Biotransformation enzymes have been found in the olfactory epithelium of vertebrates. We now show that in Drosophila melanogaster, a UDP-glucosyltransferase (UGT), as well as a short chain dehydrogenase/reductase and a cytochrome P450 are expressed specifically or preferentially in the olfactory organs, the antennae. The evolutionarily conserved expression of biotransformation enzymes in olfactory organs suggests that they play an important role in olfaction. In addition, we describe five Drosophila UGTs belonging to two families. All five UGTs contain a putative transmembrane domain at their C terminus as is the case for vertebrate UGTs where it is required for enzymatic activity. The primary sequence of the C terminus, including part of the transmembrane domain, differs between the two families but is highly conserved not only within each Drosophila family, but also between the members of one of the Drosophila families and vertebrate UGTs. The partial overlap of the conserved primary sequence with the transmembrane domain suggests that this part of the protein is involved in specific interactions occurring at the membrane surface. The presence of different C termini in the two Drosophila families suggests that they interact with different targets, one of which is conserved between Drosophila and vertebrates.

All organisms live in environments that contain potentially harmful chemicals, both natural and man-made. Extensive studies of detoxification in the vertebrate liver provide a framework to the study of detoxification mechanisms in other systems (1–3). Detoxification often occurs in two phases; in phase I, the initial compound is transformed into a more reactive species. A variety of different chemical transformations are involved, including redox reactions catalyzed by enzymes of the cytochrome P450 superfamily (4, 5) and members of the short chain dehydrogenase/reductase (SDR) family (6). Phase II reactions consist in the addition, either to a product of a phase I reaction or directly to many toxic chemicals, of a highly polar group such as UDP-glucuronosyl (catalyzed by UDP-glucuronosyltransferases) (2) or glutathione (catalyzed by glutathione S-transferases (7)). Products of phase II reactions are hydrophilic; they can no longer cross membranes and are eliminated by secretion. In addition to the elimination of environmental toxins, phase I and II biotransformation enzymes participate in the removal of toxic side products of normal metabolism (e.g. bile acids), participate in drug clearance, and play an important role in the synthesis of hormones such as prostaglandins and some steroids (3). Finally, the involvement of these enzymes in the production of carcinogens (8), drug clearance, and some hereditary diseases (9) makes an understanding of their function important for human health.

Biotransformation enzymes related to those found in vertebrates have also been found in insects and are likely to play equally important roles. Cytochrome P450s and glutathione S-transferases in particular have been implicated in insect resistance to pesticides (10). UDP-glucuronosyltransferases are part of a superfamily of UDP-glucosyltransferases (UGTs) present in plants, animals, and bacteria (11). These enzymes transfer the sugar moiety of UDP-glucose, UDP-glucuronic acid, UDP-galactose, or UDP-xylose to a variety of hydrophobic substrates (11). Insects contain UGT activities that can use UDP-glucose but not UDP-glucuronic acid as a glycosyl donor (12–14). Although no molecular information on any insect UGT was available until this work, baculoviruses infecting several species of moths have been shown to encode ecdysteroid UDP-glucosyltransferases (15, 16). These viral enzymes specifically inactivate ecdysteroids, the molting hormones of the infected hosts, and thus prolong the larval stage permissive to viral replication. Ecdyysteroid UDP-glucosyltransferases lack a C-terminal transmembrane domain and are secreted in the hemolymph where ec dysone is present (15).

Here we report that, in Drosophila, several phase I and II biotransformation enzymes are expressed preferentially in the olfactory organs, the antennae. This observation is reminiscent of the preferential or exclusive expression of a cytochrome P450, UGT, and glutathione S-transferase in the vertebrate olfactory epithelium (17–20). The presence of these enzymes in the olfactory organs of such evolutionarily distant organisms supports the notion that they play an important role in olfaction. In addition, the availability of the first sequences of UGTs from insects sheds light on the structure and function of the C-terminal domain of vertebrate UGTs.

**EXPERIMENTAL PROCEDURES**

Generation of an Appendage cDNA Library, Cloning, and Sequencing—Partial cDNA clones for AntP450, AntDH, and DmeUGt35b were initially found through random sequencing of clones in an antennae-minus-heads subtracted cDNA library that was described previously (21). All cDNA sequences discussed in this paper were obtained from
full-length cDNA clones isolated by using the partial clones as probes to screen an appendage cDNA library in Lambda-ZAP (Stratagene). Appendage RNA was generated from poly(A)+ RNA isolated from an appendage fraction (see below and Ref. 22).

**Analysis of Gene Expression—**Total RNA for Northern blots was isolated either from hand-dissected antennae or legs or from mass-produced body fractions generated as follows (22). Frozen flies were vortexed, and the resulting body parts are then sieved to yield three fractions: appendages (antennae, legs, and wings), heads (without antennae), and bodies (abdomen and thorax, decapitated and without legs or wings). Because all the proteins under study belong to multigene families, the probes used were first tested on Southern blots under identical conditions to ensure that there was no detectable cross-reactivity to related genes (not shown). The probes used are indicated in the legend to Fig. 1.

**Sequence Analysis—**Sequences were assembled and analyzed using Wisconsin Package Version 9.1, Genetics Computer Group, Madison, WI. Data base searches were performed using BLAST (23) both on the Berkeley Drosophila Genome Project and National Center for Biotechnology Information www servers (http://www.ncbi.nlm.nih.gov/BLAST/nph-newblast?Jform=0, respectively).

The central portion of the DmeUgt35a cDNA is represented by nucleotides number 1–324 of P1 clone DS07339 that has been mapped by the use of three open reading frames (ORFs) present at nucleotides number 68193 to 66570 (reverse strand) of DS075875 mapped to 86D1-D2. AntP450 is a partial cDNA sequence identical to EST number GH06928 and with significant sequence similarity to the C terminus of the cyp6 family of cytochrome P450s (24).

The three members of the DmeUgt37 family result from conceptual translation of sequences from the Drosophila Genome Project. The sequences encoding the three open reading frames (ORFs) are present at nucleotides number 68193 to 66570 (reverse strand) of P1 clone DS51087 for DmeUgt37a1; positions 22203–22207 of DS00108 for DmeUgt37b1 and nucleotides 111614 to 111388 of DS07321 for DmeUgt37c1. To generate the DmeUgt37a1 ORF, we removed a likely intron at positions 67516–67464. In the case of DmeUgt37b1, deletion of a stretch of 5 Ts at positions 22203–22207 results in the creation of an ORF with high similarity to the other Drosophila UGT37 ORFs (see text). The frameshift in the database sequence could be caused by a sequencing error or a recent mutation resulting in a pseudogene. In either case, conceptual translation of the “corrected” DmeUgt37b1 sequence represents a UGT with high similarity to the other UGT37 proteins throughout its open reading frame and therefore likely represents a real UGT, even if it no longer exists in present day laboratory canton S strains. For the purpose of this publication we will keep the asterisk to denote the ambiguity. Note that the inclusion of the DmeUgt37b1 ORF is not necessary to reach the conclusions about the domain structure of *Drosophila* UGTs that are discussed in the text.

The sequences of the three novel cDNAs discussed here have been deposited in the GenBank™ data base and their accession numbers are as follows: AntDH, AFI55535, DmeUgt35a, AF116554; and DmeUgt35b, AF116554. Non-Drosophila UGTs are designated according to the names given by the UGT Nomenclature Committee and accession numbers are given in the figures.

**RESULTS**

**Several Biodegradation Enzymes Are Expressed Preferentially in the Antennae of Drosophila**—We have previously described a subtracted cDNA library (antennae-minus-heads) enriched in cDNAs expressed specifically or preferentially in the antennae of *Drosophila melanogaster*. Analysis of a number of those cDNAs led to the discovery of several putative odorant-binding proteins with distinct expression patterns on the surface of the antennae, suggesting a role for odorant-binding proteins in olfactory discrimination (21). Here we report that, in addition to odorant-binding proteins, sampling of the antennal cDNA library has led to the discovery of cDNAs coding for a cytochrome P450 (AntP450, see “Experimental Procedures”), a UGT (DmeUgt35b, see below for explanation of the nomenclature), and a short chain dehydrogenase/reductase (AntDH). The proteins encoded by these three cDNAs are related to vertebrate enzymes involved in detoxification. In addition to DmeUgt35b isolated from our subtractive library, a second UGT-encoding cDNA (DmeUgt35a) was isolated from an appendage cDNA library by cross-hybridization to DmeUgt35b (see below).

Analysis of expression patterns was performed by Northern blots using probes specific for DmeUgt35a, DmeUgt35b, AntDH, and AntP450. In every case the probes were generated from sequences that show little similarity with other genes of the same family and the lack of cross-hybridization was verified on Southern blots performed under identical hybridization conditions (data not shown). Expression of the ubiquitously expressed rp49 gene (42) was monitored in all fractions as a loading control. 1 µg of total RNA was loaded in each lane. Probes used for the two UGTs were 5′ cDNA fragments of 560 base pairs (EcoRI-NheI) and 610 base pairs (EcoRI-NruI) for DmeUgt35a and DmeUgt35b, respectively. The probe used for AntDH was the full-length cDNA clone and that for AntP450 was the partial cDNA clone obtained in the subtracted library.

**FIG. 1.** Several novel biodegradation enzymes are preferentially expressed in *Drosophila antennae*. Northern blots were performed with RNA extracted from different parts of the fly (see “Experimental Procedures”) as indicated above each lane. Appendages: legs, third antennal segments, and wings; bodies: decapitated bodies without legs or wings; heads: heads without third antennal segments. RNA from third antennal segments and legs was obtained after manual dissection. To ensure that each signal corresponds to expression from a single gene, 32P probes were generated from relatively nonconserved regions of each gene that give rise to a single band on genomic Southern blots under identical hybridization conditions (data not shown). Expression of the ubiquitously expressed rp49 gene (42) was monitored in all fractions as a loading control. 1 µg of total RNA was loaded in each lane. Probes used for the two UGTs were 5′ cDNA fragments of 560 base pairs (EcoRI-NheI) and 610 base pairs (EcoRI-NruI) for DmeUgt35a and DmeUgt35b, respectively. The probe used for AntDH was the full-length cDNA clone and that for AntP450 was the partial cDNA clone obtained in the subtracted library.
members of this family are microsomal or even extracellular (30). Contrary to the case of the membrane-associated mouse corticosteroid 11-β-dehydrogenase (dh11_mouse in Fig. 3) there is no apparent amino-terminal signal sequence in the AntDH ORF. Because the short sequence preceding the apparent translational start in our AntDH clone does not contain any stop codon, we cannot entirely rule out the possibility that we are missing some 5’ sequences that code for a signal peptide. However, the presence of an AUG at the almost identical position as it is found in many cytoplasmic SDRs (Fig. 3) suggests that we have identified the correct amino terminus and that AntDH is a cytoplasmic protein.

Identification of Nine Putative Drosophila UGTs—When probing the appendage library with our partial Ugt clone we found two classes of clones that hybridize at different intensities. Southern blotting and sequence analysis shows that these phages correspond to two different cDNAs each encoded by a separate gene, which we will call DmeUgt35a and -b (see the last paragraph under “Results” for a justification of this nomenclature). We have mapped both sequences by in situ hybridization to cytogenetic locations 86C-D in the Drosophila genome, suggesting that these two genes have their origin in a relatively recent duplication. More recently, the Berkeley and European Drosophila genome projects have sequenced parts of both DmeUgt35 genes, refining the mapping to 86D5-10 and 86D1-2 for the a and b genes, respectively.

We have also found that several other likely Ugt sequences are present in the Drosophila genome project data. These include three genomic DNA sequences containing full-length ORFs that define three members of a second family of Drosophila UGTs, UGT37: DmeUgt37a1, -b1 and -c1 (see below for an explanation of the nomenclature). Five other likely Ugt genes are represented by partial cDNA sequences or ESTs (EST numbers GH06505, GH09393, GM04645, LD25345, and LD15335). The first four are 5’ sequences coding for NH2 termini (Fig. 4A), whereas the last one is a 3’ sequence coding for a C terminus (Fig. 5). In all, we describe nine or ten putative Drosophila UGTs (because each of the above ESTs has only been sequenced from one end, LD15335 may be identical to one of the other clones). The two DmeUgt35 cDNAs as well as the three genomic DmeUgt37 sequences appear to represent complete ORFs because they begin with ATG codons and end with stop codons at positions that match closely those expected for this family of genes (see “Experimental Procedures” for further discussion of DNA sequence analysis). In contrast to the use of alternative splicing for the generation of diversity, as is the case of the human UGT1A1 gene (2), we have found no evidence of alternative mRNA splicing, and an intron is present in only one of the three genomic sequences (see “Experimental Procedures”).

Some of the highest similarity between the five complete ORFs is found near a sequence present in all known UGTs and defined by the string: [FVA]-[LIVMF]-[TS]-[HQ]-[SGAC]-[LMVFIQ](2)-[DE]-Q, in which all amino acids that can occur at a given position are listed inside brackets (1). The presence of this sequence strongly supports the identification of these five proteins as UGTs (Fig. 4B). In addition, the five complete ORFs contain C-terminal hydrophobic domains followed by several basic residues (see below and Fig. 5). In the case of vertebrate UGTs, similar sequences have been identified as a transmembrane domain and a positively charged “stop-transfer” sequence that in combination are responsible for the anchoring of the protein to the endoplasmic reticulum membrane (1) and are necessary for enzymatic activity (31). Baculovirus ecdysteroid UDP-glucosyltransferases, which are solu-

Fig. 2. AntDH is specifically expressed in the third antennal segment. Horizontal cryosections of heads were hybridized with digoxigenin-labeled DNA probes that were visualized using standard experimental procedures with anti-digoxigenin antibodies conjugated to alkaline phosphatase (21). In the presence of a chromogenic substrate, a blue/purple precipitate is formed. The section shown is at the level of the third antennal segment and is typical of many others. In no case was signal observed in other parts of the head or in the second antennal segment (data not shown).

by this method, most likely because of its higher expression levels (data not shown). Consistent with the Northern blot analysis, in situ hybridization to AntDH mRNA is restricted to the third antennal segment; no expression is detected in the head (Fig. 2) or in the second antennal segment (not shown). This observation further supports a role for AntDH in olfaction, because the second antennal segment does not contain chemosensory hairs (25). Within the third antennal segment however, AntDH expression appears uniformly distributed (Fig. 2 and data not shown), in contrast with several odorant-binding proteins, each of which is restricted to a single morphological type of sensillum with a nonuniform distribution on the antennal surface (21, 26, 27).

AntDH Is a Short Chain Dehydrogenase/Reductase Specifically Expressed in Third Antennal Segments—The initial AntDH cDNA clone was used as a probe to isolate a full-length cDNA clone from an appendage cDNA library (see “Experimental Procedures”). A single ORF has sequence similarity to the cDNA clone from an appendage cDNA library (see “Experimental Procedures”). A single ORF has sequence similarity to the

ble and secreted in the hemolymph of the hosts lack such C-terminal transmembrane domains (15).

Finally, at least nine of the ten putative Drosophila UGTs display a region of high similarity at their very amino terminus, immediately following the signal peptides (32, 33) (Fig. 4A). In the case of the tenth putative UGT: LD15335 only the C-terminal sequence is presently known. In vertebrate UGTs this region of the molecule is involved in the formation of dimers (34), which may be the active form of the protein. In at least one case a heterodimer has enzymatic activities that differ from those of either homodimer (35), suggesting that the combinatorial association of different subunits into heterodimers may provide added functional diversity.

UGT35a and -b Contain C-terminal Transmembrane Domains Similar to Those of Vertebrate UGTs but Different from Those of the Drosophila UGT37 Protein Family—In consultation with the UGT nomenclature committee (11), we have assigned the two cDNAs we have cloned to a single family, Ugt35. The other three full-length sequences found through the genome projects fall into a second family, DmeUgt37. Both families fit the commonly accepted criteria for protein families (more than 45% overall identity within a family and less than 45% between different families, data not shown). Finally, although the C-terminal sequence of LD15335 suggests that it is a member of the DmeUgt37 family (Fig. 5), the definitive assignment of the five UGTs presently only known as ESTs (Fig. 4A) to one of these two families, or yet new ones, will require their complete sequences.

Amino acid residues present in all nine Drosophila proteins occur in the first sixty residues (Fig. 4A) as well as in the C-terminal half of the protein, particularly around the signature sequence (Fig. 4B). After the first sixty amino acids, the amino-terminal halves of the proteins are highly divergent (not shown), as is the case for vertebrate UGTs, perhaps corresponding to different substrate specificities.

Strikingly, although the C-terminal halves of all the Drosophila proteins are closely related, there is a strong discontinuity of this similarity at their very C termini. After a highly conserved segment, the sequences of the two families diverge abruptly, each encoding a different C-terminal domain containing putative transmembrane stretches and stop-transfer sequences (Fig. 5). Within each family, however, there is a high degree of sequence conservation. Interestingly, the region of sequence similarity overlaps with the likely transmembrane helix for members of the UGT35 (36) as well as UGT37 families (Fig. 5). Five of the first six amino acids in the putative transmembrane domain are identical between Drosophila UGT35b and human UGT1A1, and conservation of similarly located residues is apparent for the UGT37 family. This pattern of conservation suggests specific and different roles for the C-terminal domains of the two Drosophila families.

**DISCUSSION**

The Evolutionarily Conserved Presence of Biodegradation Enzymes Argues for an Important Function in Olfaction—The results presented here suggest that, as in the olfactory epithelia of vertebrates (17–20), several biodegradation enzymes are expressed specifically or at higher levels in the antennae of Drosophila. This conserved expression of biodegradation enzymes in olfactory organs argues for an important function in olfaction. Such a role is also consistent with the presence in the antennae of an enzyme involved in cytochrome P450 activation, NADPH P450 oxidoreductase (37).
are able to find females located many miles away by rapidly alternating between two types of behavior, upwind flight when inside a pheromone plume and casting from side to side as soon as the pheromone is no longer detected (38). The ability to monitor concentration changes without a lag requires that the half-life of odorants inside the olfactory organs be short relative to the time course of the outside fluctuations. Based on these considerations, researchers have looked for and found enzymes that can specifically metabolize pheromones in the antennae of several species of moths (39, 40). In vertebrate olfaction, a similar role has been attributed to biotransformation enzymes that are better known for their role in detoxification in the liver. The same sequence of events that eliminates toxic chemicals may be involved in odorant degradation, thereby preventing continuing stimulation of olfactory receptors. Consistent with this hypothesis, an olfactory-specific UGT, UGTolf, modifies odorants more efficiently than liver UGTs (18).

A second function for these enzymes might be in the protection of olfactory organs from environmental toxins to which they are, by necessity, preferentially exposed. Although odorant turnover and toxin degradation are not mutually exclusive functions and any given protein may be involved in both, the expression pattern of each gene may suggest the relative contribution to either function. Because detoxification occurs in many organs, proteins whose expression is highly specific to the antennae, such as AntDH and UGT35b, are likely to be involved in odorant turnover. On the other hand, proteins that have more ubiquitous expression patterns, such as UGT35a, may participate primarily in detoxification.

Parallel Conservation of Different Primary Sequences Suggests Different Functions for the C-terminal Domains of the Two Drosophila UGT Families—We report here that the Drosophila genome encodes at least nine different UGTs. The five genes for which complete coding sequences are available contain the signature motif characteristic of this superfamily and thus represent the first reported UGTs from any insect other than ecdysteroid UDP-glucosyltransferases from baculoviruses (11). In addition to the overall similarity to UGTs in a variety of organisms, both Drosophila and vertebrate proteins have at their C terminus a transmembrane domain followed by a stop-transfer sequence composed of several positively charged amino acids (1, 36). Although this domain is absent from viral-encoded ecdysteroid UDP-glucose transferases and some plant UGTs (11), mutations in either the transmembrane domain or the stop-transfer sequence of vertebrate UGT2B1 eliminate and reduce its activity, respectively (31).
domains start immediately after the LD sequence (data not shown). Hydropathy was calculated by the Kyte and Doolittle method using the respectively. The transmembrane domain indicated for members of the UGT35 family and its vertebrate relatives is the one proposed for vertebrate similarity scores. Using Pileup, but the sequences shown are contiguous and their order within the alignments has been preserved and does not correspond to arrowheads UGT37a1, -b1 and -c1. More surprising, however, is the high degree of primary sequence identity between the Drosophila members of the UGT35 family and vertebrate sequences in a stretch immediately NH2-terminal to and partially overlapping with the putative transmembrane domain (61 and 70% identity to the human UGT1A1 gene over a 31 amino acid stretch for the a and b genes, respectively). In addition, although the sequences of members of the UGT37 family are very different from those of the UGT35 family in this region, they are also highly conserved within this second family (Fig. 5). This parallel conservation of primary sequences suggests that the C-terminal domains of UGTs are involved in specific interactions at the membrane surface that differ between UGT35a, -b, and the vertebrate enzymes on one hand and the members of the UGT37 family on the other. Despite these differences, the domains of the two classes of proteins have some shared features. In all cases except for UGT37b1*, the two amino acids at positions 15 and 16 after the start of sequence divergence are LD, which are immediately followed by a series of hydrophobic residues likely to be part of the transmembrane domain (Ref. 36 and Fig. 5). In the case of UGT37b1*, a three amino acid insertion moves LD to positions 18 and 19 and the putative transmembrane domain starts at position 20. These similarities suggest that despite the divergent sequences the two different types of C-terminal domains have similar secondary structures and may therefore interact with related proteins. What is the function of these alternative C-terminal domains? The primary sequence conservation within each family suggests it may be involved in an interaction with another protein that occurs at least in part within the membrane. The enzymatic reactions catalyzed by UGTs take place in the lumen of the endoplasmic reticulum and are therefore dependent on specific transporters that allow entry of nucleotides into this subcellular compartment (41). One intriguing possibility is that the C-terminal domain of UGTs participates in interactions with specific transporters, perhaps corresponding to the specificities of these enzymes for different glycosyl donors. However, because permeabilization of membranes with detergent does not restore activity to proteins with mutations in the C-terminal domain (31), substrate transport cannot be its only function. Although scans of the existing data bases have not revealed any UGT from any organism with a C-terminal domain similar to that of the UGT37 family, the ongoing sequence of the human genome may yet uncover such genes. Alternatively, if this domain arose after the divergence of the ancestors of insects and vertebrates, it may constitute an insect-specific domain and therefore a possible target for rational pesticide design. The availability of the genes coding for all these enzymes in Drosophila should allow the test of their involvement in olfaction as well as a dissection of the function of the UGT C-terminal domains using both biochemical and reverse genetic approaches.

Acknowledgments—We thank Yashi Ahmed and Ren-Ping Zhou for helpful comments on the manuscript and express gratitude also to Michael Rosbash in whose laboratory this work was initiated.

REFERENCES
Biodegradation Enzymes in the Antennae of Drosophila

257–224


Preferential Expression of Biotransformation Enzymes in the Olfactory Organs of
Drosophila melanogaster, the Antennae
Qin Wang, Gaiiti Hasan and Claudio W. Pikielny

doi: 10.1074/jbc.274.15.10309

Access the most updated version of this article at http://www.jbc.org/content/274/15/10309

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 42 references, 13 of which can be accessed free at
http://www.jbc.org/content/274/15/10309.full.html#ref-list-1