Functional Characterization and Expression Analysis of Members of the UDP-GalNAc:Polypeptide N-Acetylgalactosaminytransferase Family from Drosophila melanogaster*

Kelly G. Ten Hagen‡§, Duy T. Tran‡, Thomas A. Gerken‡, David S. Stein, and Zhenyu Zhang

From the §Departments of Biological Chemistry, NIDDK, National Institutes of Health, Bethesda, Maryland 20892, and the ¶Departments of Pediatrics and Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106-4948, and (§Section of Molecular Cell and Developmental Biology, Institute for Cellular and Molecular Biology, University of Texas, Austin, Texas 78312

Here we report the cloning and functional characterization of eight members of the UDP-GalNAc-polypeptide N-acetylgalactosaminytransferase gene family from Drosophila melanogaster (polypeptide GalNAc transferase = pgant1–8). Full-length cDNAs were isolated from a Drosophila embryonic library based on homology to known ppGaNTases. Alignments with characterized mammalian isoforms revealed strong sequence similarities between certain fly and mammalian isoforms, highlighting putative orthologues between the species. In vitro activity assays demonstrated biochemical transferase activity for each gene, with three isoforms requiring glycosylated substrates. Comparison of the activities of Drosophila and mammalian orthologues revealed conservation of substrate preferences against a panel of peptide and glycopeptide substrates. Furthermore, Edman degradation analysis demonstrated that preferred sites of GalNAc addition were also conserved between certain fly and mammalian orthologues. Semi-quantitative PCR amplification of Drosophila cDNA revealed expression of most isoforms at each developmental stage, with some isoforms being less abundant at certain stages relative to others. In situ hybridization to Drosophila embryos revealed specific staining of pgant5 and pgant6 in the salivary glands and pgant5 in the developing hindgut. Additionally, pgant5 and pgant6 expression within the egg chamber was restricted to the follicle cells, cells known to be involved in egg formation and subsequent embryonic patterning. The characterization reported here provides additional insight into the use of this model system to dissect the biological role of this enzyme family in vivo during both fly and mammalian development.

The recent completion of genome sequencing for many eukaryotic organisms as well as advances in proteomic analysis have placed new emphasis on the study of post-translation modifications. Glycosylation is by far the most abundant modification seen in nature. Protein substrates may be modified by the addition of N-linked oligosaccharides, O-linked fucose, O-linked xylose, O-linked galactose, O-linked mannose, O-linked GlcNAc, or O-linked GalNAc followed by subsequent sugar additions, resulting in the formation of very complex structures (1–3). Mucin-type O-linked glycosylation involves the transfer of GalNAc from the nucleotide sugar UDP-GalNAc to the hydroxyl group of either serine or threonine in protein substrates. A large, evolutionarily conserved family of enzymes, known as the UDP-GalNAc:polypeptide N-acetylgalactosaminytransferases (ppGaNTases) (EC 2.4.1.41), is responsible for this catalysis (reviewed in Ref. 1). To date, 13 distinct mammalian isoforms of this enzyme family have been functionally characterized, with each displaying a unique combination of expression patterns and well as substrate specificity. These enzymes have been shown to fall into two general categories as follows: those that will transfer GalNAc onto unmodified (as well as modified) peptide substrates (peptide transferases) and those that require the prior addition of a GalNAc on a peptide substrate before they will add additional GalNAc moieties (glycopeptide transferases) (4).

Estimates based on homology searches of the human and mouse genome data bases predict a total of 24 members of this enzyme family in mammals (1). A total of nine genes are predicted in Caenorhabditis elegans, with four being functionally confirmed as transferases (5). Searches of the Drosophila genome data base indicate that up to 14 polypeptide GalNAc transferase (pgants) may be present. Of those, two have been cloned and functionally characterized (6, 7). One of these genes, pgant35A, was found to be required for viability in Drosophila (6, 7). Studies are currently underway to delete the mouse orthologue of pgant35A (ppGaNTase-T11) in hopes of gaining insight into the function of this gene in mammals. However, mouse strains deficient in ppGaNTase-T1 (8) and -T13 (9, 10) do not display definitive phenotypes that would provide information as to the biological function of these gene products in vivo. Given the potentially large size of the mammalian ppGaNTase family as well as overlapping substrate preferences and expression patterns, individual gene knockouts may not show distinct phenotypes. This potential problem necessitates the use of model organisms (where the family size is smaller and genetic analysis more tractable) to aid in dissecting the role of this enzyme family in vivo.

In an effort to begin to use Drosophila as a model system in experimental biology, we focused on the characterization of a panel of eight members of the fly GalNAc transferase family. By using functional expression analysis in Drosophila embryos, we were able to determine the presence of functional GalNAc transferase activity and expression patterns as well as substrate specificity. These enzymes are predicted to have profound effects on cell surface biology and have been implicated in the formation of very complex structures. In the future it will be important to analyze these proteins in an in vivo context to determine their role in the development of this model organism.
which to study the ppGaNTases, we performed genome data base searches and screened cDNA libraries to identify and characterize potential members of the ppGaNTase enzyme family. Sequence analysis of fly and mammalian transferases indicates that certain isoforms were present prior to the divergence of these species. In *in vitro* analysis of the cloned isoforms demonstrates biochemical activity and reveals functional conservation between certain *Drosophila* and mammalian orthologues.

**EXPERIMENTAL PROCEDURES**

**Isolation of ppGaNTase Full-length cDNAs**—The amino acid consensus sequence SPTMAGGLFAVNRKYFQHLGEY, derived from the conserved region of previously characterized mammalian ppGaNTases, was used to perform a tBLASTn search against the existing *D. melanogaster* genome data base present in NCBI to identify all potential members of this enzyme family. The 14 predicted *D. melanogaster* ppGaNTase gene sequences obtained were aligned to identify highly conserved regions on which to base degenerate probes to screen cDNA libraries. The primers MAGGLF-S (dATGGCCGGCGGNCTGTTTGCC- CTG) and WGGEN-A (dATCTCCANATTCTGCGGCCCCCA) were used to amplify a 100-bp fragment from *D. melanogaster* genomic DNA that should hybridize to all 14 predicted isoforms. This amplified genomic fragment was then radioactively labeled with Digoxigenin DNA Labeling System (Invitrogen) and used to probe a *D. melanogaster* Canton-S embryo (2–14-h-old) UniZap XR cDNA library (Stratagene catalog no. 937602). Hybridizations were performed in 5× SSPE, 5% formamide at 42 °C with washes in 2× SSC, 0.5% SDS for 5 min at room temperature and for 15 min at 65 °C. Positively hybridizing clones were further screened with isoform-specific primers to eliminate duplicate clones, hybridizing in 5× SSPE, 5% formamide at 30 °C with wash in 2× SSC, 0.5% SDS for 5 min at room temperature and 15 min at 42 °C. Isoform-specific primers were as follows: FlyA (dAGCGCGGAAT- GCAGCAGGGATG) for pgant1, FlyC (dAGCGCGGAATCAGCGGGGAG) for pgant2, FlyD (dAGCGCGGAATGCGGGGAG) for pgant3, FlyG (dAGCGCGGAATGCAGCCGGGAG) for pgant4, FlyH (dATCGGATCGGCGGAGG; for pgant5, FlyJ (dATCGGATCGGCGGAGG) for pgant6, FlyK (dATCGGATCGGCGGAGG) for pgant7, FlyL (dATCGGATCGGCGGAGG) for pgant8, and FlyM (dATCGGATCGGCGGAGG) for pgant9. Clones were then sequenced using the T3 and T7 primers to identify each isoform obtained. Clones containing complete open reading frames were completely sequenced on both strands (Lark Technologies, Inc.). cDNA clones of pgant3 (GH09147) (GenBank accession number AF145655), pgant4 (AT-25481) and pgant8 (RE06471) (GenBank accession number) were purchased from Research Genetics (Invitrogen).

**Amino Acid Alignments and Similarity Determinations**—Amino acid alignments were performed using the following parameters: neighbor joining; tie breaking = systematic; Poisson-correction; proportional gap distribution. Numbers at the nodes (Fig. 2) represent reproducibility of particular groupings and are expressed as percentage of recovery in 1000 replicates. Nodes that appeared less than 50% of the time were not retained. Sequences comprising the conserved domains used in Figs. 1 and 2 began with the consensus sequence SPTMAGGLFAVNRKYFQHLGEY, a 340-amino acid region in the putative catalytic domain of each ppGaNTase, and corresponds to the putative catalytic domain based on structural modeling and mutagenesis studies (22).

**PCR Analysis**—cDNA panels at four different dilutions (1, 10, 100, and 1000×) from various *D. melanogaster* tissues and stages of development were obtained from OriGene Technologies, Inc. The following regions were amplified from each gene using the specified primer pairs: 330-bp region of pgant1 using primers Fly-1230S (dGCTCTGTTGTT- GTTGAATGATGG) and Fly-1565AS (dAGGACATTGTGGTTGTTG- AAGGAG); a 282-bp region of pgant3 using primers FlyB-1489S (dGTGCGGACCTGAGCCAGACCAAGG) and FlyB-1711AS (dGCTGGCGGTG- GTTGAATGACCTGAC); a 255-bp region of pgant2 using primers FlyC- 724AS (dGACGATTTAGCTGCGGACAACTGAGA) and FlyC-970AS (dTCGCGGACCTGAGCCAGACCAAGG); a 206-bp region of pgant4 using primers FlyD-918S (dATCGGAGAAATAGGCGGAGG) and FlyD-1241AS (dCGGAG- AAACATCCGAGCACTACC); a 334-bp region of pgant5 using primers Flyg-1085S (dDTAGCATGTTGCTTACAGGCGGAGG) and Flyg-1419AS (dCGGAGAAACATCCGAGCACTACC); a 345-bp region of pgant6 using primers Flyg-633S (dCACCGAATACCATGGCGGAGG) and Flyg-892AS (dCGCGGACCTGAGCCAGAACTGAC); and a 357-bp region of pgant7 using primers Flyg-966S (dCTCGTTGCGACATATGGGAGG) and Flyg-1323AS (dGACGTGTCGCTTCCCTTCCAGC) and Flyg-1419AS (dCGGAGAAACATCCGAGCACTACC) for the primer pair for 1 μl. In the case of each gene, sequenced to verify that no PCR-induced mutation had occurred. To clone pgant1, an 1.8-kb site was introduced into the pgant1 cDNA (Fly-4a) and cloned into the same vector. Amplification using the primer pairs for pgant3, pgant4, and pgant6 was performed according to the manufacturer’s instructions. Reaction products were electrophoresed in a 1% TAE-agarose gel and photographed on a Bio-Rad Fluor-ST™ MultiImager.

**Generation of Secretion Constructs for *Drosophila* ppGaNTase**—Each cDNA was cloned into an SV40 expression vector (pIMFK4) to generate a recombinant fusion protein containing an insulin secretion signal, a metal-binding site, a heart muscle kinase site, a FLAG™ epitope tag, and the transferase gene of interest. Fusions of each transferase gene begin after the transmembrane domain by introduction of an *Mlu*I site, so that the recombinant protein is efficiently secreted into the media. PCR products used in the cloning of each gene were sequenced to verify that no PCR-replication had occurred. To clone pgant1, an 1.7-kb fragment was isolated from the original cDNA clone for pgant1 (Fly-4a) and cloned into the same vector. Amplification using the primer pairs for pgant3, pgant4, and pgant6 was performed according to the manufacturer’s instructions. Reaction products were electrophoresed in a 1% TAE-agarose gel and photographed on a Bio-Rad Fluor-ST™ MultiImager.

**Generation of Secretion Constructs for *Drosophila* ppGaNTase**—Each cDNA was cloned into an SV40 expression vector (pIMFK4) to generate a recombinant fusion protein containing an insulin secretion signal, a metal-binding site, a heart muscle kinase site, a FLAG™ epitope tag, and the transferase gene of interest. Fusions of each transferase gene begin after the transmembrane domain by introduction of an *Mlu*I site, so that the recombinant protein is efficiently secreted into the media. PCR products used in the cloning of each gene were sequenced to verify that no PCR-replication had occurred. To clone pgant1, an 1.7-kb fragment was isolated from the original cDNA clone for pgant1 (Fly-4a) and cloned into the same vector. Amplification using the primer pairs for pgant3, pgant4, and pgant6 was performed according to the manufacturer’s instructions. Reaction products were electrophoresed in a 1% TAE-agarose gel and photographed on a Bio-Rad Fluor-ST™ MultiImager.
CATCGGCCA) and FlyE-900AS (dGAAGATCTGAGAGAGGAGGAGCTG- 
AGT) and FlyJ-1919S (dATAACCGGTGACCGGACGTGG- 
ATG) and FlyJ-596AS (dAGATCGCCGTCCAGAGAATCCAT- 
CTG); after digestion with DmBamHI, this fragment was cloned into the 
MidUbgII sites of pMKF4 to generate the vector, pMKF4-J. The 
expression vector for pgant7 was generated by amplification of a 350-bp 
fragment from the pgant7 cDNA (Fly-42a) using the primers FlyG- 
MidUbgI (dATAACCGGTAGACCGGACGTGCGCCAGCGAGGCGCGA- 
TAT) and FlyG-BglII (dGAAGATCTGAGAGAGGAGGAGCTG- 
AGT) and cloned into the BsmBI/BglII fragment of pBSFyE was then cloned into the 
BsmBI/BglII sites of pMKF4 to generate the vector, pMKF4-J. A 1.7-kb 
Pci/KpnI (blunt) fragment from Fly-42a was then inserted into the 
Pci/KpnI (blunt) sites of pMKF4-J to generate the expression vector, pMKF4-J.

Functional Expression Assays of Secreted Recombinant ppGaNTases in COS7 Cells—COS7 cells were grown to 80% confluency and 
transfected with pMKF4, pF4-T7, pF4-mT2, pF4-mT1, pF4-FyA, pF4-FyB, pF4-FyC, pF4-FyD, pF4-FyE, pF4-FyF, pF4-FyG, 
or pF4-FyJ and 64 μl of LipofectAMINE (Invitrogen) as described 
previously (14). Recombinant enzymes were purified using FLAG-affin- 
y agarose (Sigma), labeled with [γ-32P]ATP using heart muscle kinase 
(Sigma) and quantitated by Tricine SDS-PAGE as described previously 
(14). Enzyme activity is expressed as dpm/h.

RESULTS
cDNA clones for eight novel D. melanogaster UDP-GalNAc: 
opositive N-acetylgalactosaminyltransferase genes were 
obtained from either the D. melanogaster cDNA collection 
through Research Genetics (pgant3, pgant4, and pgant8) (In- 
vitrogen) or were isolated from a D. melanogaster Canton S 
embryonic cDNA library (pgant1, pgant2, pgant5, pgant6, and 
pgant7) (Stratagene). Library screening was performed using 
degenerate PCR probes based on consensus sequences derived 
from putative D. melanogaster pgGaNTases found in the data 
base. All novel cDNA clones were sequenced completely on both 
strands. One cDNA isolated (pgant7) was nearly identical to the 
previously described dGalNAc-T2 (7) except for three nucleo- 
cid differences, one of which resulted in a proline replacing 
a serine at position 84. Conceptual translation of each cDNA 
reveals type II membrane proteins consisting of an N-terminal 
cytoplasmic region, a hydrophobic/transmembrane region, 
a stem region and a putative catalytic region. Fig. 1 shows an 
amino acid alignment of the putative catalytic region of the 
Drosophila isoforms and mouse pgGaNTase-T1; highlighted 
regions denote areas of extensive conservation between differ- 
ent isoforms as well as between species. Drosophila isoforms 
contain the conserved regions corresponding to the GT1 and 
GalNAc-T motifs defined previously in mammalian trans- 
ferases (22).

The Drosophila isoforms cloned here as well as five putative 
transferases from the FlyBase/GalFly (CG30463, CG10000, 
CG7304, CG31776, and CG7579) were compared with all mam- 
mal isoforms characterized to date (ppGaNTases-T1-T14) 
by best tree and bootstrap analysis in order to assess related- 
ness across species. The best tree in Fig. 2A provides a phylo- 
genetic analysis with distances between pairs shown on each 
branch of the tree. The bootstrap analysis (Fig. 2B) provides an 
assessment of the confidence in the groupings shown by deter- 
mining the percentage of times that these groupings occur in 
1000 replicates. The phylogenetic trees obtained demonstrate 
the evolutionary divergence of members of this enzyme family 
both within and across species. Although clear subgroups of 
family members from both species emerge, there are also clear 
orthologous pairs between mammals and flies. Only pairs that 
were seen in both trees were considered to be orthologous. For 
example, rat ppGaNTase-T7 is more closely related to the 
Drosophila isoform, PGANT 7, than to any other mammalian 
transferase characterized to date (67% aa similarity within the 
conserved region). Additionally, PGANT2 and mammalian 
pp-GaNTase-T2 are most closely related at the sequence level 
(86% aa similarity within the conserved region), as are 
PGANT3A and mammalian pgGaNatas-T11 (71% aa simi- 
larity within the conserved region). Although mammalian 
pp-GaNTase-T1 and -T13 share extensive aa similarity in this 
region (93%), each also shares a great degree of similarity with 
Drosophila isoforms defined previously in mammalian trans- 
ferases (22).
The resultant recombinant proteins contained a FLAG epitope tag for enrichment and a kinase site for labeling the expressed proteins. Recombinant isoforms were partially purified from the cell media and quantitated relative to one another by SDS-PAGE as described previously (15). However, no recombinant protein or transferase activity was recovered for PGANT3 and PGANT8 following FLAG purification, suggesting that the N-terminal portion of the proteins containing the FLAG site had been lost. Therefore, in vitro transferase activity for these two recombinant proteins was assessed directly from the media against a panel of peptides as described (14). Equal relative amounts of the remaining recombinant proteins enriched by FLAG affinity (PGANT1, -2, and -5–7) were tested against the same panel of peptides, with the exception of PGANT4, where maximal amounts were used to detect activity relative to the other isoforms.

As shown in Fig. 3, all Drosophila isoforms cloned displayed biochemical activity in vitro, indicating that each is capable of functioning as a UDP-GalNAc-polypeptide transferase. Initial velocities for each transferase against a panel of peptides are shown. Like the mammalian transferases characterized previously, the fly transferases fall into two general categories of activity: those that will transfer GalNAc onto peptide substrates (peptide transferases), and those that require the prior addition of a GalNAc before they will transfer further GalNAc residues (glycopeptide transferases). As seen in Fig. 3, PGANT1, PGANT2, PGANT3, PGANT5, and PGANT8 all demonstrate peptide transferase activity, by virtue of the fact that they transfer GalNAc onto various combinations of unmodified peptides. These same enzymes display activity on glycopeptide substrates as well (MUC5AC-3, MUC5AC-13, and/or MUC5AC-3/13), although PGANT3 activity on these substrates was very low. By comparison of initial velocities, PGANT1 and PGANT2 appear to prefer the mono-glycosylated substrate, MUC5AC-3 (Fig. 3A and B); and PGANT8 prefers both EA2 and the diglycosylated MUC5AC-3/13, albeit at very low levels (Fig. 3E).

In contrast to the activities described above, PGANT4, PGANT6, and PGANT7 did not act appreciably on unmodified peptides but rather transferred GalNAc to previously modified glycopeptide substrates. Consistent with this result, preliminary characterization of a cDNA clone homologous to PGANT7 (dGalNAc-T2) suggested a preference for previously glycosylated substrates as well (7). Whereas PGANT4 and PGANT6 showed preferential transfer to the diglycopeptide (MUC5AC-3/13) (Fig. 3, C and D), PGANT7 acted on the monoglycopeptide (MUC5AC-3), the diglycopeptide, and to a lesser degree a second monoglycopeptide (MUC5AC-13) (Fig. 3A). The low levels of activity seen for PGANT4 with Muc1B and rMUC2 were not above background values obtained in reactions containing no acceptor substrate. Table I summarizes information about all putative Drosophila UDP-GalNAc transferases and their activities (where known).

We next wished to compare acceptor substrate preferences and sites of GalNAc addition between Drosophila and mammalian orthologous pairs (Figs. 3, F–H, and 4). The substrate specificity of the orthologous pair consisting of PGANT2 and mammalian ppGaNTase-T2 is shown in Fig. 3F. Both enzymes show a rather broad substrate specificity, adding GalNAc to both peptide and glycopeptide substrates. Initial velocities for both enzymes were highest with the monoglycopeptide, MUC5AC-3. When the glycosylation sites of MUC5AC-3 were mapped after incubation with each enzyme, it was determined that both PGANT2 and ppGaNTase-T2 added additional GalNacs to threonines at positions 9, 10, 12, and 13 (Fig. 4A); ppGaNTase-T2 also gave small amounts of addition at serine 5 (Fig. 4A). Neither enzyme utilized threonine 2, serine 11, or serine 14.

Comparison of PGANT5 and ppGaNTase-T1 substrate preferences revealed that both act on peptide and glycopeptide

FIG. 1. Amino acid sequence alignments of Drosophila PGANT proteins and murine ppGaNTase-T1 (mT1). Amino acid sequences were aligned within the putative catalytic domain, beginning with the consensus sequence FNXXSXD and extending to the C terminus. Shaded blocks indicate regions of similarity or identity. A consensus sequence is given below the alignments for positions that are greater than 50% conserved among the isoforms displayed. The GT1 and Gal/GalNAc-T motifs are shown.
substrates; however, ppGaNTase-T1 has a broader substrate specificity, acting on a larger number of substrates (Fig. 3G). Analysis of the sites of addition of PGANT5 and ppGaNTase-mT1 on the EA2 substrate revealed glycosylation by both enzymes at threonines 7, 11, and 12, demonstrating similar preferences for sites of addition (Fig. 4B). Neither PGANT5 nor ppGaNTase-mT1 added GalNAc to threonine 3, serine 5, or threonine 6.

Fig. 3H compares the substrate preferences of the orthologous pair composed of PGANT7 and mammalian ppGaNTase-T7. Both enzymes in this pair act only on the three glycopeptide substrates in the panel, prefering the monoglycopeptide, MUC5AC-3, and the diglycopeptide, MUC5AC-3/13. Analysis of the sites of GalNAc addition on MUC5AC-3/13 by each enzyme (Fig. 4C) revealed that each preferentially transferred GalNAc to threonine 2, immediately N-terminal to the previously glycosylated threonine 3 in this substrate. Threonine 12 and serine 11, which are also immediately N-terminal to a glycosylated residue (threonine 13), were modified by both enzymes as well. Transfer of GalNAc to positions vicinal to a pre-existing GalNAc by glycopeptide transferases has been reported previously (4). ppGaNTase-rT7 also displayed smaller percentages of addition to threonine 10 and serine 14. Neither PGANT7 nor ppGaNTase-T7 utilized serine 5 in this substrate. Previous work demonstrated similar activity for the orthologous pair composed of Drosophila PGANT35A (dGalNAc-T1) and mammalian ppGaNTase-T11 across a panel of peptides (7). Examining the sites of addition of each enzyme on the peptide substrate EA2 demonstrates that both PGANT35A and ppGaNTase-T11 preferentially glycosylated threonine 7, with lesser amounts of glycosylation also seen at threonines 6 and 11 (Fig. 4D). Moreover, relative amounts of addition at each site by each enzyme were similar. Previous work on PGANT35A (dGalNAc-T1) demonstrated that both enzymes added GalNAc to threonines 6 and 7, although relative occupancy at each site was not given (7). Cumulatively, the results
described here point to functional conservation between orthologous pairs such that certain substrate preferences are retained and similar propensities for specific sites of addition within substrates are preserved.

Expression Analysis of pgant35 Genes—To address the temporal expression pattern of pgant genes, we used semi-quantitative PCR amplification of panels containing cDNA from staged D. melanogaster embryos, larvae, pupae, and adult males and females. Primers were designed from within the coding region of each gene and used to amplify gene-specific products from panels containing four dilutions of each cDNA (1, 10, 100, and 1000) (Fig. 5). A representative picture is shown for each isoform. It is worth noting that expression seen during early embryonic stages may represent maternal contribution as opposed to active zygotic transcription as oocytes typically receive maternal RNA for many genes prior to fertilization. PCR amplification demonstrates that pgant1, pgant3, pgant5, pgant6, and pgant7 transcripts are found throughout various stages of embryonic, larval, and pupal development as well as in the adult head and body of both males and females; additionally, each gene shows a slight to moderate increase in expression as larval development proceeds. pgant2 displays a more restricted expression pattern, with low levels of expression in the male and female body and correspondingly low levels during early embryonic stages, indicating a paucity of maternal RNA and zygotic transcription during this time period. Significant expression of pgant2 is then seen as early as embryonic stage 8–12 h and reaches maximal levels in the pupae and male head. pgant4 expression is also low during early embryonic stages but displays a dramatic increase at 12–24 h of embryonic development. pgant4 continues to be expressed through to adulthood but displays greatly lower levels in the female adult as compared with the male, again likely contributing to the dearth of transcripts seen in the early embryonic stages. Expression of pgant8 is low during early embryonic stages but increases during 12–24 h of embryogenesis through larval development and continues to be expressed throughout adulthood, albeit at slightly lower levels in males than females. No product is seen in the negative control for each panel, which lacked cDNA (Fig. 5, lane 13 in all panels). Fig. 5 also shows amplification of previously characterized pgant35A (6) for comparison and the rp49 control (bottom panels).
In Situ Hybridization Analysis of Drosophila Embryos and Egg Chambers—To obtain insights into the function of the various Drosophila UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases during development, we carried out RNA in situ hybridization studies to assess whether the various isoforms exhibited informative cell- or tissue-specific patterns of gene expression during embryogenesis. By using digoxigenin-labeled DNA probes, we did not observe specific

**FIG. 4.** Site of GalNAc addition by orthologous pairs on selected substrates. Sites of GalNAc addition were mapped by Edman degradation as described under “Experimental Procedures.” Potential glycosylation sites are shown on the horizontal axis and percent glycosylation of each threonine (Thr) and serine (Ser) residue is shown on the vertical axis. The peptide sequence of substrates examined is shown at the top of each panel. The peptide sequence of substrates examined is shown at the top of each panel. **A**, PGANT2 and mammalian ppGaNTase-T2 with MUC5AC-3. **B**, PGANT5 and mammalian ppGaNTase-T1 with EA2. **C**, PGANT7 and mammalian ppGaNTase-T7 with MUC5AC-3/13. **D**, PGANT35A and mammalian ppGaNTase-T1 with EA2.

**TABLE I**

Summary of Drosophila UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases

<table>
<thead>
<tr>
<th>CG no.</th>
<th>Gene designation</th>
<th>Protein designation</th>
<th>Coding region</th>
<th>Location</th>
<th>Biochemical activity</th>
<th>Mammalian orthologue</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG3843</td>
<td>NA</td>
<td>NA</td>
<td>650&lt;sup&gt;c&lt;/sup&gt;</td>
<td>2R; 53C11-C14</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>CG10000</td>
<td>NA</td>
<td>NA</td>
<td>558&lt;sup&gt;e&lt;/sup&gt;</td>
<td>3R; 98E3</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>CG7304</td>
<td>NA</td>
<td>NA</td>
<td>563&lt;sup&gt;d&lt;/sup&gt;</td>
<td>3L; 71F1</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>CG31776</td>
<td>NA</td>
<td>NA</td>
<td>601&lt;sup&gt;e&lt;/sup&gt;</td>
<td>2L; 23F3</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>CG7579</td>
<td>NA</td>
<td>NA</td>
<td>498&lt;sup&gt;d&lt;/sup&gt;</td>
<td>3L; 71F1</td>
<td>NA</td>
<td>NA</td>
</tr>
</tbody>
</table>

<sup>a</sup>Homologous to dGalNAc-T2 (7).
<sup>b</sup>From Ref. 6 and homologous to dGalNAc-T1 (7).
<sup>c</sup>NA, not applicable.
<sup>d</sup>Predicted from FlyBase.

panel) to demonstrate the presence of cDNA in each sample well.
patterns of expression for most of the transferases tested. However, pgant5 and pgant6 produced specific and strong staining in the salivary glands of embryos (Fig. 6, A and B, respectively). Strong expression of pgant6 was initially detected in the salivary glands of stage 12 embryos, becoming stronger at stage 13 (Fig. 6B) and continuing through the remainder of embryogenesis. pgant5 also exhibited conspicuous expression in the hindgut of the stage 16 embryo (Fig. 6A). pgant5 and pgant6 were further seen to be expressed during oogenesis, in the somatically derived follicle cells that surround the developing oocyte (Fig. 6, C and D). These cells are involved in the maturation of the oocyte and construction of the egg shell, as well as playing a role in subsequent embryonic pattern formation.

**DISCUSSION**

In this report we describe the cloning and expression of eight additional members of the *D. melanogaster* UDP-GalNAc: polypeptide N-acetylgalactosaminyltransferase gene family. Like their mammalian counterparts, these genes encode type II membrane proteins and demonstrate biochemical transferase activity *in vitro*. Phylogenetic analysis of 14 fly isoforms (9 confirmed transferases plus 5 putative transferases) and 14 previously characterized mammalian isoforms demonstrates the existence of clear orthologous pairs between the two species. The groupings obtained here were similar to those obtained by Schwientek et al. (7) using 13 fly isoforms (2 confirmed transferases plus 11 putative transferases) and 12 mammalian isoforms. For example, PGANT7 is more closely related to mammalian ppGaNTase-T7 at the aa sequence level within the conserved domain than either is to other isoforms within the same species. Additionally, PGANT2 is most closely related to mammalian ppGaNTase-T2 and, as described previously, PGANT35A is mostly closely related to mammalian ppGaNTase-T11 (6, 7). In these three cases, a fly and a mammalian isoform are more similar to one another than either is to any other isoform within their respective species. Although mammalian ppGaNTase-T1 is most closely related to mammalian ppGaNTase-T13 (93% aa similarity in the conserved region), both are very similar to *Drosophila* PGANT5 (81% similarity).

The phylogenetic data presented in Fig. 2 suggest the existence of ancestral isoforms of this family that were present prior to the divergence of deuterostomes and protostomes. These results imply that this family, as well as specific members of this family, have been under evolutionary pressure resulting in their maintenance throughout the evolution of eukaryotic organisms. Indeed, one previously characterized member of this gene family has been shown to be essential for viability in *D. melanogaster* (6, 7). These observations suggest that the functions of this family are likely required for higher eukaryotic development and viability as well.

Biochemical analysis of these isoforms indicates that, as has been observed for their mammalian counterparts, these transferases are Type II membrane proteins that are able to catalyze the transfer of GalNAc from UDP-GalNAc to acceptors. In addition, the phylogenetic analysis suggests that the evolution of this family has been under selective pressure, perhaps driven by the need for precise control of glycosylation in eukaryotic development.

**FIG. 5.** PCR panels of pgant gene expression. cDNA panels from embryonic, larval, pupal, and adult stages of *D. melanogaster* (Origene Technologies) were PCR-amplified to detect the expression of the pgant genes. Lanes are as follows: lane 1, 0–4-h embryo; lane 2, 4–8-h embryo; lane 3, 8–12-h embryo; lane 4, 12–24-h embryo; lane 5, 1st instar larva; lane 6, 2nd instar larva; lane 7, 3rd instar larva; lane 8, pupa; lane 9, male head; lane 10, female head; lane 11, male body; lane 12, female body; lane 13, PCR-negative control. The genes examined are shown to the left of each panel. Arrows denote the position of the specific PCR fragment for each gene. Bands seen below the specific PCR fragment in some panels are primers. The rp49-positive control is shown in the bottom panel.

**FIG. 6.** In situ hybridization of pgant5 and pgant6 to *Drosophila* embryos and egg chambers. pgant5 (A and C) and pgant6 (B and D) were hybridized to embryos (A and B) as well as egg chambers (C and D). The arrows in A and B show expression of pgant5 and pgant6 in developing salivary glands. The arrows in C and D highlight expression seen in the follicle cell layer surrounding the developing oocyte. A, stage 16 embryo. B, stage 13 embryo. C, stage 11 egg chamber. D, stage 10B egg chamber.
been found in mammals, there are members that require substrates previously modified by the addition of GalNAc (glycopeptide transferases) as well as those that will transfer to unmodified peptides (peptide transferases). In Drosophila, PGANT4, -6, and -7 represent glycopeptide transferases, while PGANT1-3, -5, and -8 are peptide transferases. Phylogenetic analysis based on the sequence of the putative catalytic region shows that these fly glycopeptide transferases are present in the same subgroup as the two known mammalian glycopeptide transferases (Fig. 2). Therefore, there exists a general form of functional conservation between fly and mammalian isoforms within this subgroup on the phylogenetic tree. Analysis of these latest glycopeptide transferases may allow the detection of sequences specific to this group that are responsible for their unique substrate requirements.

The orthologous pairs identified in this study show degrees of specific functional conservation as well. The glycopeptide transferases, PGANT7 and ppGaNTase-T7, both show a preference for addition to sites immediately N-terminal (threonines 2, 12, and serine 11) to the position of pre-existing GalNAcs in the glycopeptide substrate. In contrast to the glycopeptide transferases described above, PGANT2 and ppGaNTase-T2 transfer GalNAc to a broad array of substrates, including both peptides and glycopeptides. The sites of addition by PGANT2 and ppGaNTase-T2 on a glycopeptide substrate do not include the residues N-terminal to the pre-existing GalNAc but other residues in the center of the substrate (threonines 9, 10, 12, and 13). Other potential sites in the substrate are not used by either enzyme. These results highlight functional similarities between members of orthologous pairs as well as elucidate differences between peptide and glycopeptide transferases.

While the orthologous pair consisting of PGANT5 and ppGaNTase-T1 did not show strong similarities in terms of substrate preference, both did add GalNAc preferentially to threonines 7, 11, and 12 of the EA2 peptide. Similarly, the orthologous pair of PGANT35A and ppGaNTase-T11, which had been shown to act similarly on a panel of peptides (7), shared sites of addition as well as comparable relative levels of addition at threonines 6, 7, and 11 of EA2. Again, other potential sites of addition within the respective substrates were not used. Therefore, certain similarities and preferences in activity exist between orthologous pairs. Future biochemical analyses on additional substrates will provide more insight into the extent of the functional conservation between orthologous pairs. Whether this functional conservation is the result of passive similarity remaining since the time of divergence or a more active selection for certain essential biochemical activities remains to be determined.

Five additional putative UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferases are annotated in FlyBase, yet were not isolated from the Drosophila embryo cDNA library (CG30463, CG10000, CG7304, CG31776, and CG7579). The degenerate PCR probe used should have hybridized to all 14 predicted isoforms. However, it is possible that levels of some isoforms are low during embryogenesis, resulting in a poor representation of cDNA clones in the library screened. Isolation, expression, and biochemical characterization of these putative isoforms will be necessary to confirm their specific activities and whether or not they represent functional transferases.

Expression analysis of cDNA from various stages of Drosophila development revealed that most isoforms described here were expressed throughout embryonic, larval, pupal, and adult stages, with increasing levels during larval development. Spatial expression of each isoform within specific cells and tissues was also examined in developing oocytes and egg chambers. Whereas most isoforms did not show specific patterns of expression, pgant5 and pgant 6 were expressed very specifically in the developing salivary glands of embryos; pgant5 was also expressed in the developing hindgut. Additionally, pgant5 and pgant6 were expressed in the follicle cells surrounding the developing oocyte. Salivary glands and follicle cells are active secretory tissues in the fly. Salivary glands will eventually produce large amounts of salivary gland secretion proteins (Sgs) that are synthesized (Sgs1 and Sgs4) during late larval stages to allow the adherence of larvae to surfaces appropriate for pupariation (36).

Within the egg chamber, follicle cells are involved in establishing polarity of the developing oocyte (37). Follicle cells are also responsible for secreting the structural components of the egg (chorion and vitteline membrane) and certain mucin-like molecules (hemomucin) (38). The presence of pgant5 and pgant6 transcripts may indicate their potential involvement in the glycosylation of certain molecules involved in these processes. Information regarding specific spatial expression of these isoforms will enable us to perform a more directed search for potential substrates in appropriate cells/tissues and provide insights into biological functions during development.

In summary, we have functionally characterized eight additional members of the ppGaNTase family in D. melanogaster. Future studies will be directed toward defining isoforms required for viability and identifying their native substrates. Given both the sequence similarity and functional conservation seen between Drosophila and mammalian isoforms, we hope to eventually elucidate the role of O-linked glycosylation in higher eukaryotic development.

Acknowledgments—We thank Dr. Lawrence Tabak for continued support during the course of these studies. We also thank Jason Goltz and Jason Rarick for technical assistance.

REFERENCES