Mandel, U., Mirgorodskaya, E., Roepstorff, P., Burchell, J., Taylor-Papadimitriou, J., Hollingsworth, M. A., Merkx, G., van Kessel, A. G., Eiberg, H., Steffensen, R., and Clausen, H. (1998) J. Biol. Chem. 273, 30472–30481
3. Ten Hagen, K. G., Hagen, F. K., Balys, M. M., Beres, T. M., Van Wyckhuysse, B., and Tabak, L. A. (1998) J. Biol. Chem. 273, 27749–27754
4. Bennett, E. P., Hassan, H., Mandel, U., Hollingsworth, M. A., Akisawa, N., Ikenatsu, Y., Merkx, G., van Kessel, A. G., Olofsson, S., and Clausen, H. (1999) J. Biol. Chem. 274, 25362–25370
