

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

Received for publication, April 11, 2003, and in revised form, June 20, 2003  
Published, JBC Papers in Press, June 26, 2003, DOI 10.1074/jbc.M303836200

Kelly G. Ten Hagen<sup>‡§</sup>, Duy T. Tran<sup>‡</sup>, Thomas A. Gerken<sup>¶</sup>, David S. Stein<sup>||</sup>, and Zhenyu Zhang<sup>||</sup>

From the <sup>‡</sup>Section of Biological Chemistry, NIDDK, National Institutes of Health, Bethesda, Maryland 20892, the <sup>¶</sup>Departments of Pediatrics and Biochemistry, Case Western Reserve University, Cleveland, Ohio 44106-4948, and <sup>||</sup>Section of Molecular Cell and Developmental Biology, Institute for Cellular and Molecular Biology, University of Texas, Austin, Texas 78712

Here we report the cloning and functional characterization of eight members of the UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase 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 mod-

ification 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*-acetylgalactosaminyltransferases (ppGaNTases)<sup>1</sup> (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

\* This work was supported in part by National Institutes of Health Grant GM52761 (to D. S.). 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.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY268063, AY268064, AY268065, AY268066, AY268067, and AY268068.

§ To whom correspondence should be addressed: Section of Biological Chemistry, NIDDK, National Institutes of Health, 9000 Rockville Pike, Bldg. 50, Rm. 4120, Bethesda, MD 20892. Fax: 301-480-4214; E-mail: Kelly.Tenhagen@nih.gov.

<sup>1</sup> The abbreviations used are: ppGaNTase, UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferase; aa, amino acids; Tricine, *N*-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine.

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 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 *Drosophila 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 (dATGGCCGGCGGNGCTGTTGCC-AT) and WGGEN-AS (dATCTCCANATTCTCGCCGCCCA) 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 using the Random Primers 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, 50% 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, 50% formamide at 30 °C with washes 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 (dAGCGGAATCGCCACGGGGATG) for pgant1, FlyC (dTGCGGATGAGCTGTGAGTAGTG) for pgant2, FlyE (dTGCCGAGCATGTGAGCGGTGAGG) for pgant5, FlyF (dCGCAAGGAATGCCACTGCAGAGG) for pgant6, FlyG (dTGCTTTGCATCCCGCAGGTGAGG) for pgant7, and FlyI (dAGCCTCAAGTATTTGTGGCAAG) for CG30463. 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 sequences were aligned, one pair at a time, using the pairwise ClustalW (1.4) algorithm in MacVector (Oxford Molecular Group). The following alignment modes and parameters were used: slow alignment, open gap penalty = 10, extended gap penalty = 0.1, similarity matrix = blosum, delay divergence = 40%, hydrophilic penalties, and residue-specific penalties. Best tree and bootstrap analyses were performed using the following parameters: neighbor joining; tie breaking = systematic; Poisson-correction; proportional gap distribution. Numbers at the nodes (Fig. 2B) 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 begin with the consensus sequence FNXXXSD in the putative catalytic domain (aa position 84 in ppGaNtase-mT1 (11), 104 in ppGaNtase-hT2 (12), 150 in ppGaNtase-mT3 (13), 102 in ppGaNtase-mT4 (14), 454 in ppGaNtase-rT5 (15), 142 in ppGaNtase-hT6 (16), 175 in ppGaNtase-rT7 (17), 149 in ppGaNtase-hT8 (18), 119 in ppGaNtase-hT9 (19), 113 in ppGaNtase-rT10 (4), 119 in ppGaNtase-hT11 (7), 103 in ppGaNtase-h12 (20), 83 in ppGaNtase-h13 (10), 79 in ppGaNtase-h14 (21), 114 in PGANT35A (6), 116 in PGANT1, 152 in PGANT2, 118 in PGANT3, 161 in PGANT4, 142 in PGANT5, 170 in PGANT6, 110 in PGANT7, 95 in PGANT8, 176 in CG30463, 75 in CG10000, 75 in CG7304, 103 in CG31776, and 18 in CG7579 and end at a conserved residue (amino acid position 421 in ppGaNtase-mT1, 436 in ppGaNtase-hT2, 495 in ppGaNtase-mT3, 434 in ppGaNtase-mT4, 792 in ppGaNtase-rT5, 487 in ppGaNtase-hT6, 520 in ppGaNtase-rT7, 490 in ppGaNtase-hT8, 458 in ppGaNtase-hT9, 445 in ppGaNtase-rT10, 454 in ppGaNtase-hT11, 435 in ppGaNtase-hT12, 420 in ppGaNtase-hT13, 410 in ppGaNtase-hT14, 452 in PGANT35A, 463 in PGANT1, 484 in PGANT2, 459 in PGANT3, 497 in PGANT4, 479 in PGANT5, 504 in PGANT6, 454 in PGANT7, 432 in PGANT8, 512 in CG30463, 425 in CG10000, 420 in CG7304, 440 in CG31776, and 351 in CG7579). The

segment of conserved sequences is ~340 amino acids in length in the various isoforms 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: a 335-bp region of pgant1 using primers FlyA-1230S (dGGCTCTGGT-TTGGATGGATGAG) and FlyA-1565AS (dAGGACATTTGTGTTGGTG-AAGGAG); a 282-bp region of pgant3 using primers FlyB-1489S (dT-TGTCACGAGAGTGAAGCGAG) and FlyB-1771AS (dCGTTGCGGTT-CTTTAGCATCTTG); a 255-bp region of pgant2 using primers FlyC-724S (dCGAGTGATTGCAACGACAAG) and FlyC-979AS (dTCAGGT-TCCAGTCAAAGCCG); a 259-bp region of pgant4 using primers FlyD-981S (dTAGCGGAAATAAGGACGGAGC) and FlyD-1241AS (dCAGG-GAACATCCAACAGCATAACC); a 334-bp region of pgant5 using primers FlyE-1085S (dATGAGATTGGCTCCTACGACGAGG) and FlyE-1419AS (dCAGATACCAGCGGAAACTCTTG); a 259-bp region of pgant6 using primers FlyF-633S (dCAACGAATACCTGAGCGTGCTG) and FlyF-892AS (dCGTGGGAATCAAGGAAGATGAG); a 357-bp region of pgant7 using primers FlyG-966S (dCTGTTCGCCATCAATAGGGAG) and FlyG-1323AS (dGCAGTTCAGTCGTTCTTCAGC); an 845-bp region of pgant35A using primers FlyH-1330-S (dCCTCATCAAGTCGGAGAA-CG) and FlyH-2175-AS (dAGGCACAGCAACTTGTCCAG); a 213-bp region of pgant8 using primers FlyJ-1496S (dGATTCTATTGCTTGGAGGTGCG) and FlyJ-1709AS (dCACTGTGCTTATGGTTCTTCGG). Primer pairs were used under the following PCR conditions: 35 cycles of 95 °C for 1 min, 5 °C below the lowest  $T_m$  for the primer pair for 1 min, and 72 °C for 1 min followed by 1 cycle of 72 °C for 10 min. rp49 control PCR amplifications were performed according to the manufacturer's primers and instructions. Reaction products were electrophoresed in a 1% TAE-agarose gel and photographed on a Bio-Rad Fluor-S™ MultiImager.

**Generation of Secretion Constructs for *Drosophila* ppGaNtases**—Each cDNA was cloned into an SV40 expression vector (pIMKF4) 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-induced mutation had occurred. To clone pgant1, an *Mlu*I site was introduced into a fragment of the pgant1 cDNA (Fly-4a) by PCR amplification using the primers FlyA-59S (dATAACGCGTTCAGCAGAAATGGTTCCACCA) and FlyA-925AS (dGAA-GATCTGTACTGAAAGTCGTTGGCATCG); after digestion with *Mlu*I, the product was cloned into the *Mlu*I/*Bgl*II (blunt) sites of pIMKF4 to generate pF4-FlyA700. A 1.7-kb *Bsm*BI/*Sac*I fragment was isolated from the original cDNA clone for pgant1 (Fly-4a) and cloned into the same sites of pF4-FlyA700 to generate the pgant1 expression vector, pF4-FlyA. The expression vector for pgant3 was constructed by PCR amplification of a fragment from the pgant3 cDNA (GH09147) using the primers FlyB-*Mlu*-S (dATAACGCGTTCAGGCGGGGACGCGGAG) and FlyB-*Bgl*II-AS (dGAAGATCTTGGGCGTCGAGGAAGGTCAGCA-C); this fragment was digested with *Mlu*I/*Bgl*II and cloned into the *Mlu*I/*Bgl*II sites of pIMKF4 to generate the vector pF4FlyB-640. A *Bsp*TI/*Xba*I (blunt) fragment from the pgant3 cDNA clone was then cloned into the *Bsp*TI/*Sac*I (blunt) sites pF4FlyB-640 to generate the expression vector pF4-FlyB. The expression vector for pgant2 was generated by PCR amplification of a 374-bp fragment from the pgant2 cDNA clone (FlyC-1a) using the primers FlyC-*Mlu*IS (dATAACGCGT-CAAGTGGTCTGGCACCAGGTT) and FlyC-*Bgl*IIAS (dGAAGATCT-CACTCGACCTGTCTAATCTTCGCCA); after digestion with *Mlu*I, this product was cloned into the *Mlu*I/*Bgl*II (blunt) sites of pIMKF4 to generate the vector, pF4FlyC-374. A 1.7-kb *Bsu*361/*Xba*I (blunt) fragment from the pgant2 cDNA clone was then cloned into the *Bsu*361/*Sac*I (blunt) sites of pF4FlyC-374 to generate the expression vector pF4-FlyC. To generate the pgant4 expression vector, an *Mlu*I site was introduced into a fragment of the pgant4 cDNA (AT25481) by PCR amplification using the primers FlyD-185S (ATAACGCGTTGAAAAA-TGCGGCGGAGCTGA) and FlyD-1000AS (dAATCTTCATGCGAAATA-GTATCCACCA); this product was digested with *Mlu*I/*Bgl*II and cloned into the *Mlu*I/*Bgl*II sites of pIMKF4 to generate the vector, pF4-FlyD820. A 1.9-kb *Afl*II/*Sma*I fragment from AT25481 was then cloned into the *Afl*II/*Bgl*II (blunt) sites of pF4-FlyD820 to generate the expression vector, pF4-FlyD. The expression vector for pgant5 was generated by amplification of an 800-bp fragment from the pgant5 cDNA clone, pBSFlyE, using the primers FlyE-82S (dATAACGCGTACTCGGACTG-

CATCGGCA) and FlyE-900AS (dGAAGATCTAATCACATCAATAATCGGGCAC). This fragment was then digested with *MluI/BglII* and cloned into the *MluI/BglII* sites of pIMKF4 to generate the vector, pF4FlyE-819. A 1.4-kb *BsmBI/BglII* fragment from pBSFlyE was then cloned into the *BsmBI/BglII* sites of pF4FlyE-819 to generate the expression vector, pF4-FlyE. The pgant6 expression vector was constructed by cloning the 570-bp *MluI/BglII* PCR fragment, obtained using primers FlyF-*MluIS* (dATAACGCGTTCTCATCTACTCCGGACACCAACA) and FlyF-*BglIIAS* (dGAAGATCTGCATCAGCACGCTCAGGTATTC) with the pgant6 cDNA (FlyF-1a), into the *MluI/BglII* sites of pIMKF4 to generate the vector, pF4FlyF-570. Then a 1.9-kb *BstBI/NdeI*(blunt) from pBSFlyE was cloned into the *BstBI/NotI*(blunt) sites of pF4FlyF-570 to generate the expression vector, pF4-FlyF. The expression vector for pgant7 was generated by amplification of a 350-bp fragment from the pgant7 cDNA (Fly-42a) using the primers FlyG-*MluIS* (dATAACGCGTACAAGCGCGTCCAGGAGGCGTAT) and FlyG-*BglIIAS* (dGAAGATCTTGAAGACAATGATGACGCTCGTGC); this fragment was digested with *MluI/BglII* and cloned into the *MluI/BglII* sites of pIMKF4 to generate the vector, pF4FlyG-349. A 1.7-kb *PciI/XhoI*(blunt) fragment from Fly-42a was then inserted into the *PciI/SacI*(blunt) sites of pF4FlyG-349 to generate the expression vector, pF4-FlyG. A fragment from the pgant8 cDNA (RE06471) was amplified using the primers FlyJ-191S (dATAACGCGTTGGAGGGGAGCGTGATG) and FlyJ-596AS (dGAGGATCCCGTCCACCAGGATAACCTCTCG); after digestion with *MluI/BamHI*, this fragment was cloned into the *MluI/BglII* sites of pIMKF4 to generate the vector, pF4FlyJ-405. A 1.7-kb *BstBI/BamHI*(blunt) fragment from RE06471 was then cloned into the *BstBI/SacI*(blunt) sites of pF4FlyJ-405 to generate the expression vector, pF4-FlyJ.

**Functional Expression Assays of Secreted Recombinant ppGaNtases in COS7 Cells**—COS7 cells were grown to 90% confluency and transfected with pIMKF4, pF1-rT7, pF3-mT2, pF3-mT1, pF3-mT11, pF4-FlyA, pF4-FlyB, pF4-FlyC, pF4-FlyD, pF4-FlyE, pF4-FlyF, pF4-FlyG, or pF4-FlyJ and 64  $\mu$ M of LipofectAMINE (Invitrogen) as described previously (14). Recombinant enzymes were purified using FLAG-affinity agarose (Sigma), labeled with [ $\gamma$ -<sup>32</sup>P]ATP using heart muscle kinase (Sigma) and quantitated by Tricine SDS-PAGE as described previously (15). Gels were dried under vacuum, exposed to film (XAR, Eastman Kodak Co.), and quantitated on a Personal Molecular Imager FX (Bio-Rad). Enzyme assays were conducted using equivalent amounts of FLAG-purified PGANT1, -2, and -5-7 based on gel densitometric measurements. Maximal amounts of FLAG-purified PGANT4 were used in assays as its activity was low relative to the other FLAG-purified isoforms. Assays for PGANT3 and PGANT8 were performed using cell media, as FLAG purification resulted in no detectable protein and loss of activity (likely due to an N-terminal cleavage of the recombinant proteins, resulting in loss of the FLAG site). All enzymes were tested against the following peptide and glycopeptide substrates: EA2 (PTTDSTTPAPPTK) (23); MUC1b (PDTRPAPGSTAPPAC) (24); rMUC-2 (SPTTSTPISSTPQPTS) (25); mG-MUC (QTSSPNTGKTSTISTI) (26); MUC5AC (GTTTSPVPTTSTTSSAP) (27); MUC5AC-3 (GTT\*PSPVPTTSTTSSAP) (where the \* denotes a GalNAc modified residue); MUC5AC-13 (GTTTSPVPTTSTT\*SAP); IgAh (PSTPPTPSPSTPPTPSPS) (28); and Drosocin (GKPRPYSRPTSHPRIRV) (29). Glycopeptide substrates were the kind gift of H. Hang and C. Bertozzi. All reactions were performed in duplicate at 37 °C for 1 h. Reactions were performed in 25- $\mu$ l final volumes consisting of the following: 500  $\mu$ M acceptor substrate, 7.3  $\mu$ M [<sup>14</sup>C]UDP-GalNAc (54.7 mCi/mmol; 0.02mCi/ml), 44  $\mu$ M cold UDP-GalNAc, 10 mM MnCl<sub>2</sub>, 40 mM cacodylate (pH 6.5), 40 mM 2-mercaptoethanol, and 0.1% Triton X-100. Reaction products were purified using anion exchange chromatography (AG 1-X8, Bio-Rad). Reactions using media or FLAG-purified material from cells transfected with vector alone yielded background values for each substrate that were subtracted from each experimental value. Adjusted experimental values for each substrate were then averaged, and standard deviations were calculated. Enzyme activity is expressed as dpm/h.

**Amino Acid Sequencing**—Reactions used to determine the position of GalNAc addition were performed over 96 h at 37 °C under the following conditions: 15  $\mu$ g of acceptor substrate, 1 mM leupeptin, 1 mM aprotinin, 1 mM E-64, 1 mM phenylmethylsulfonyl fluoride, 1.25 mM AMP, 10 mM MnCl<sub>2</sub>, 40 mM cacodylate (pH 6.5), 40 mM 2-mercaptoethanol, 0.1% Triton X-100, 440  $\mu$ M UDP-GalNAc. Additional enzyme and UDP-GalNAc (44 nmol) were added after each 24-h period. Reactions were purified over an AG 1-X8 and then over a G-15 Sephadex column in 15 mM acetic acid (pH 4) (Amersham Biosciences). Pulsed liquid phase Edman degradation amino acid sequencing was performed on an Applied Biosystems Procise 494 protein sequencer (Applied Biosystems,

Foster City, CA). The extent of serine and threonine glycosylation by GalNAc was determined as described previously (30) after correcting for the relative recoveries and overlap of the individual amino acid phenylthiohydantoin derivatives.

**RNA in Situ Hybridizations to Embryos and Ovaries**—Whole mount *in situ* hybridizations to overnight collections of embryos and to dissected ovaries from Oregon R wild-type females were performed according to Tautz and Pfeifle (31). Digoxigenin-labeled DNA probes were prepared by random primed labeling of purified DNA fragments derived from cDNAs corresponding to the various pgant genes under study. Embryo staging was performed according to Ref. 32. Egg chamber staging is from Ref. 33).

## RESULTS

cDNA clones for eight novel *D. melanogaster* UPD-GalNAc: polypeptide *N*-acetylgalactosaminyltransferase genes were obtained from either the *D. melanogaster* cDNA collection through Research Genetics (pgant3, pgant4, and pgant8) (Invitrogen) 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* ppGaNtases 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 nucleotide differences, one of which resulted in a proline replacing a serine at aa 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 ppGaNtase-T1; highlighted regions denote areas of extensive conservation between different isoforms as well as between species. *Drosophila* isoforms contain the conserved regions corresponding to the GT1 and Gal/GalNAc-T motifs defined previously in mammalian transferases (22).

The *Drosophila* isoforms cloned here as well as five putative transferases from the FlyBase/GadFly (CG30463, CG10000, CG7304, CG31776, and CG7579) were compared with all mammalian isoforms characterized to date (ppGaNtases-T1-T14) by best tree and bootstrap analysis in order to assess relatedness across species. The best tree in Fig. 2A provides a phylogenetic analysis with distances between pairs shown on each arm of the tree. The bootstrap analysis (Fig. 2B) provides an assessment of the confidence in the groupings shown by determining 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 ppGaNtase-T2 are most closely related at the sequence level (86% aa similarity within the conserved region), as are PGANT35A and mammalian ppGaNtase-T11 (71% aa similarity within the conserved region). Although mammalian ppGaNtase-T1 and -T13 share extensive aa similarity in this region (93%), each also shares a great degree of similarity with PGANT5 (81%).

To determine whether the novel *Drosophila* isoforms represent functional transferases, the truncated coding region of each isoform was cloned into a mammalian expression vector and transfected into COS7 cells as described previously (14).

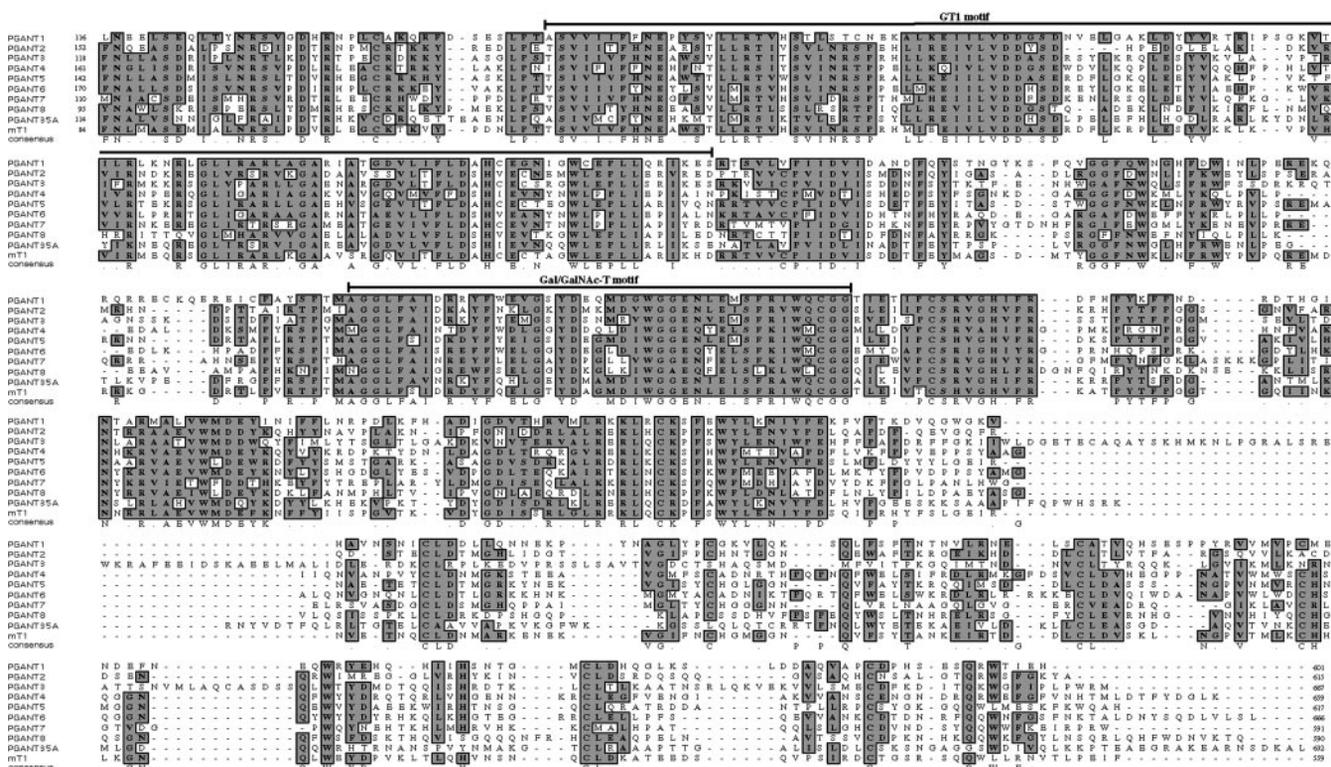


FIG. 1. Amino acid sequence alignments of *Drosophila* PGANT proteins and murine ppGaNTase-T1 (*mTI*). Amino acid sequences were aligned within the putative catalytic domain, beginning with the consensus sequence FNXXXSD and extending to the C terminus. Shaded blocks indicate regions of similarity or 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.

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.

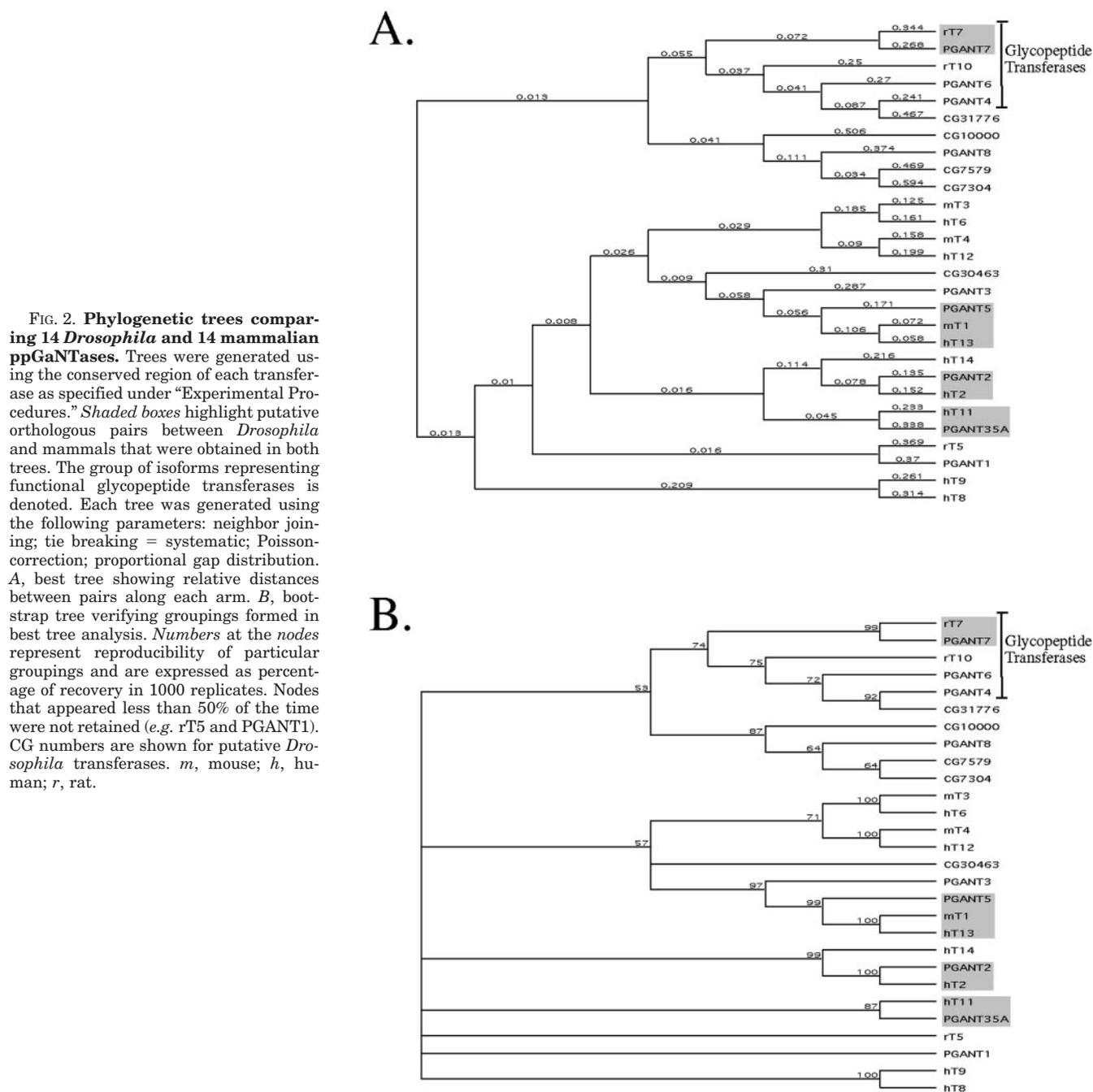
As shown in Fig. 3, all *Drosophila* isoforms cloned displayed biochemical activity *in vitro*, indicating that each is capable of functioning as a UDP-GalNAc-polypeptide transferase. Initial velocities for each transferase against a panel of peptides are shown. Like the mammalian transferases characterized 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); PGANT3 and PGANT5 prefer EA2 (Fig.

3, A 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

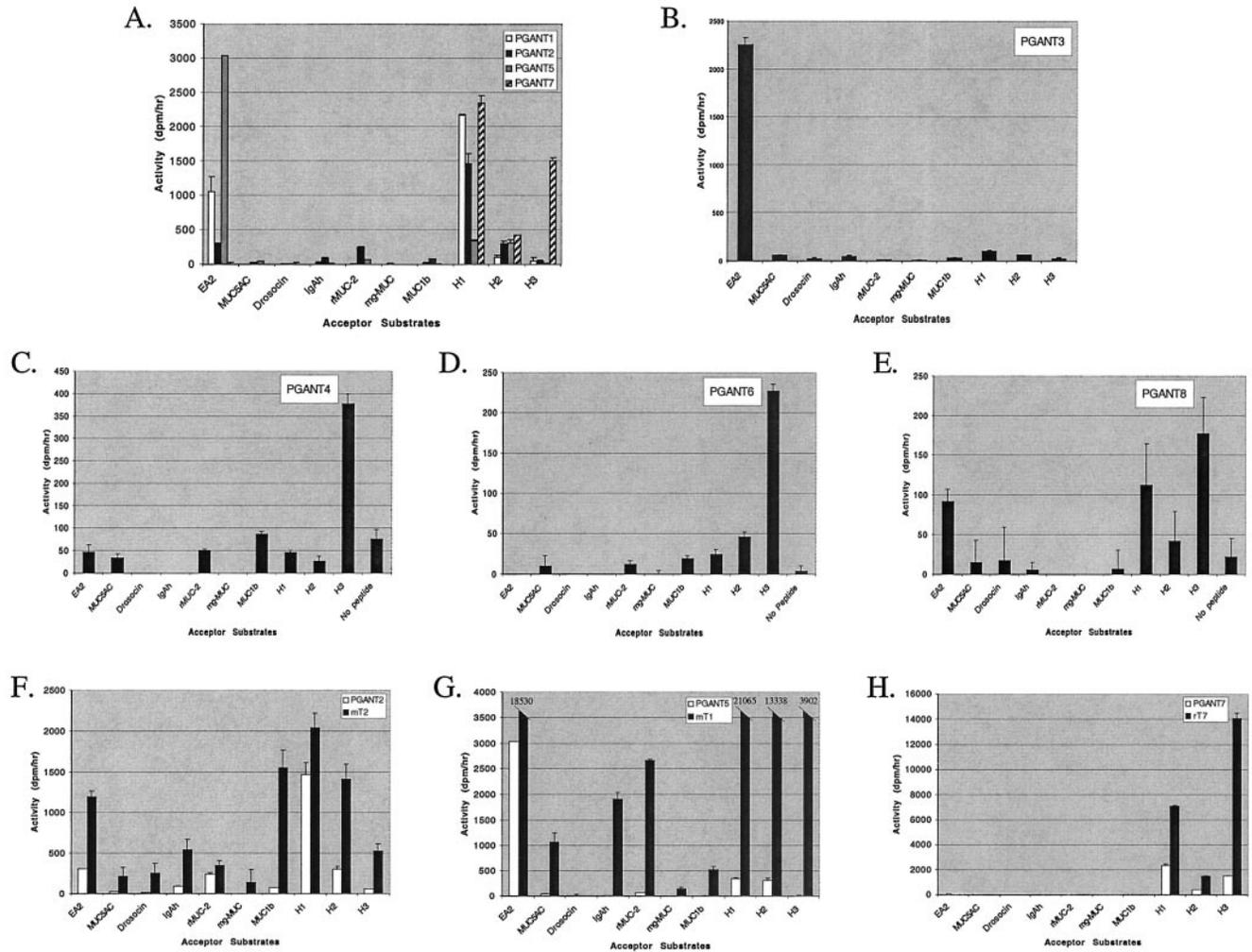


substrates; however, ppGaNTase-T1 has a broader substrate specificity, acting on a larger number of substrates (Fig. 3G). Analysis of the sites of addition of PGANT5 and ppGaNTase-mT1 on the EA2 substrate revealed glycosylation by both enzymes at threonines 7, 11, and 12, demonstrating similar preferences 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, preferring 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



**FIG. 3. Initial rates of recombinant PGANT proteins and mammalian orthologues against a panel of peptide and glycopeptide substrates.** Vertical axes indicate rates of transfer expressed as dpm/h. Acceptor substrates are shown along the *x* axis in each panel. Comparison of activity for PGANT1, -2, -5, and -7 (A) and PGANT6 (D) using equivalent relative amounts of FLAG-enriched recombinant proteins as described under “Experimental Procedures.” B, activity of PGANT3 using cell culture media. C, activity of FLAG-enriched recombinant PGANT4 (not quantitated relative to PGANT1, -2, and -5–7). E, activity of PGANT8 using cell culture media. Assays using recombinant protein directly from the cell culture media are shown for PGANT3 and PGANT8 as FLAG enrichment resulted in loss of activity. Comparison of activity of orthologous pairs composed of PGANT2 and mammalian ppGaNTase-T2 (F); PGANT5 and mammalian ppGaNTase-T1 (G); and PGANT7 and mammalian ppGaNTase-T7 (H) are shown in the bottom panels. Values for T1 with EA2, H1, H2, and H3 are shown numerically at the top of each respective bar. Error bars indicate standard deviations. Background values (No peptide) are shown in C–E where isoform activity is low. H1, MUC5AC-3; H2, MUC5AC-13; H3, MUC5AC-13.

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

TABLE I  
Summary of *Drosophila* UDP-GalNAc:polypeptide *N*-acetylgalactosaminyltransferases

pgant = polypeptide GalNAc transferase. CG numbers only are given for those putative isoforms not yet confirmed biochemically.

CG no.	Gene designation	Protein designation	Coding region	Location	Biochemical activity	Mammalian orthologue
CG8182	<i>pgant1</i>	PGANT1	<i>aa</i> 601	2R; 52A6	Peptide/glycopeptide transferase	
CG3254	<i>pgant2</i>	PGANT2	615	2L; 23F1	Peptide/glycopeptide transferase	ppGaNtase-T2
CG4445	<i>pgant3</i>	PGANT3	667	2R; 42E7–42F1	Peptide/glycopeptide transferase	
CG31956	<i>pgant4</i>	PGANT4	659	2L; 23F5	Glycopeptide transferase	
CG31651	<i>pgant5</i>	PGANT5	617	2L; 25C4-C6	Peptide/glycopeptide transferase	ppGaNtase-T1/T13
CG2103	<i>pgant6</i>	PGANT6	666	3L; 62F4	Glycopeptide transferase	
CG6394	<i>pgant7<sup>a</sup></i>	PGANT7	591	X; 17B6	Glycopeptide transferase	ppGaNtase-T7
CG7297	<i>pgant8</i>	PGANT8	590	3L; 71F1	Peptide/glycopeptide transferase	
CG7480	<i>pgant35<sup>b</sup></i>	PGANT35A	632	2L; 35A4	Peptide/glycopeptide transferase	ppGaNtase-T11
CG30463	NA <sup>c</sup>	NA	650 <sup>d</sup>	2R; 53C11-C14	NA	
CG10000	NA	NA	558 <sup>d</sup>	3R; 98E3	NA	
CG7304	NA	NA	563 <sup>d</sup>	3L; 71F1	NA	
CG31776	NA	NA	601 <sup>d</sup>	2L; 23F3	NA	
CG7579	NA	NA	498 <sup>d</sup>	3L; 71F1	NA	

<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.

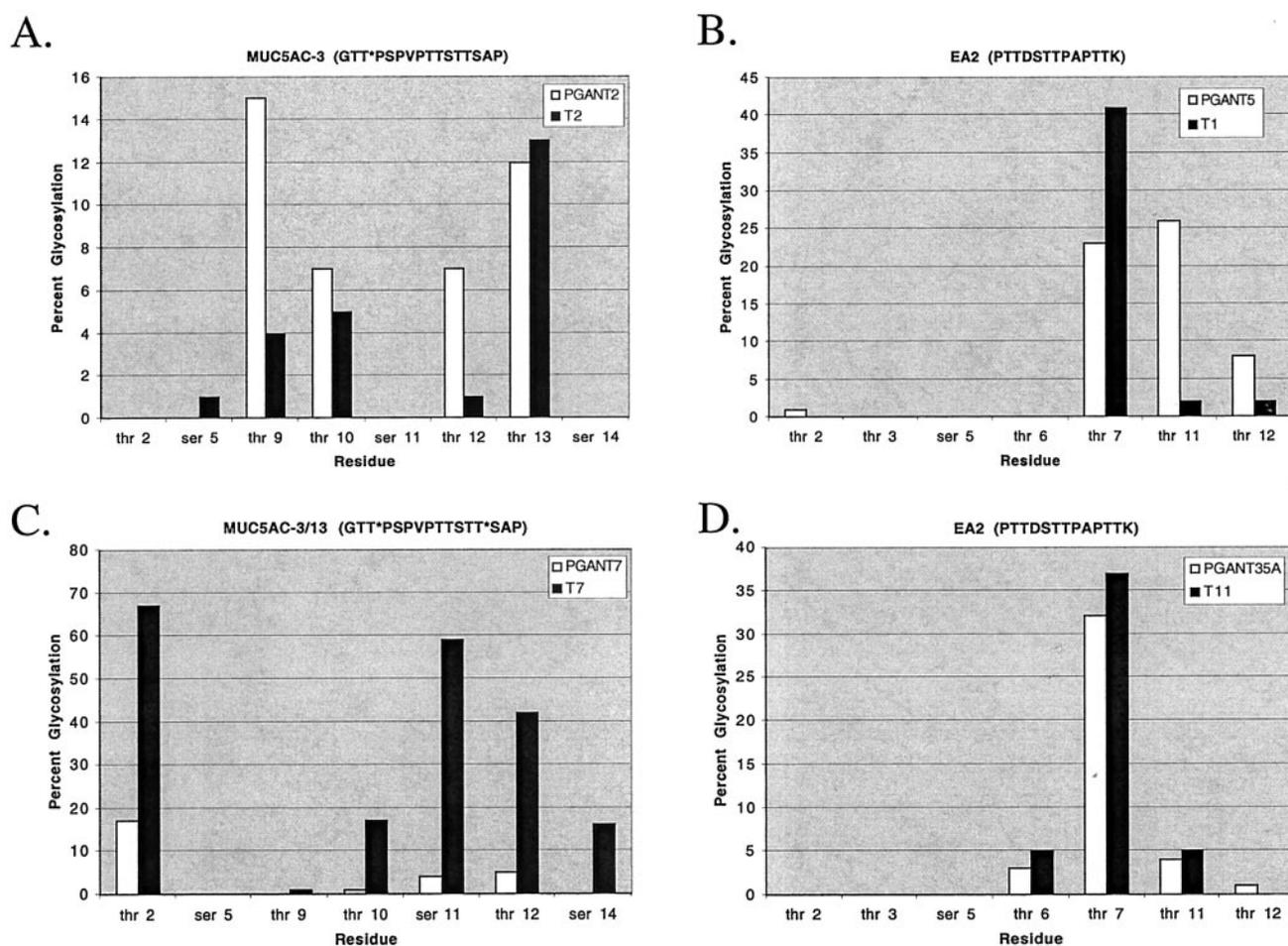


FIG. 4. Site of GalNAc addition by orthologous pairs on selected substrates. Sites of GalNAc addition were mapped by Edman degradation as described under "Experimental Procedures." Potential glycosylation sites are shown on the horizontal axis and percent glycosylation of each threonine (Thr) and serine (Ser) residue is shown on the vertical axis. The peptide sequence of substrates examined is shown at the top of each panel. T\* indicates a threonine previously modified with GalNAc. 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-T11 with EA2.

panel) to demonstrate the presence of cDNA in each sample well.

*In Situ Hybridization Analysis of Drosophila Embryos and Egg Chambers*—To obtain insights into the function of the various *Drosophila* UDP-GalNAc:polypeptide *N*-acetylgalac-

tosaminyltransferases 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

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 embryogen-

esis. *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.

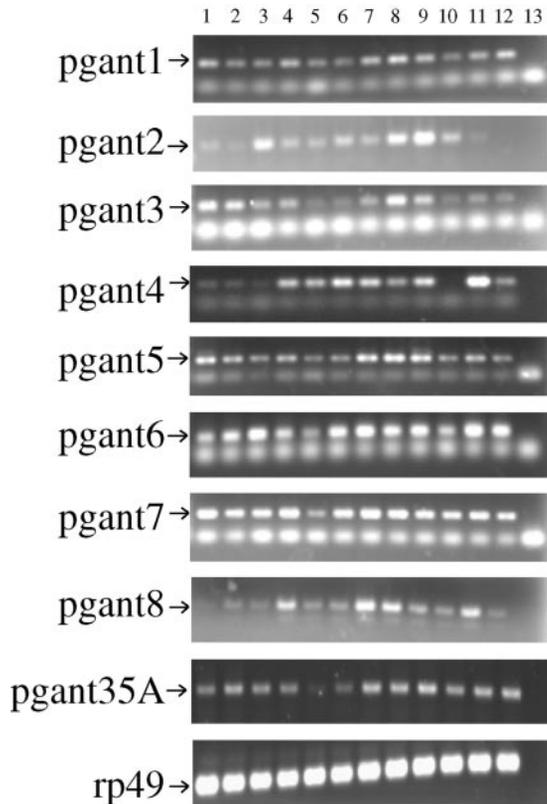


FIG. 5. PCR panels of *pgant* gene expression. cDNA panels from embryonic, larval, pupal, and adult stages of *D. melanogaster* (Origene Technologies) were PCR-amplified to detect the expression of the *pgant* genes. Lanes are as follows: lane 1, 0–4-h embryo; lane 2, 4–8-h embryo; lane 3, 8–12-h embryo; lane 4, 12–24-h embryo; lane 5, 1st instar larvae; lane 6, 2nd instar larvae; lane 7, 3rd instar larvae; lane 8, pupae; 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.

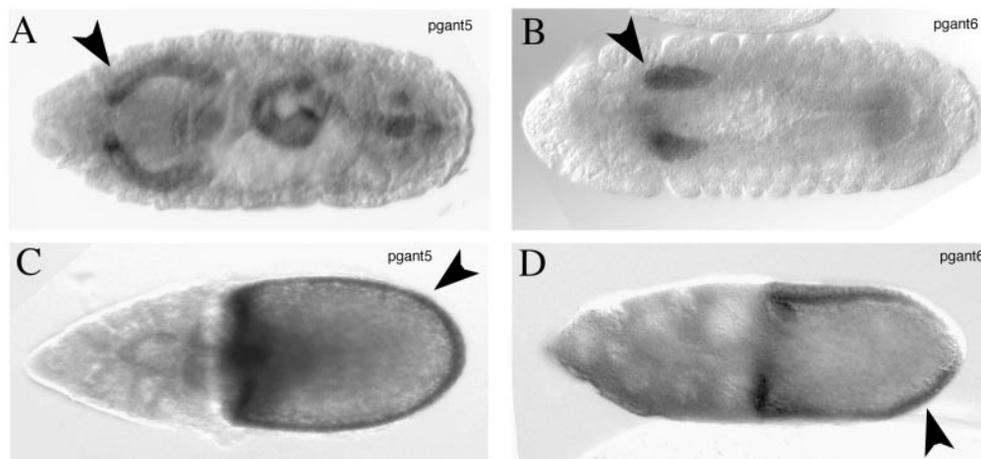


FIG. 6. *In situ* hybridization of *pgant5* and *pgant6* to *Drosophila* embryos and egg chambers. *pgant5* (A and C) and *pgant6* (B and D) were hybridized to embryos (A and B) as well as egg chambers (C and D). The arrows in A and B show expression of *pgant5* and *pgant6* in developing salivary glands. 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.

## 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 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 ex-

pression, pgant5 and pgant6 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 (Sgs3 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 vitellin 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

1. Ten Hagen, K. G., Fritz, T. A. & Tabak, L. A. (2003) *Glycobiology* **13**, R1–R16
2. Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G. & Marth, J. (1999) *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
3. Varki, A. (1998) *Trends Cell Biol.* **8**, 34–40
4. Ten Hagen, K. G., Bedi, G. S., Tetaert, D., Kingsley, P. D., Hagen, F. K., Balys, M. M., Beres, T. M., Degand, P. & Tabak, L. A. (2001) *J. Biol. Chem.* **276**, 17395–17404
5. Hagen, F. K. & Nehrke, K. (1998) *J. Biol. Chem.* **273**, 8268–8277
6. Ten Hagen, K. G. & Tran, D. T. (2002) *J. Biol. Chem.* **277**, 22616–22622
7. Schwientek, T., Bennett, E. P., Flores, C., Thacker, J., Hollmann, M., Reis, C. A., Behrens, J., Mandel, U., Keck, B., Schafer, M. A., Haselmann, K., Zubarev, R., Roepstorff, P., Burchell, J. M., Taylor-Papadimitriou, J., Hollingsworth, M. A. & Clausen, H. (2002) *J. Biol. Chem.* **277**, 22623–22638
8. Westerman, E. L., Ellies, L. G., Hagen, F. K., Marek, K. W., Sutton-Smith, M., Dell, A., Tabak, L. A. & Marth, J. D. (1999) *Glycobiology* **9**, 1121
9. Hennet, T., Hagen, F. K., Tabak, L. A. & Marth, J. D. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 12070–12074
10. Zhang, Y., Iwasaki, H., Wang, H., Kudo, T., Kalka, T. B., Hennet, T., Kubota, T., Cheng, L., Inaba, N., Gotoh, M., Togayachi, A., Guo, J., Hisatomi, H., Nakajima, K., Nishihara, S., Nakamura, M., Marth, J. D. & Narimatsu, H. (2003) *J. Biol. Chem.* **278**, 573–584
11. Hagen, F. K., VanWuyckhuysse, B. & Tabak, L. A. (1993) *J. Biol. Chem.* **268**, 18960–18965
12. White, T., Bennett, E. P., Takio, K., Sørensen, T., Bonding, N. & Clauser, H. (1995) *J. Biol. Chem.* **270**, 24156–24165
13. Zara, J., Hagen, F. K., Ten Hagen, K. G., VanWuyckhuysse, B. C. & Tabak, L. A. (1996) *Biochem. Biophys. Res. Commun.* **228**, 38–44
14. Hagen, F. K., Ten Hagen, K. G., Beres, T. M., Balys, M. M., VanWuyckhuysse, B. C. & Tabak, L. A. (1997) *J. Biol. Chem.* **272**, 13843–13848
15. Ten Hagen, K. G., Hagen, F. K., Balys, M. M., Beres, T. M., VanWuyckhuysse, B. C. & Tabak, L. A. (1998) *J. Biol. Chem.* **273**, 27749–27754
16. Bennett, E. P., Hassan, H., Mandel, U., Hollingsworth, M. A., Akisawa, N., Ikematsu, Y., Merckx, G., van Kessel, A. G., Olofsson, S. & Clausen, H. (1999) *J. Biol. Chem.* **274**, 25362–25370
17. Ten Hagen, K. G., Tetaert, D., Hagen, F. K., Richet, C., Beres, T. M., Gagnon, J., Balys, M. M., VanWuyckhuysse, B., Bedi, G. S., Degand, P. & Tabak, L. A. (1999) *J. Biol. Chem.* **274**, 27867–27874
18. White, K. E., Lorenz, B., Evans, W. E., Meitinger, T., Strom, T. M. & Econs, M. J. (2000) *Gene (Amst.)* **246**, 347–356
19. Toba, S., Tenno, M., Konishi, M., Mikami, T., Itoh, N. & Kurosaka, A. (2000) *Biochim. Biophys. Acta* **1493**, 264–268
20. Guo, J. M., Zhang, Y., Cheng, L., Iwasaki, H., Wang, H., Kubota, T., Tachibana, K. & Narimatsu, H. (2002) *FEBS Lett.* **524**, 211–218
21. Wang, H., Tachibana, K., Zhang, Y., Iwasaki, H., Kameyama, A., Cheng, L., Guo, J., Hiruma, T., Togayachi, A., Kudo, T., Kikuchi, N. & Narimatsu, H. (2003) *Biochem. Biophys. Res. Commun.* **300**, 738–744

22. Hagen, F. K., Hazes, B., Raffo, R., deSa, D. & Tabak, L. A. (1999) *J. Biol. Chem.* **274**, 6797–6803
23. Albone, E. F., Hagen, F. K., VanWuyckhuysse, B. C. & Tabak, L. A. (1994) *J. Biol. Chem.* **269**, 16845–16852
24. Gendler, S. J., Lancaster, C. A., Taylor-Papadimitriou, J., Duhig, T., Peat, N., Burchell, J., Pemberton, L., Lalani, E.-N. & Wilson, D. (1990) *J. Biol. Chem.* **265**, 15268–15293
25. Ohmori, H., Dohrman, A. F., Gallup, M., Tsuda, T., Kai, H., Gum, J. R., Jr., Kim, Y. S. & Basbaum, C. B. (1994) *J. Biol. Chem.* **269**, 17833–17840
26. Shekels, L. L., Lyftogt, C., Kieliszewski, M., Filie, J. D., Kozak, C. A. & Ho, S. B. (1995) *Biochem. J.* **311**, 775–785
27. Guyonnet-Dupérat, V., Audié, J. P., Debailleul, V., Laine, A., Buisine, M. P., Galiege-Zouitina, S., Pigny, P., Degand, P., Aubert, J. P. & Porchet, N. (1995) *Biochem. J.* **305**, 211–219
28. Hiki, Y., Odani, H., Takahashi, M., Yasuda, Y., Nishimoto, A., Iwase, H., Shinzato, T., Kobayashi, Y. & Maeda, K. (2001) *Kidney Int.* **59**, 1077–1085
29. McManus, A. M., Otvos, L., Jr., Hoffman, R. & Craik, D. L. (1999) *Biochemistry* **38**, 705–714
30. Gerken, T. A., Gilmore, M. & Zhang, J. (2002) *J. Biol. Chem.* **277**, 7736–7751
31. Tautz, D. & Pfeifle, C. (1989) *Chromosoma* **98**, 81–85
32. Campos-Ortega, J. A. & Hartenstein, V. (1985) *The Embryonic Development of Drosophila Melanogaster*, Springer-Verlag, Berlin
33. King, R. C. (1970) *Ovarian Development in Drosophila melanogaster*, Academic Press, New York
34. Venter, J. C., *et al.* (2001) *Science* **291**, 1304–1351
35. Rubin, G. M., *et al.* (2000) *Science* **287**, 2204–2215
36. Andrew, D. J., Henderson, K. D. & Seshiah, P. (2000) *Mech. Dev.* **92**, 5–17
37. Ray, R. P. & Schupbach, T. (1996) *Genes Dev.* **10**, 1711–1723
38. Theopold, U., Samakovlis, C., Erdjument-Bromage, H., Dillon, N., Axelsson, B., Schmidt, O., Tempst, P. & Hultmark, D. (1996) *J. Biol. Chem.* **271**, 12708–12715