

# Structure-Function Analysis of the UDP-*N*-acetyl-D-galactosamine: Polypeptide *N*-acetylgalactosaminyltransferase

ESSENTIAL RESIDUES LIE IN A PREDICTED ACTIVE SITE CLEFT RESEMBLING A LACTOSE REPRESSOR FOLD\*

(Received for publication, October 22, 1998, and in revised form, December 5, 1998)

Fred K. Hagen<sup>‡</sup>, Bart Hazes<sup>§</sup>, Roberto Raffo<sup>‡</sup>, Deborah deSa<sup>‡</sup>, and Lawrence A. Tabak<sup>‡¶</sup>

From the <sup>‡</sup>Center for Oral Biology, Rochester Institute of Biomedical Sciences, University of Rochester, Rochester, New York 14642 and the <sup>§</sup>Department of Medical Microbiology and Immunology, University of Alberta, 1-41 Medical Sciences Building, Edmonton, Alberta T6G 2H7, Canada

Mucin-type *O*-glycosylation is initiated by a family of UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferases (ppGaNTases). Based on sequence relationships with divergent proteins, the ppGaNTases can be subdivided into three putative domains: each putative domain contains a characteristic sequence motif. The 112-amino acid glycosyltransferase 1 (GT1) motif represents the first half of the catalytic unit and contains a short aspartate-any residue-histidine (DXH) or aspartate-any residue-aspartate (DXD)-like sequence. Secondary structure predictions and structural threading suggest that the GT1 motif forms a 5-stranded parallel  $\beta$ -sheet flanked by 4  $\alpha$ -helices, which resembles the first domain of the lactose repressor. Four invariant carboxylates and a histidine residue are predicted to lie at the C-terminal end of three  $\beta$ -strands and line the active site cleft. Site-directed mutagenesis of murine ppGaNTase-T1 reveals that conservative mutations at these 5 positions result in products with no detectable enzyme activity (D156Q, D209N, and H211D) or <1% activity (E127Q and E213Q). The second half of the catalytic unit contains a DXXXXXWGGENXE motif (positions 310–322) which is also found in  $\beta$ 1,4-galactosyltransferases (termed the Gal/GalNAc-T motif). Mutants of carboxylates within this motif express either no detectable activity, 1% or 2% activity (E319Q, E322Q, and D310N, respectively). Mutagenesis of highly conserved (but not invariant) carboxylates produces only modest alterations in enzyme activity. Mutations in the C-terminal 128-amino acid ricin-like lectin motif do not alter the enzyme's catalytic properties.

UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferases (EC 2.4.1.41; ppGaNTases)<sup>1</sup> comprise a large family of metal-dependent enzymes that initiate mucin-type *O*-glycosylation. During catalysis, GalNAc is transferred from UDP-GalNAc to selected serine and threonine residues of proteins destined for the extracellular environment. The family of ppGaNTases have been conserved

during evolution (1). Since *O*-glycosylation proceeds in a stepwise manner, the expression and specificity of ppGaNTase isozymes represent key regulatory factors in defining the repertoire of *O*-glycans expressed by a cell. *In vitro* studies have shown that some ppGaNTases (ppGaNTase-T1, -T2, and -T3) glycosylate a broad spectrum of peptides (2, 3), whereas others exhibit more restricted (ppGaNTase-T4 and -T5) or unique (ppGaNTase-T3) substrate preferences (2, 4, 5). The structural basis for this enzyme specificity is unknown, as no experimental structure is currently available for ppGaNTases or for any other mammalian glycosyltransferase.

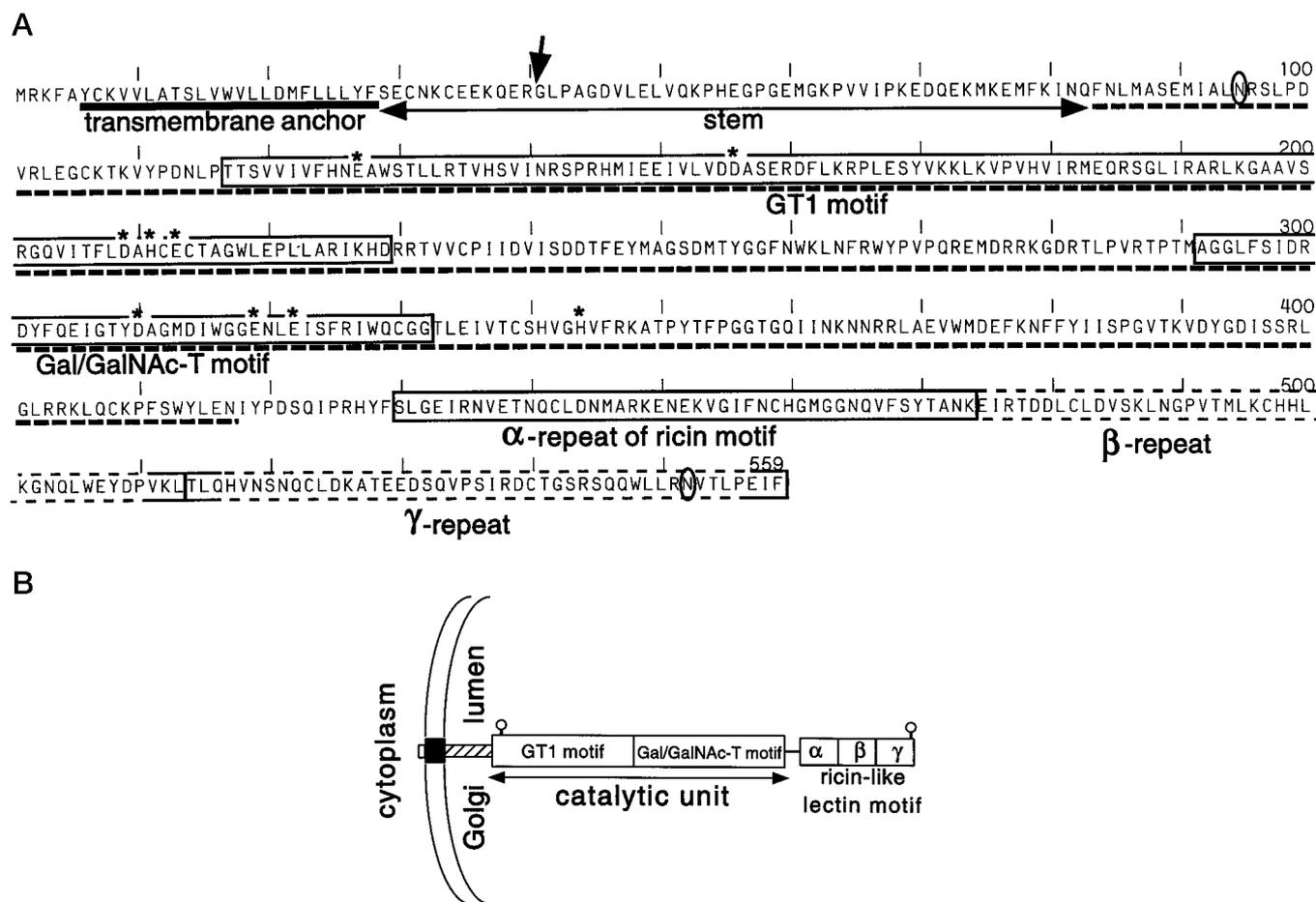
Recently, several groups have identified an aspartate-any residue-aspartate (DXD) (or aspartate-any residue-histidine (DXH)) sequence motif that is common to many glycosyltransferases (6–9). All known ppGaNTases contain such a DXH or DXD-like motif. Based on fold recognition, the DXD motif of the  $\alpha$ 1,3-fucosyltransferase is predicted to correspond to aspartate 100 in the active site of the bacteriophage T4 DNA  $\beta$ -glucosyltransferase (7, 10). Mutagenesis of aspartate residues in other DXD-containing enzymes (mannosyltransferase, glucosyltransferase, and chitin synthase 2) demonstrates that they are essential to catalytic activity (7, 9, 11). Such carboxylic acid residues may be involved in different aspects of the catalytic process. First, glycosyltransferases (including ppGaNTases) that retain the anomeric configuration of the sugar-nucleotide bond are thought to work via a double displacement mechanism, which would require two carboxylic acid side chains (12). Second, carboxylates could be involved in binding of substrate. Third, the activity of ppGaNTases requires the binding of metal ions that can adopt an octahedral geometry; ppGaNTase-T1 has a preference for Mn<sup>2+</sup> (13–15). Coordination of Mn<sup>2+</sup> typically involves multiple aspartate residues although histidine and glutamate may also be involved (16). However, the position of key aspartate, glutamate, and histidine residues within a glycosyltransferase model is not known.

In the current study, threading analysis suggests that the central 340-amino acid region of mammalian and *Caenorhabditis elegans* ppGaNTases share a common structural fold that consists of two domains, each of which contains a parallel  $\beta$ -sheet, flanked by  $\alpha$ -helices. This fold most closely resembles the lac repressor protein, a member of the periplasmic binding protein family fold, which itself is structurally related to bacteriophage T4 DNA  $\beta$ -glucosyltransferase. We have used site-directed mutagenesis to mutate highly conserved and invariant aspartate, glutamate, and histidine residues and find that those residues required for enzymatic activity are predicted to line the active site cleft of a model based on the lac repressor crystal structure.

\* This work was supported in part by National Institutes of Health Grants DE08108 (to L. A. T.) and T35 DE07189. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed. Tel.: 716-275-0770; Fax: 716-473-2679; E-mail: Lawrence\_Tabak@urmc.rochester.edu.

<sup>1</sup> The abbreviations used are: ppGaNTase, UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase;  $\beta$ 4GalT,  $\beta$ 1,4-galactosyltransferase;  $\alpha$ 3GalT,  $\alpha$ 1,3-galactosyltransferase; GT1, glycosyltransferase motif 1; GalT, galactosyltransferase; PAGE, polyacrylamide gel electrophoresis; aa, amino acid.



**FIG. 1. Sequence and schematic representation of ppGaNTase.** *A*, sequence of the murine ppGaNTase-T1. *Solid underline* indicates the transmembrane anchor, *horizontal arrow* indicates the stem, *broken underlining* indicates the region conserved throughout the ppGaNTase family from worms to mammals, and the *solid boxes* indicate sequence motifs that were mutated in this study. The invariant carboxylate and histidine residues are indicated with an *asterisk*. The  $\beta$ - and  $\gamma$ -repeats with homology to the carbohydrate-binding sites of ricin are indicated with a *dashed box* and were not mutated. *N*-Glycosylation sites that were experimentally mapped (33) are *circled*. The soluble form of ppGaNTase-T1 is created by proteolytic processing at residue 40 (*vertical arrow*). *B*, schematic representation of the cytoplasmic, transmembrane, and luminal regions of ppGaNTases. A short N-terminal cytoplasmic tail is separated from the luminal catalytic domain by a transmembrane anchor (*solid box*) and a stem region (*cross-hatched box*) which varies in length and sequence among ppGaNTase isoforms. The catalytic unit of approximately 340 amino acids can be subdivided into two halves. The N-terminal half of the catalytic domain is represented by a GT1 sequence motif. The C-terminal half of the catalytic domain contains a Gal/GalNAc-T sequence motif. The far C-terminal end of the enzyme is a region with sequence and predicted structural homology to the plant lectin ricin (22). *Open circles* indicate positions of *N*-glycosylation sites.

#### EXPERIMENTAL PROCEDURES

**Secondary Structure Predictions and Threading Analysis**—The multiple sequence alignments of ppGaNTases were prepared with ClustalW (17). The secondary structure prediction was created by the JPREP server (<http://circinus.ebi.ac.uk:8081/>) using all sequences listed in Fig. 2A plus the other closely related ppGaNTase isoforms. The sequence homology between the ppGaNTase and  $\beta$ 1,4-galactosyltransferase ( $\beta$ 4GalT) families was discovered by PSI-BLAST (18) using the ppGaNTase family as the seed. The structural homology to the periplasmic binding protein family was proposed based on threading analyses with the program THREADER (19) using the human ppGaNTase-T1 isoform as the search sequence. All runs were carried out using default settings and scanning the full structural data base provided with the program.

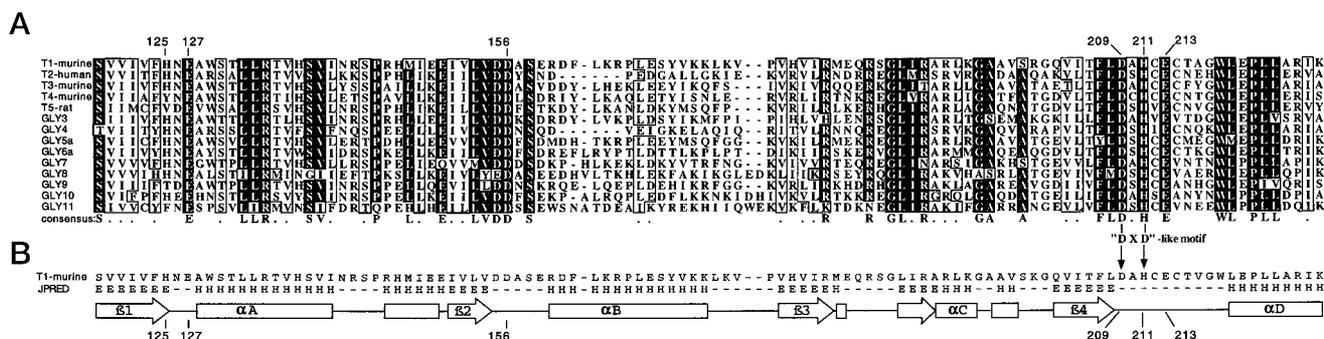
**Accession Numbers for ppGaNTases Sequences**—The sequences which include bovine, human, murine, porcine, and rat-T1 amino acid sequences are >98% identical and are available with accession numbers 539755, 2135942, 2149049, 1339955, and 1709559. Human-T2 is 2135941. Human and murine-T3 are 1617312 and 1575723. Murine-T4 is 2121220; rat-T5 is 3510639. *C. elegans* ppGaNTase sequence homologs GLY3, GLY4, GLY5a, GLY6a, GLY7, GLY8, GLY9 are AF031833, AF031834, AF031835, AF031838, AF031841, AF031842, AF031843, respectively. Full-length GLY10, GLY11 cDNA sequence is not complete.<sup>2</sup>

**Accession Numbers for  $\beta$ 1,4-Galactosyltransferases**—The numbers

for human  $\beta$ 4GalT-I, -II, -III, -IV, -V, and -VI are 86952, 2995442, 313298, 3132900, 2924555, and 3132904, respectively. Chicken CK- $\beta$ 4GalT-I and -II are 1469908 and 1469906, respectively. Snail  $\beta$ 4GalT is 2494837. *C. elegans* CE- $\beta$ 4GalT-A and -B are 1359573 and 4820871.

**Plasmid Constructs and Mutagenesis**—Mouse ppGaNTase-T1 isoform was amplified and cloned from a first-strand cDNA synthesis reaction of mouse kidney total RNA, using the following sense and antisense oligonucleotide primers Mlu-mT1 (CACACGCGTTGCCT-GCTGGTGACGTTCTAGAGCTAGT) and Bam-mT1 (ATGCGGATC-CAGCCCAGTCAATCCTTCCTT), respectively. Final cDNA sequence was obtained by sequencing both strands of multiple polymerase chain reaction-derived clones. The cDNA was cloned into a M13 vehicle and uracil-containing single-stranded DNA was prepared in the *Escherichia coli* strain CJ236. Mutagenesis, according to the procedure described by Kunkel (20), was first used to create a *Mlu*I cloning site in the cDNA encoding the stem region of the enzyme at amino acid 40, which is the position of the native N terminus for the secreted form of the enzyme. This construct, designated M13-mT1-Mlu, was used as a uracil-containing parent vector for all mutagenesis experiments. Mutants were identified by restriction enzyme screening and verified by DNA sequencing using a LiCOR 4000 DNA sequencer. Positive clones were subcloned into the COS7 cell expression vector pIMKF1, which contains an insulin secretion signal and a FLAG antibody recognition sequence (4). All expression constructs with mutants were verified by DNA sequencing. The initial mutations introduced *Bss*III cloning sites in the stem region at amino acid codon positions 56 and 73, which allowed for the construction and expression of truncated recombinant enzymes. While the

<sup>2</sup> F. K. Hagen, unpublished data.



**FIG. 2. Sequence and secondary structure predictions of the GT1 sequence motif in mammalian and *C. elegans* ppGaNTases.** *A*, ClustalW alignment of the GT1 motif is shown for 14 ppGaNTases. Amino acids 119 to 227 (representing 108 of the 112 aa which comprise the GT1 motif) from the murine ppGaNTase-T1 are indicated in the *top line*. *C. elegans* ppGaNTases are indicated as Gly proteins 3–11. Consensus sequences are defined, using a threshold of 85% conservation and are indicated on the consensus line and within the alignment as sequences in *reverse video print*. Periods on the consensus line indicate positions that contain a high sequence similarity but no identity. Key aspartates, glutamates, and histidines that are invariant are indicated by numbered positions, indicating their location in the murine ppGaNTase protein. *B*, secondary structure predictions were performed by the JPRED server. Rectangles indicate  $\alpha$ -helices and arrows indicate  $\beta$ -strands. The DXD-like sequence is a small aspect of this whole motif and is conserved as a DXH sequence in ppGaNTases.

$\Delta$ 1–55 and  $\Delta$ 1–73 deletion constructs were active, the  $\Delta$ 1–73 enzyme was labile and further deletions were not pursued. All remaining point mutants were constructed in the  $\Delta$ 1–40 background vector.

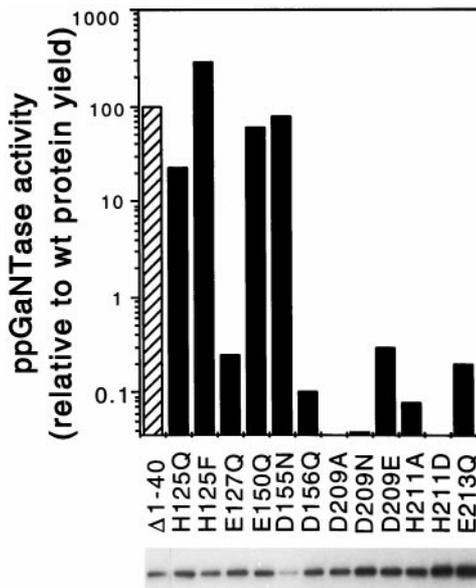
**Expression and Immunopurification of Recombinant Proteins**—Recombinant ppGaNTases were expressed as secreted products from COS7 cells and partially purified from cell culture medium as described previously (4). To normalize the quantity of recombinant proteins employed for enzyme assays, aliquots of the immuno-purified enzymes were  $^{32}$ P-labeled with heart muscle kinase and separated by SDS-PAGE, and the level of product was estimated by PhosphorImager analysis (Molecular Dynamics) as described previously (5). Unless noted, expression levels of all single point mutants of ppGaNTase-T1 were within the same order of magnitude of wild-type levels, and activities appeared to be stable during the time course of the *in vitro* glycosylation assays.

**ppGaNTase Assays**—Enzyme activity was measured by quantifying the transfer of [ $^{14}$ C]GalNAc to the peptide acceptor EA2 (PTTDDSTT-PAPTTP) using conditions previously described (4). Apparent  $K_m$  values were obtained for UDP-GalNAc, using a nucleotide-sugar concentration series ranging from 2.5 to 80  $\mu$ M with the EA2 peptide held at 500  $\mu$ M. The apparent  $K_m$  measurements for EA2 peptide used a peptide concentration series ranging from 31.25–1000  $\mu$ M with a final concentration of 50  $\mu$ M UDP-GalNAc. All transfections and enzyme assays were performed in duplicate and repeated twice.

## RESULTS AND DISCUSSION

**General Structural Features of the ppGaNTases**—In common with other glycosyltransferases that localize to the Golgi complex, the ppGaNTases are type II membrane proteins, characterized by a short (4–24 aa) N-terminal cytoplasmic tail, followed by a small (15–25 aa) transmembrane anchor, which is tethered to a large (>450 aa) segment in the lumen of the Golgi by a stem region of variable length. In Fig. 1A, the amino acid sequence of murine ppGaNTase-T1 is presented together with the structural features described above and the sequence motifs and structural elements to be described in the subsequent sections.

To define the catalytic unit of the ppGaNTase-T1, three approaches were taken: 1) putative domains of potential functional importance were delineated by identifying ppGaNTase sequence motifs that were related to motifs present in distant protein families (detected by either PSI-BLAST or Hidden Markov Modeling); 2) secondary structure predictions and fold recognition was used to propose a structural model for each motif (using JPRED and THREADER); and 3) mutagenesis was used to test the role of the conserved features of each motif in enzyme function. In this manner, three putative structural domains were identified in the luminal regions of ppGaNTases and the boundaries of the stem were defined (Fig. 1B). The putative stem regions of ppGaNTases are diverse both in terms



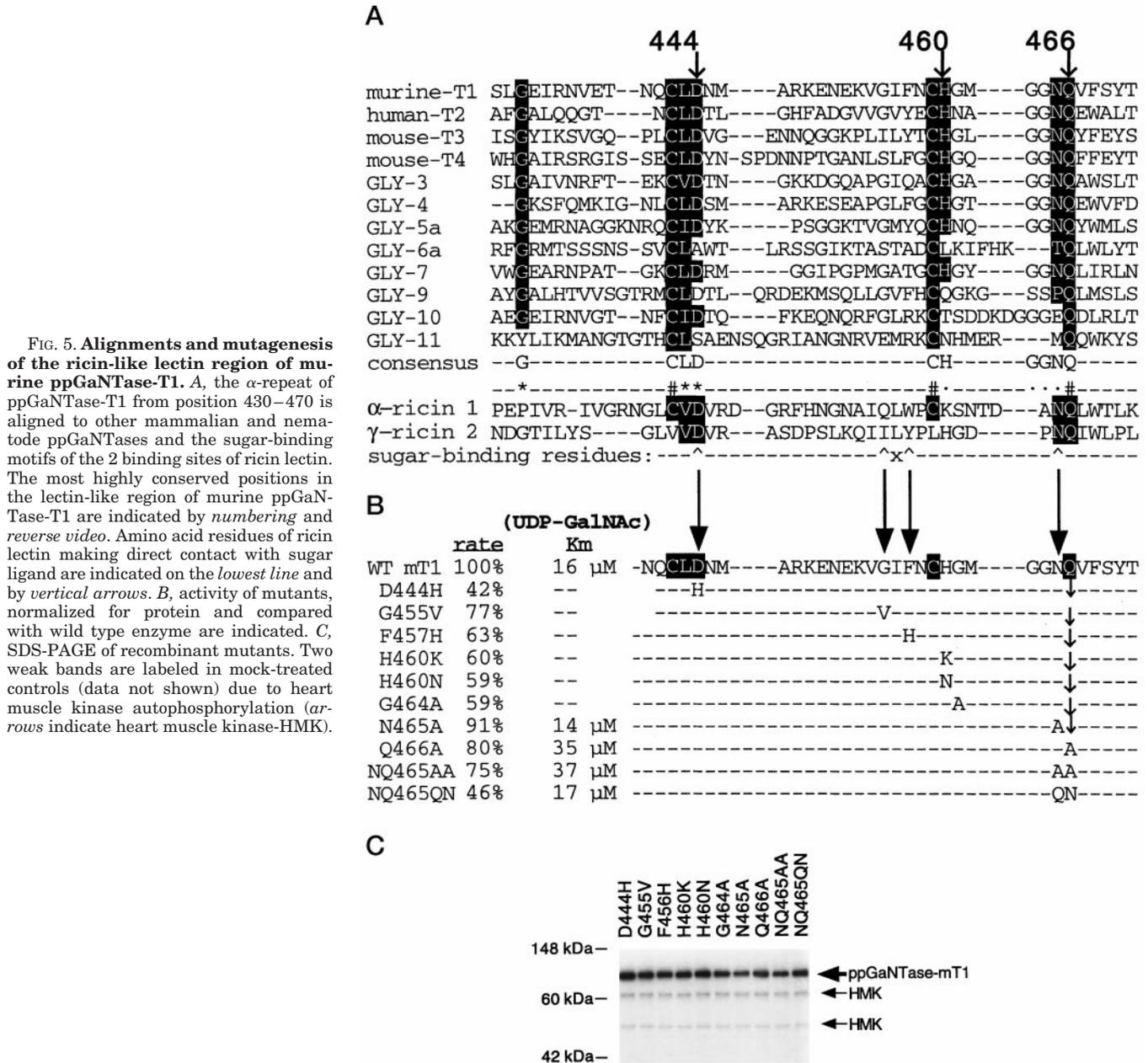
**FIG. 3. Enzyme activity of GT1 motif mutants.** The murine ppGaNTase was mutated at positions described in the legend to Fig. 2. Recombinant mutants were expressed, immunopurified, and quantitated, as described under “Experimental Procedures.” The yield of recombinant protein was determined by SDS-PAGE analysis (*lower panel*), and the enzyme activity was corrected for recombinant protein expression. The *bar chart* indicates the percent enzyme activity relative to the  $\Delta$ 1–40 wild type murine ppGaNTase-T1 (*Hatched bar*). *Solid bars* indicate the normalized activity of point mutants in the GT1 sequence motif.

of composition and length, ranging from 55 to 416 aa for ppGaNTase-T1 and -T5, respectively. Proteolytic cleavage within the stem region of many type II membrane proteins leads to the release of a soluble circulating species. From amino acid sequencing, it is known that the ppGaNTase-T1 is cleaved at residue 40 (21 and Fig. 1A).

In contrast to the highly variable stem region, all known ppGaNTases share a highly conserved block of approximately 340 aa, which span about two-thirds of the luminal portion of the enzyme (Fig. 1, A and B). Contained within this block of 340 aa are two unique sequence motifs. The first sequence motif is common to a wide range of glycosyltransferases<sup>3</sup> and hence termed the glycosyltransferase 1 (GT1) motif. The second se-

<sup>3</sup> B. Hazes and F. K. Hagen, manuscript in preparation.





**FIG. 5. Alignments and mutagenesis of the ricin-like lectin region of murine ppGaNTase-T1.** A, the  $\alpha$ -repeat of ppGaNTase-T1 from position 430–470 is aligned to other mammalian and nematode ppGaNTases and the sugar-binding motifs of the 2 binding sites of ricin lectin. The most highly conserved positions in the lectin-like region of murine ppGaNTase-T1 are indicated by *numbering* and *reverse video*. Amino acid residues of ricin lectin making direct contact with sugar ligand are indicated on the *lowest line* and by *vertical arrows*. B, activity of mutants, normalized for protein and compared with wild type enzyme are indicated. C, SDS-PAGE of recombinant mutants. Two weak bands are labeled in mock-treated controls (data not shown) due to heart muscle kinase autophosphorylation (*arrows* indicate heart muscle kinase-HMK).

the DXD sequence that has recently been described for many other glycosyltransferases (6–9).

ppGaNTases, which are retaining glycosyltransferases, are thought to work via a double-displacement mechanism involving at least two carboxylic acid residues (12). Amino acid sequence alignment of this region reveals 4 invariant carboxylic acid residues and 1 invariant histidine. Conservative mutagenesis of any one of the 4 invariant aspartates and glutamates in the GT1 motif to their amide forms resulted in a >99.8% loss of enzyme activity (Fig. 3). This indicates that a carboxylate is crucial at 4 specific sites: E127Q, D156Q, D209N, and E213Q. A D156N mutant was not made because this would have created a potential *N*-glycosylation site. Aspartate 209 was also mutated to an alanine and glutamate (D209A and D209E) to examine the effect of changing the size and charge of this essential residue. Mutants D209N and D209A, while stably expressed, had no detectable enzyme activity (<0.04%), whereas D209E exhibited a very low but significant level of activity (0.29%) relative to wild type. This latter result suggests that the negative charge of the aspartate carboxyl group is

absolutely necessary for catalytic activity and that its spatial position is extremely important. As a control, we mutated two carboxylate side chains that were highly conserved (but not invariant) to their amide forms (E150Q and D155N). Both mutants exhibited a wild type level of enzyme activity. In the case of D155N, the specific activity of the enzyme was not significantly affected; however, the expression of recombinant protein from COS7 cells was markedly compromised (Fig. 3). To confirm that this low level of protein production was not due to a cloning artifact, four independent mutants were constructed, sequenced, and expressed in COS7 cells. In each case, the D155N mutation resulted in a low level of protein expression, suggesting that the preference for aspartate at position 155 of the GT1 motif serves a structural rather than functional role.

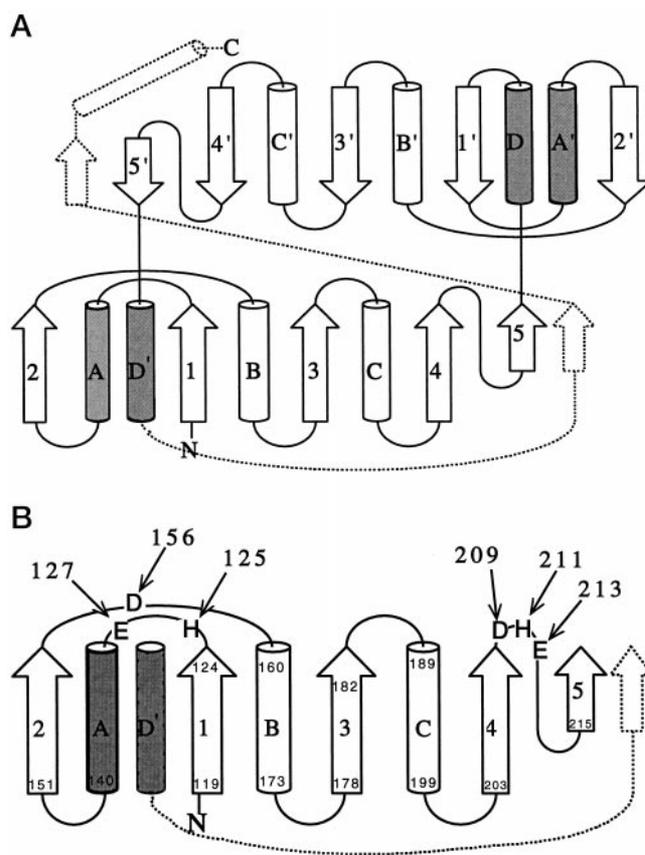
Sequence comparison indicates that the most highly conserved position in the GT1 motif corresponds to aspartate 209 in ppGaNTase-T1. The aspartate 209 analogue in three other GT1-containing glycosyltransferases (chitin synthase 2, MNN1  $\alpha$ 1,3-mannosyltransferase, and *Clostridium sordellii* lethal

toxin) has been mutated and found to be critical for enzyme activity (7, 9, 11). In most GT1-containing transferases, this critical position is part of a DXD motif. ppGaNTases, however, are exceptional in that there is strict conservation of a DXH motif instead (see Fig. 2A). The second conserved residue in this motif (H211 for ppGaNTases-T1 and aspartate for most other GT1 family members) is also critical for function. We have previously mutated H211 in bovine ppGaNTase-T1 and demonstrated that a H211A mutant is inactive (24). A common theme among DXD motif-containing proteins is their ability to bind nucleoside diphosphate sugars and coordinate manganese ions (7). One possibility is that H211 is a ligand for the metal ion. Because aspartate is the most common ligand for manganese, we converted the ppGaNTase DXH motif into the prototypical DXD motif by creating a murine ppGaNTase-T1 H211D mutant. However, this mutant had no detectable activity (<0.04% relative to wild type, Fig. 3), indicating that the aspartate could not functionally substitute for a histidine in DXH. To test if the H211D mutation altered the metal ion requirements of the enzyme, the H211D mutant was assayed in the presence of 1 mM Mg<sup>2+</sup>, Mn<sup>2+</sup>, Co<sup>2+</sup>, and Fe<sup>2+</sup>. No enzyme activity could be detected with these alternative divalent metal ions (data not shown).

Previously, we had reported that a H125A mutation inactivates recombinant bovine ppGaNTase-T1 (24). With a larger data set of ppGaNTases, it is now clear that this position, albeit highly conserved (found in 10 out of 14 ppGaNTase isoforms), is not invariant. Both the H125Q and H125F mutants were active, with the H125F mutant having a near 3-fold greater activity than wild type (Fig. 3). These results suggest that an aromatic group at position 125 enhances enzyme activity. Interestingly, a phenylalanine or tyrosine is preferred at this position in the GT1 motif of many other glycosyltransferases, suggesting that it serves a function common to the family.<sup>3</sup>

**Identification of Essential Residues within the Gal/GalNAc-T Motif**—The Gal/GalNAc-T sequence motif (Fig. 1) was identified by PSI-BLAST analysis, which revealed a 41-aa segment that has a distant sequence similarity to the  $\beta$ 4GalT family of enzymes (Fig. 4, A and B). This 41-amino acid segment, the Gal/GalNAc-T motif, contains three carboxylates in a DXXXXXWGGENXE sequence motif (Asp-310 to Glu-322) that are invariant in the ppGaNTase family. Mutation of these carboxylates to their amide forms (D310N, E319Q, and E322Q) reduces enzyme activity to 2, <0.04, and 1%, respectively (Fig. 4C). The WGGENXE sequence corresponds to region II (WGGEDDD) in the  $\beta$ 4GalT family (25, 26) and displays weaker sequence similarity with  $\alpha$ 1,3-galactosyltransferase ( $\alpha$ 3GalT), Blood Group B  $\alpha$ 3GalT, and the  $\alpha$ 1,3-GalNAc-transferases Blood Group A and Forssman synthetase (9). Region II of the  $\beta$ 4GalT family is thought to interact with both the sugar donor UDP-galactose and the sugar acceptor (27, 28). While aspartate 310 is invariant among ppGaNTases, only the positions containing the tryptophan, glycine, and glutamate residues (Glu-319 and Glu-322) in the WGGENXE sequence are conserved within the galactosyltransferase family as a WGGEDDD sequence motif. This suggests that the critical residues Glu-319 and Glu-322 may be involved in similar functions in both the ppGaNTase and  $\beta$ 4GalT families, as both families have individual isoforms capable of binding both UDP-Gal and UDP-GalNAc (29).

As a control, four highly conserved but not invariant residues have been mutated (N320A, D375A, D375N, E376Q, H341A, H341L, H341V, H341K, and H341R), but these mutations had little effect on enzymatic activity (Fig. 4C). With the accumulation of a larger sequence data base of ppGaNTase isoforms (1–5), it is now apparent that histidine 341 is not as conserved



**FIG. 6. Topology diagram of lactose repressor and hypothetical structure of the GT1 region of ppGaNTase-T1.** A, crystal structure of lactose repressor (31) indicated that the ligand-binding site is formed by two symmetrical domains that face each other and interlock by contributing  $\alpha$ -helices (see  $\alpha$ -helix D and D') to each domain. Each domain is composed of a parallel  $\beta$ -sheet, formed by five or six  $\beta$ -strands.  $\beta$ -Strands are indicated as broad arrows and lie in the plane of the paper, in a 2-1-3-4-5 order. The  $\alpha$ -helices (shown as cylinders) flank the  $\beta$ -sheet, such that the white  $\alpha$ -helices (B and C) are positioned behind the page and the dark  $\alpha$ -helices (A and D) are positioned in front of the page, thus forming a three-layered sandwich. Residues important for binding or interacting with substrate line the binding site at positions located at the C-terminal ends of the  $\beta$ -strands (arrowheads) or loop region just prior to the next  $\alpha$ -helix. The N and C termini of the lactose repressor are indicated with a N and C, respectively. The dotted line indicates a C-terminal extension that is not as well conserved within the periplasmic binding protein family. B, the hypothetical structure of the N-terminal half of the ppGaNTase-T1 is predicted by structural threading. Invariant carboxylic acid and histidine residues of the GT1 and DXD-like motifs, essential for enzyme activity, line the predicted face of the active site. THREADER predicts a significant match with the N-terminal half of the catalytic domain and agrees with the assignment of  $\beta$ -strands and  $\alpha$ -helices, predicted by different criteria (JPRED) in Fig. 2. The amino acid positions at the start and end of each strand, helix, or loop are given based on the JPRED prediction (see Fig. 2B) and the THREADER results.

as previously thought (24). In contrast, histidine 344, is invariant among all known ppGaNTases and is essential for catalytic activity (24).

**The Ricin-like Sequence of ppGaNTases Is Not Essential for Catalytic Activity**—The presence of a ricin-like sequence in the C terminus of ppGaNTase-T1 raised the possibility that this region may represent the binding site for the sugar donor UDP-GalNAc or for partially glycosylated peptide substrates (22, 23). In ClustalW alignments of the first of three repeats in the ricin-like motif, it is apparent that the key structural elements and carbohydrate-binding residues of ricin are preserved in the C-terminal end of most but not all ppGaNTases (Fig. 5A). For murine ppGaNTase-T1, Asp-444 and Asn-465 model to the conserved positions of the  $\alpha$ -repeat, which are

important for H-bonding to hydroxyl groups of galactose in ricin (22). The phenylalanine at position 457 aligns to an aromatic position in ricin that forms an aromatic stacking interaction with galactose (22). However, mutation of these positions in murine ppGaNtase-T1 did not substantially inhibit catalytic activity of recombinant enzyme (D444H, N465A, and F457H) in an *in vitro* enzyme assay (Fig. 5B). In contrast, the analogous amino acid substitutions in ricin reduced sugar binding by at least an order of magnitude (30). Furthermore, the double mutant N465A/Q465A, which should affect main chain hydrogen bonding in the ricin domain, as well as hydrogen bonding to galactose, was characterized by only a 2-fold increase in apparent  $K_m$  for UDP-GalNAc. Collectively, these findings indicate that the C-terminal lectin domain does not participate in an essential catalytic function as measured by *in vitro* glycosylation reactions.

**Essential Amino Acid Residues of ppGaNtases Lie Near the Predicted C-terminal Ends of  $\beta$ -Strands in a Lac Repressor Fold Topology**—A match between the GT1 motif of ppGaNtase and the lactose repressor fold was detected by threading analysis of some 1900 structures (score 4.21; empirical data suggests that scores  $>3.5$  are very significant (THREADER Users Manual)). The lac repressor fold consists of 2 symmetrical domains, in which each domain is composed of a 5-stranded parallel  $\beta$ -sheet flanked by 4  $\alpha$ -helices (Fig. 6A) (31). The  $\beta$ -strands align in a  $\beta 2$ - $\beta 1$ - $\beta 3$ - $\beta 4$ - $\beta 5$  order and these are flanked by 2  $\alpha$ -helices on either side. A proposed alignment of the ppGaNtase-T1 sequence to the known three-dimensional structure of the lactose repressor is shown in Fig. 6B. The assignment of  $\beta$ -strands and  $\alpha$ -helices agrees with the JPRED predictions depicted in Fig. 2B. The topology diagram of the lactose repressor depicts a binding cleft formed by two symmetrical domains that face each other. Amino acid side chains interacting with the sugar ligand are positioned at the C-terminal ends of the strands that line the binding cleft between the two domains (Fig. 6A). The threading of the ppGaNtase to the lactose repressor fold demonstrates that the aa positions essential for catalytic activity (127, 156, 209, 211, and 213) would also line the same face of the binding cleft (Fig. 6B). This mutagenesis data supports a model in which these positions form part of the sugar donor, acceptor, or  $Mn^{2+}$ -binding site. The crystal structure of the bacteriophage T4 DNA  $\beta$ -glucosyltransferase reveals that the first domain adopts a similar fold as the lactose repressor protein (32), suggesting that a lactose repressor fold may be conserved among glycosyltransferases.

The threading analysis with the second ppGaNtase domain revealed a weak match to the second domain of the lac repressor. Essential glutamates of the WGGENXE sequence motif lie in the active site cleft at the predicted C-terminal end of strand  $\beta 3$  (data not shown). Therefore, both halves of the catalytic unit contain 3 to 4 carboxylates and 1 histidine residue that are critical for enzyme activity.

The lac repressor structure model suggests that both the first and second domains of the catalytic unit in ppGaNtases contribute critical residues for the binding of the sugar donor, sugar acceptor, and manganese ion. To determine if binding affinities for the sugar donor and acceptor could be dissociated from one another, kinetic parameters were measured for UDP-GalNAc and the EA2 peptide for ppGaNtase-T1 mutants that retained enzyme activity. For each mutant tested, the  $K_m$  values for both UDP-GalNAc and peptide acceptor were simultaneously affected by a single mutation, suggesting an interdependence of binding of both substrates (data not shown). The lactose repressor and the bacteriophage T4 DNA  $\beta$ -glucosyl-

transferase crystal structure indicate that substrate binding shifts the equilibrium between open and closed forms of the protein (32). Accordingly, mutations that interfere with the binding of one substrate in ppGaNtases may affect binding of the other substrate through effects on domain opening or closure. Chemical modification studies suggest that the ppGaNtase-T1 undergoes a conformational change upon UDP-GalNAc binding (33). Therefore, it is possible that an equilibrium between the open and closed states is used as a mechanism to regulate ppGaNtase activity.

In the absence of a definitive experimental structure, our data provide a starting point for the creation of ppGaNtase-specific inhibitors and mutagenesis studies to both enhance ppGaNtase activity and modify substrate specificity. Studies of this type are currently in progress.

**Acknowledgments**—We thank Christine Cagnina and Brian Van Wuyckhuysse for excellent technical assistance. We also thank Meng Qian for help in preparing this manuscript.

#### REFERENCES

- Hagen, F. K., and Nehrke, K. (1998) *J. Biol. Chem.* **273**, 8268–8277
- Wandall, H. H., Hassan, H., Mirgorodskaya, E., Kristensen, A. K., Roepstorff, P., Bennett, E. P., Nielsen, P. A., Hollingsworth, M. A., Burchell, J., Taylor-Papadimitriou, J., and Clausen, H. (1997) *J. Biol. Chem.* **272**, 23503–23514
- Zara, J., Hagen, F. K., Ten Hagen, K. G., Van Wuyckhuysse, B. C., and Tabak, L. A. (1996) *Biochem. Biophys. Res. Commun.* **38**, 38–44
- Hagen, F. K., Ten Hagen, K. G., Beres, T. M., Balys, M., VanWuyckhuysse, B. C., and Tabak, L. A. (1997) *J. Biol. Chem.* **272**, 13843–13848
- Ten Hagen, K. G., Hagen, F. K., Balys, M. M., Beres, T. M., VanWuyckhuysse, B., and Tabak, L. A. (1998) *J. Biol. Chem.* **273**, 27749–27754
- Van Die, I., Bakker, H., and Van den Eijnden, D. H. (1997) *Glycobiology* **7**, v-ix
- Wiggins, C. A. R., and Munro, S. (1998) *Proc. Natl. Acad. Sci. U. S. A.* **95**, 7945–7950
- Breton, C., Bettler, E., Joziassse, D. H., Geremia, R. A., and Imberty, A. (1998) *J. Biochem.* **123**, 1000–1009
- Busch, C., Hofmann, F., Selzer, J., Munro, S., Jeckel, D., and Aktories, K. (1998) *J. Biol. Chem.* **273**, 19566–19572
- Breton, C., Oriol, R., and Imberty, A. (1996) *Glycobiology* **6**, R7–R12
- Nagahashi, S., Sudoh, M., Ono, N., Sawada, R., Yamaguchi, E., Uchida, Y., Mio, T., Takagi, M., Arisawa, M., and Yamada-Okabe, H. (1995) *J. Biol. Chem.* **270**, 13961–13967
- Sinnott, M. L. (1990) *Chem. Rev.* **90**, 1171–1202
- Sugiura, M., Kawasaki, T., and Yamashina, I. (1982) *J. Biol. Chem.* **257**, 9501–9507
- Elhammer, A., and Kornfeld, S. (1986) *J. Biol. Chem.* **261**, 5249–5255
- Wang, Y., Abernethy, J. L., Eckhardt, A. E., and Hill, R. L. (1992) *J. Biol. Chem.* **267**, 12709–12716
- Karlin, S., Zhu, Z.-Y., and Karlin, K. D. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 14225–14230
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402
- Jones, D. T., Taylor, W. R., and Thornton, J. M. (1992) *Nature* **358**, 86–89
- Kunkel, T. A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 488–492
- Hagen, F. K., VanWuyckhuysse, B., and Tabak, L. A. (1993) *J. Biol. Chem.* **268**, 18960–18965
- Hazes, B. (1996) *Protein Sci.* **5**, 1490–1501
- Imberty, A., Piller, V., Piller, F., and Breton, C. (1997) *Protein Eng.* **10**, 1353–1356
- Wragg, S., Hagen, F. K., and Tabak, L. A. (1997) *Biochem. J.* **328**, 193–197
- Lo, N.-W., Shaper, J. H., Pevsner, J., and Shaper, N. L. (1998) *Glycobiology* **8**, 517–526
- Almeida, R., Amado, M., David, L., Levery, S. B., Holmes, E. H., Merckx, G., von Kessel, A. G., Rygaard, E., Hassan, H., Bennett, E., and Clausen, H. (1997) *J. Biol. Chem.* **272**, 31979–31991
- Aoki, D., Appert, H. E., Johnson, D., Wong, S. S., and Fukuda, M. N. (1990) *EMBO J.* **9**, 3171–3178
- Boeggeman, E. E., Balaji, P. V., and Qasba, P. K. (1995) *Glycoconj. J.* **12**, 865–878
- Do, K. Y., Do, S. I., and Cummings, R. D. (1995) *J. Biol. Chem.* **270**, 18447–18451
- Frankel, A. E., Burbage, C., Fu, T., Tagge, E., Chandler, J., and Willingham, M. C. (1996) *Biochemistry* **35**, 14749–14756
- Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., and Lu, P. (1996) *Science* **271**, 1247–1254
- Vrieland, A., Ruger, W., Driessen, H. P. C., and Freemont, P. S. (1994) *EMBO J.* **13**, 3413–3422
- Wragg, S., Hagen, F. K., and Tabak, L. A. (1995) *J. Biol. Chem.* **270**, 16947–16954