Characterization of a Subfamily of Beetle Odorant-binding Proteins Found in Hemolymph*

Laurie A. Graham‡§, Dyanne Brewer¶, Gilles Lajoie¶, and Peter L. Davies‡

In insects, hydrophobic odorants are transported through the sensillar lymph to receptors on sensory neurons by odorant-binding proteins (OBPs). The beetle Tenebrio molitor, which is a pest of stored grain products, produces a set of 12-14 kDa OBP-like proteins in its hemolymph. The structure of one of these proteins and that of a moth pheromone-binding protein have been solved. Both proteins have at least six α-helices with an internal, hydrophobic, ligand-binding pocket, but the beetle OBP lacks one of the disulfide bonds immediately adjacent to this pocket. To explore this difference and to sample lacks one of the disulfide bonds immediately adjacent to hydrophobic, ligand-binding pocket, but the beetle OBP

Internal binding pocket. In insects, the lymph within the antennal sensilla contains odorant-binding proteins (OBPs),1 which are thought to deliver the odorants to receptors on the sensory dendrites. The first OBP described was the pheromone-binding protein (PBP) of the moth Antheraea polyphemus, which is abundant in the antennae of adult males (1). Subsequently, additional PBPs as well as OBPs that are thought to bind non-sexual odors, such as those released by food sources, have been identified in a large number of insect species (for review, see Ref. 2). High affinity binding has been demonstrated between a number of PBPs/ OBPs and putative ligands (for examples, see Refs. 3, 4, and 5).

Individual insect species produce numerous OBPs, which has become increasingly evident in part from genomic and cDNA sequencing projects. For example, as many as 38 isoforms have been identified in the fly Drosophila melanogaster with deduced molecular masses ranging from 11 to 24 kDa for the monomeric isoforms (6, 7, 8). The primary sequences of these and other insect OBPs are not usually well conserved, but most members share 6 Cys and a similar patterning of hydrophobic and hydrophilic residues that defines the helical regions. Different roles have been postulated for two D. melanogaster isoforms, LUSH and PBPRP-2. LUSH+ flies do not avoid high concentrations of ethanol, which suggests that LUSH has a direct role in sensory perception (9). In contrast, PBPRP-2 is only found in antennal lymph external to the sensory dendrites, which led to the suggestion that it may not be directly involved in sensory perception but instead may have a role in odorant clearance (10). However, little is known about the function of most OBPs.

It would appear that OBPs are not restricted to sensory roles as some are present in non-sensory tissues. For example, OBPs have been found in the hemolymph of the medfly Ceratitis capitata (11) and the mealworm beetle Tenebrio molitor (12). Additional examples include sericotropicin from the brain of the moth Galleria mellonella (13) and the two B proteins from the accessory sex gland of T. molitor (Tmo/B1 and Tmo/B2 (14)). Although clear functions have not been ascribed to these proteins, it is likely that they also bind small hydrophobic molecules.

The structures of two OBPs have been solved. The 12-kDa T. molitor hemolymph protein (THP12, now renamed THP12a) (15) and the PBP from the moth Bombyx mori (16, 17) are both hexahelical with two or three disulfide bonds linking adjacent helices. Although their primary sequences are only 11% identical, they are clearly homologous as the largely amphipathic helices of both proteins pack in a similar manner (1.6 Å root mean square deviation between backbone atoms) to form a
cavity lined with hydrophobic residues (8). The crystal structure revealed that the *B. mori* PBP (*BmorPBP*) binds the sex pheromone bombykol in this cavity. However, there are interesting differences as well. NMR analysis showed that at low pH, the significantly longer C-terminal region of *BmorPBP* forms an additional helix that can enter this cavity (17). This might serve as a possible mechanism to displace the ligand.

Structures are available for other proteins that are thought to have similar functions to insect OBPs. For example, the chemosensory protein of the moth *Mamestra brassicae* is also hexahelical with two disulfide bridges (18). However, chemosensory proteins do not share sequence similarity with OBPs; the helices are arranged differently, and the disulfide bonds appear to stabilize loops rather than to link helices. The OBPs of vertebrates are drastically different as they consist largely of an eight-stranded β-barrel (19).

The non-sensory OBPs mentioned above, four *D. melanogaster* isoforms (8), and a number of as yet unpublished insect ESTs that have been deposited in GenBank™ lack one of the three disulfide bonds found in most OBPs, which links the third and sixth helix and is adjacent to the cavity. We refer to this group of proteins here as 4-Cys isoforms because they contain only 4 of the 6 conserved Cys. Previous work (12) suggested that additional OBPs were present in *T. molitor* hemolymph. We have investigated these isoforms to assess their diversity, to determine their relationship to the other OBPs, and to obtain additional OBPs for functional and structural studies.

**EXPERIMENTAL PROCEDURES**

Isoform Purification—Hemolymph (2.5 ml) was collected from *T. molitor* larvae reared under control conditions as described previously (20). Gel-exclusion chromatography and analytical HPLC were done as described by Liou et al. (21) except that the gradient for HPLC was 10–30% B (80% acetonitrile, 0.05% trifluoroacetic acid) in 10 min followed by 30–50% B in 40 min. The actual ⅓ B at elution was determined by timing the delay between a rapid change in ⅓ B and the corresponding change in absorbance. Selected fractions were lyophilized, resuspended, and then reduced in 200 μl of 6 M urea, 50 mM Tris- HCl (pH 6.8), 5 mM dithiothreitol by incubation at 50 °C for 30 min. Cys residues were then carboxamidomethylated with 25 mM iodoacetamide by incubation at 50 °C for 30 min. Cys residues were then carboxamidomethylated with 25 mM iodoacetamide (Sigma) at 50 °C for 30 min. Modifying proteins were reisolated by HPLC as above and digested in 100 μl of 25 mM NH₄HCO₃ (pH 7.9) using 1 μg of bovine pancreatic trypsin (Sigma) at 37 °C for 16 h.

Intact Protein Molecular Mass Determination—Electrospray ionization mass spectrometry was performed on a Micromass Q-TOF2 mass spectrometer (Micromass, Manchester, United Kingdom) in positive ion mode. Average molecular masses of intact proteins were determined by flow injection analysis using a Waters CapLC system with a carrier solvent of 50:50 acetonitrile:water containing 0.1% formic acid at a flow rate of 30 μl/min. Spectra were acquired in an m/z range of 600–2000 using a capillary voltage of 3.5 kV, a cone voltage of 50 V, and a desolvation temperature of 250 °C. The instrument was initially calibrated using a standard solution of horse heart myoglobin (5 pmol/μl, 16,951.49 Da, Sigma). The multiply charged raw data were baseline-subtracted and deconvoluted using MaxEnt1 (22). Acquisition and data analysis were all performed using the MassLynx 3.5 software package supplied by Micromass.

**Peptide Sequence Determination**—Peptide sequence information was obtained on tryptic digests of HPLC-purified protein fractions using a nanospray ionization source on the Q-TOF2 instrument. Concentrated digest samples (3–5 μl) were sprayed from borosilicate capillaries (Type F, Micromass). The time-of-flight analyzer was calibrated using an MS/MS spectrum of [Glu]fibrinopeptide B (Sigma). Survey and MS/MS spectra were acquired in an m/z range of 50–2000 using a cone voltage of 35 V and capillary voltages ranging from 700 to 900 V to optimize spray. Data-dependent acquisition parameters were set to select doubly and triply charged precursor ions for MS/MS analysis. Fragmentation was achieved using argon as the collision gas and varying the collision energy depending on the charge state and the m/z value of the precursor ion. MS/MS spectra were processed by background subtraction and deconvoluted using the MaxEnt3 module of MassLynx 3.5.

Cloning of cDNAs—The following degenerate oligonucleotides were synthesized based on high confidence peptide sequences of two isoforms: sense primer encoding ATEAGDT, 5′-GCG AON GAR GCN GGN GAY AC-3′, and antisense primer encoding PEETAFQT, 5′-GT YTG RAA NGG NGT YTC YTG NGG-3′. Approximately 2 × 10^5 plaque-forming units of a fat body cDNA library (23) were used as the templates in anchor PCR reactions using each of the above primers in combination with the appropriate vector primer (T3 or T7). The first round of amplification was carried out using 1 μm degenerate primer, 100 nm anchor primer, 2.5 units of Taq DNA polymerase (MBI Fermentas), 200 μm each deoxyxylonucleotide triphosphate, 1.5 mM MgCl₂, and the supplied buffer. Cycle conditions were as follows: initial denaturation of phage/primer mixture at 99 °C (5 min); hold at 80 °C while enzyme, buffer, and deoxyxylonucleotide triphosphates were added; and 30 cycles of 95 °C for 1 min, 52 °C for 1 min, and 72 °C for 4 min. Reamplification of 1 μl of each reaction for 25 cycles as above, Taq was removed by proteinase K digestion (20 μg/150 μl for 30 min at 37 °C) followed by phenol/chloroform extraction. After precipitation, DNA was blunt-ended with T4 DNA polymerase (New England Biolabs) as described by Sambrook et al. (24) and gel-purified (Qiagen). Fragments were subcloned into pCR®-Blunt II-TOPO (Invitrogen) and sequenced (Cortec, Kingston, Ontario, Canada). Non-degenerate primers were used to amplify only the coding portions of the two subcloned fragments and the previously isolated THP12α cDNA (12). These fragments were ^32P-labeled and used to screen the cDNA library at low stringency as described previously (21) except that all washes were done at 50 °C. Plaque-purified phage were in *vitro* excised using R408 helper phage (Stratagene), and sequencing was performed on purified plasmid (Qiagen miniprep kit, Qiagen).

**Phylogenetic Analysis**—New protein sequences were manually aligned with selected sequences from a previous alignment (12) derived using a secondary structure mask. The distance matrix was used to generate a neighbor-joining tree (25) using ClustalX (version 8.1) (26) with 1000 bootstrap trials.

**RESULTS**

**Isoform Purification and Characterization**—The proteins in *T. molitor* hemolymph can be separated into two major peaks by gel-exclusion chromatography (Fig. 1, inset). The larger of the two peaks eluted around the void volume, and the second peak was centered around 12–14 kDa. The trailing peak has been shown to contain both antifreeze proteins (21) and a set of proteins of similar masses that showed variable cross-reactivity to THP12α antiserum (12). Further fractionation of this second peak by reversed-phase HPLC (Fig. 1) resulted in over 20 different protein peaks, which eluted between 33 and 44% B. These proteins were devoid of antifreeze activity and
were well separated from the antifreeze protein cluster, which eluted earlier (21). Electrospray ionization MS analysis was also done on the second gel-exclusion peak prior to HPLC fractionation (see supplementary material). The average masses determined by MS are shown in order of peak height, with boldface indicating a relative height of over 50% of maximum. Isoforms that were further characterized by modification (see Table I) and MS/MS (see Fig. 4) are underlined. Asterisks denote masses consistent with cDNA sequences (Fig. 4). The major components (>14% of maximum peak height), detected by MS analysis prior to HPLC fractionation, are numbered in order of decreasing peak height in white on a black background.

Table I

| B | Initial mass (Da) | Alkylated mass (Da) | No. of Cys | No. of fragments (% coverage) | No. of amino acids (% coverage) | cDNA/predicted mass/no. of fragments matched |
|---|------------------|---------------------|------------|------------------------------|-------------------------------|---------------------------------------------|
| 33.4 | 12,246 | 12,478 | 4 | 5 (62) | 30 (28) | 12b (12,247) 5/5 |
| 33.9 | 12,032 | 12,264 | 4 | 5 (53) | 43 (40) | 12d (12,052) 4/5 |
| 37.2 | 12,024 | 12,256 | 4 | 4 (38) | 31 (29) | 12d (12,052) 3/4 |
| 38.7 | 12,767 | 12,999 | 4 | 4 (36) | 36 (31) | 13a (12,840) 4/4 |
| 39.1° | 12,489 | 12,721 | 4 | 1 (11) | 5 (5) | 12e (12,402) 1/1 |
| 39.6 | 12,840 | 13,072 | 4 | 7 (72) | 57 (50) | 13a (12,840) 7/7 |
| 42.1 | 13,492 | 13,725 | 4 | 6 (60) | 53 (45) | 13d (13,492) 6/6 |

° Number of fragments for which sequence and/or mass were obtained and percentage of protein represented by these fragments.

Total number of amino acids determined by MS/MS (see Fig. 4 for details) and percentage of protein represented by these residues.

Predicted average masses that match the observed average mass are shown in bold. The number of tryptic fragments that matched the deduced sequence (Fig. 4) and/or monoisotopic mass of predicted fragments is indicated.

Fragments 1, 2, 3, and 6 matched 12d (numbered from the N terminus relative to the cDNA sequence), but fragment 4 was 37 Da lower.

Fragments 1, 2, 6 matched 12d, but fragment 5 was 6 Da lower.

Matched fragments 2, 3, 4, and 6.

Two isoforms present in this fraction (12,489 and 12,387 Da) were not separated by HPLC chromatography following reduction and modification.

Selected fractions throughout the profile were reduced and modified with iodoacetamide. The mass increases of the eight carboxymethylated isoforms examined indicated that they all contained four Cys residues (Table I). These fractions were digested with trypsin, and a representative mass survey spectrum of the resulting fragments from the 12,840-Da isoform (Fig. 1) is shown in Fig. 2. Major fragments were observed as both singly and doubly charged species and corresponded to masses predicted from a theoretical digest of a protein deduced from a cDNA sequence obtained subsequently. Many of the minor species appear to correspond to incomplete digestion products. For example, fragment 5 is preceded by two Lys residues. Its peak at 802.5 m/z corresponds to the doubly charged fragment. A second doubly

Fig. 1. Purification of THP isoforms present in larval hemolymph. Pooled fractions from an S100 size-exclusion column (inset) were fractionated by reversed-phase HPLC on a C18 analytical column (B = 0.05% trifluoroacetic acid, 80% acetonitrile). The average masses determined by MS are shown in order of peak height, with boldface indicating a relative height of over 50% of maximum. Isoforms that were further characterized by modification (see Table I) and MS/MS (see Fig. 4) are underlined. Asterisks denote masses consistent with cDNA sequences (Fig. 4). The major components (>14% of maximum peak height), detected by MS analysis prior to HPLC fractionation, are numbered in order of decreasing peak height in white on a black background.

Insect Odorant-binding Protein Homologs
charged peak at 866.5 m/z (fragment 5a) results from cleavage after the first rather than the second preceding Lys. A number of doubly charged tryptic fragments from seven different fractions were selected for collisional fragmentation and further analysis. Representative MS/MS mass profiles from two 12-amino acid peptides from different proteins are shown in Fig. 3. The complete series of y ion fragments (27) was observed for these peptides, permitting unambiguous identification of all residues. Other peaks, such as those originating from b ions and internal fragments, were also observed. The two peptides appeared to be derived from the same region of homologous proteins as nine of the 12 residues were identical.

Sequence and/or fragment masses were obtained for seven fractions (Table I). The fragments represented 11–72% of each intact protein, and 5–50% of the residues in each protein were sequenced by MS/MS. Because the sequences of the 12,840- and 13,492-Da isoforms appeared unique, a more extensive MS/MS analysis was performed on these components. The fraction containing the 12,387-Da isoform was contaminated with significant amounts of the 12,489-Da isoform, so it was not analyzed in detail. The sequences obtained from MS/MS experiments were deduced primarily from the y ion series in the spectra and are shown relative to the cDNA sequences in Fig. 4. For some fragments, the entire sequence could be accurately deduced. Partial sequence was obtained when one or more of the y fragment ions were not clearly observed in a spectrum. However, even in these cases, the mass differences between observed peaks (for example, fragment 3, THP12b, y7–y11) were consistent with the masses of the portion deduced from the cDNA sequences (Pro-Asp-Lys). Leu and Ile were not distinguished as their masses are isobaric. Also, because the masses of Gin and Lys are very similar, Lys, at which cleavage did not occur (those adjacent to acidic residues), could not be distinguished from Gin (for example, fragment 3, THP12b). Despite these limitations, mass and sequences corresponding to theoretical tryptic digests were clearly identified in most cases.

The sequences of the fragments from the ≈12-kDa isoforms eluting at 33.4, 33.9, and 37.2% B (Table I) were quite
similar to the previously characterized THP12a (Fig. 4). Therefore, we reasoned that the corresponding cDNAs could be obtained using THP12a as a probe. The isoforms at 38.7 and 39.6% B appeared similar to each other but quite different from THP12a, and the isoform at 42.1% B was unique. Therefore, oligonucleotides were designed to the highest confidence sequences with the lowest codon degeneracy that did not contain Leu/Ile or Lys/Gln (Fig. 4) from the isoforms at 39.6 and 42.1% B. These were used to amplify cDNA fragments from the library by anchor PCR.

Characteristics of cDNA Sequences—The larval fat body cDNA library was screened at low stringency using the cDNA encoding THP12a. A total of eight unique cDNAs (data not shown) encoding 5 additional isoforms (THP12b–f, Fig. 4) were obtained. Two of these encoded the previously cloned THP12a, and the isoform at 42.1% B was unique. Therefore, oligonucleotides were designed to the highest confidence sequences with the lowest codon degeneracy that did not contain Leu/Ile or Lys/Gln (Fig. 4) from the isoforms at 39.6 and 42.1% B. These were used to amplify cDNA fragments from the library by anchor PCR.

The larval fat body cDNA library was screened at low stringency using the cDNA encoding THP12a. A total of eight unique cDNAs (data not shown) encoding 5 additional isoforms (THP12b–f, Fig. 4) were obtained. Two of these encoded the previously cloned THP12a, and although one had 3 silent changes in the coding region and a single change within the 3'-untranslated region (3'-UTR), the second was merely polyadenylated 10 bases further down. The two cDNAs that encode THP12b differ at a single silent position. Although THP12b only differs from THP12a at two amino acid positions, there are up to 8 silent changes and 5 changes within the 3'-UTR, including two insertions/deletions of one and four bases, at which the four cDNA sequences differed. The 11–14 differences (2.9–3.6% divergence) between the cDNAs encoding THP12a and THP12b suggest either that they were derived from a recently duplicated gene or that considerable genetic diversity exists within the insect colony. The additional four isoforms encode THP12c–f, all of which are more distinct and are presumably derived from four separate loci as they share only 60–85% amino acid sequence identity (Figs. 4 and 5).

The cDNA library was also screened at low stringency using the coding portion of the anchor PCR products generated above. The four additional isoforms obtained were named THP13a–d as their masses are closer to 13 kDa. The six unique cDNAs obtained using the PCR product corresponding to THP13d included two that encoded THP13b and THP13c with 1 missense and 3 silent differences. The other

---

**Fig. 4. Alignment of THP isoform sequences deduced from cDNA.** Signal peptides are italicized, N-terminal residues are in purple, Cys is red, and the potential helix breakers, Pro and Gly, are green. Shading is used where over 80% of the residues are hydrophobic (yellow) or hydrophilic (blue). Asterisks and dots are used to denote positions of identity or similarity, respectively. The six α-helices of THP12a are indicated by gray cylinders. Buried residues with side chains that line the binding pocket are indicated with p (pocket), and those that form the hydrophobic core are indicated by c (core) (based upon an energy-minimized model of the known structure of THP12a, data not shown). Black rectangles outline tryptic fragments matching the predicted masses, and the MS/MS-determined residues are in bold. Arrows show the positions of primers used to obtain PCR fragments for screening of the cDNA library. Matches between the masses of intact deduced and observed proteins are in bold. For some proteins, the mass and sequence of tryptic fragments matched, but the mass of the intact protein did not (*2 = 12,489 and 12,387 rather than 12,402; *3 = match + 12,767). In other cases, the mass of a single fragment (see Table I) and the whole protein mass differed (*1 = 12,032 and 12,024 rather than 12,052). The sequences of two other non-sensory OBPs, the B1 protein from T. molitor male accessory gland, and the male-specific serum protein (MSSP) from the medfly are shown as well.
four encoded THP13d with 6–11 base differences at 13 positions (11 silent, 2 within the 3′-UTR) and variation in the polyadenylation site. The polyadenylation signal is only separated from the stop codon by a single base. Three more cDNAs were obtained by screening with the PCR product corresponding to THP13a, including one corresponding to THP13c. The other two encode THP13a with 3 base differences (2 silent, 1 within the 3′-UTR). In addition, one was polyadenylated following a second consensus polyadenylation signal.

Characteristics of the New Isoforms—The cDNAs encode 10 different THP isoforms sharing 32–99% sequence identity (Figs. 4 and 5). A representative sequence encoding each new isoform has been deposited in GenBank™ with accession numbers AY153772–AY153780. The sequence for THP12a was deposited previously with GenBank™/EBI accession number U24237. The deduced masses of the mature proteins encoded by nine of these 10 deduced proteins matched the experimentally determined masses of native proteins (Fig. 1).

Although we obtained intact protein and tryptic fragment masses as well as sequence data for seven proteins, only three deduced proteins (THP12b, THP13a, and THP13d) exactly matched observed proteins (Table I and Fig. 4). The other four proteins appear quite similar to those deduced from three cDNAs (THP12d, THP12e, and THP13a) as indicated by sequence and tryptic fragment mass matches. However, the intact protein masses differed, and in some cases, the mass of a single tryptic fragment also differed. These mass differences cannot easily be explained by post-translational modifications or by proteolysis of the N or C termini of the cDNA-encoded species, but they could arise from one or a few polymorphisms. The presence of numerous similar but largely uncharacterized species (Fig. 1) and the fact that only three of seven proteins appear to exactly correspond to a cDNA suggest that we have probably isolated fewer than half of all the possible cDNA sequences for this protein family.

All of the isoforms contain the 4 Cys residues, which link helices 1 and 3 as well as 5 and 6 but do not contain the additional pair found in 6-Cys OBPs. Only THP12f contains an additional Cys residue in an ectopic location with no apparent partner. The similarities between the isoforms suggest that they adopt similar protein folds. For example, there is a similar patterning of hydrophobic and hydrophilic residues between isoforms, consistent with the amphipathic helical regions of the known structure. Also, both Gly and Pro are well conserved at several positions and are found primarily in the presumed loop regions.

The only differences between the predicted and observed masses were due to disulfide bond formation or the conversion of the N-terminal Glu residue to pyroglutamate (THP13b–d), a modification that was also seen with the antifreeze proteins present in T. molitor hemolymph (21). The signal-peptide cleavage sites were accurately predicted using the program SignalP (28). The exception was THP13a, for which most of the signal-peptide sequence is unknown. However, enough sequence and mass information was available to determine the actual N-terminal residue, and the three residues that precede it (AOA) correspond well to the −3, −1 rule for small residues. Four isoforms (THP12a–d) contain the consensus N-glycosylation signal NXS at the C terminus but
Insect Odorant-binding Protein Homologs

would not be glycosylated as additional C-terminal residues are required for this to occur (29). However, THP12f does contain additional C-terminal residues, which may explain why a matching mass was not observed for this sequence. It may also form an intermolecular disulfide bond through its additional, unpaired Cys, but this is unlikely as this residue is expected to reside in the interior of the protein (Fig. 4).

The pattern of amino acid substitution, particularly between the 12- and 13-kDa proteins, appears to be far from random (Fig. 4). Amino acids that have important roles in protein structure, such as the 4-Cys residues, which form two disulfide bridges, as well as two Gly residues within turns are absolutely conserved. Numerous surface residues, including 4 acidic and 7 basic residues (Fig. 4, see asterisks), are also conserved and appear to be involved in forming salt bridges. However, residues found in the interior of the protein are less conserved, particularly those that line the binding pocket (Fig. 4, denoted by P). This suggests that the binding pocket may have been subject to divergent evolution and that the 12- and 13-kDa groups or subsets may bind to different classes of compounds.

Phylogenetic Comparisons—The cDNA sequences of two pairs of isoforms (THP12a versus THP12b and THP13b versus THP13c) differ by 3.6% or less, suggesting that they may be encoded at the same locus. The same cannot be said of the other six isoforms, which differ by over 15% at the amino acid level. Therefore, there is a minimum of eight different genes encoding this gene family in the beetle and likely 2–3 × this number. The lower molecular mass isoforms (THP12a–f) are more similar to each other than to the higher molecular mass isoforms (THP13a–d), which form a second grouping (Fig. 5). However, these hemolymph isoforms do not appear to be monophyletic as the THP13 group is more similar to the T. molitor accessory gland proteins, TmolB1 and -B2, than to the THP12 group. Nevertheless, these T. molitor isoforms are more similar to each other than to OBPs of other insects, including any of the 38 isoforms (partial data shown) found in D. melanogaster (8). Therefore, OBPs appear to be evolving rapidly, and the T. molitor isoforms were duplicated following the divergence of the major insect orders.

The data neither support nor contradict a common origin for all 4-Cys isoforms. However, the 4-Cys isoforms that have been recovered from Diptera, Coleoptera, and Lepidoptera cluster within each order (Fig. 5, dotted boxes), suggesting that the third disulfide bond is neither easily lost nor gained during evolution. The only exception would appear to be one D. melanogaster OBP (Dmel89A), a 6-Cys isoform that clusters with 4-Cys isoforms. A noticeable trend is that the 6-Cys isoforms appear to be expressed in antennae, whereas 4-Cys isoforms are frequently expressed elsewhere. More expression data will reveal whether this relationship will hold.

DISCUSSION

T. molitor hemolymph was known to contain a number of low molecular mass proteins including OBPs (12), but it was unknown whether the numerous proteins observed following HPLC fractionation were a result of conformational differences and/or post-translational modification of a few sequences. The combined MS sequencing/cDNA cloning approach proved to be a rapid, efficient way to sample isoform diversity across a group of ~20 peaks within the HPLC profile, and the evidence suggests that the mass differences observed result from differences in amino acid sequences. We have obtained cDNAs encoding 10 unique proteins, ranging from 12,052 to 13,492 Da, which encode 4-Cys OBPs present in the hemolymph of the beetle T. molitor. These cDNAs are encoded by at least 8 different genes, but there are likely 2–3-fold more. Some of these additional genes might belong to the THP12 family because cDNAs encoding isoforms with similar fragment sequences (12,024- and 12,032-Da isoforms) were not obtained; also, many other masses were observed within the THP12 isoform-containing region of the HPLC profile (for example, 12,334 and 12,206) (Fig. 1). In addition, 8–10 bands were observed in a low stringency Southern blot of genomic DNA from individual insects probed with THP12a cDNA (12). Others may be more distinct, such as the 13,096- and 12,381-Da isoforms, for which no additional information was obtained, as they lie outside the region of the HPLC profile showing high cross-reactivity to anti-THP12a antibodies (12). Overall, this analysis has provided a good estimate of the isoform diversity within this family of proteins and has revealed that OBPs are the most abundant smaller proteins (within the 6–20-kDa range) in hemolymph.

Currently, most known insect OBPs contain 6 Cys, and in D. melanogaster, where the whole genome is known, only four of the 38 OBPs belong to the 4-Cys group (8). Three are known from the hemolymph of the medfly C. capitata (11). The number of 4-Cys isoforms and indeed the number of OBPs in general are increasing dramatically as a result of the various genome and EST sequencing projects. However, T. molitor is the only beetle for which 4-Cys isoforms have been recovered. It is evident that the 4-Cys isoforms from T. molitor, including those found in the accessory sex gland, have arisen by gene duplication after the major insect orders arose over 300 million years ago (30) as they are more similar to each other than to any of the 38 OBPs found in D. melanogaster. This indicates that this gene family is undergoing rapid evolution and that the genes are being duplicated frequently.

The 4-Cys OBPs found in T. molitor hemolymph were more variable than we initially expected, showing as little as 32% amino acid identity. Despite this, the sequences aligned very well as there were only a few single amino acid deletions or insertions plus some variability in the length of the N and C termini. The differences, especially between more divergent isoforms, are not consistent with major changes to the helical structure of the protein but are consistent with changes to the binding pocket. Therefore, it is possible that these OBPs have arisen to carry a number of different compounds specific to T. molitor or to beetles and that functionally im-
Insect Odorant-binding Protein Homologs

important residues in the vicinity of the binding pocket are subject to divergent evolution. Alternatively, natural selection may have produced binding pockets of similar shape in different isoforms from different insects for the purpose of carrying the same compound. Unfortunately, the rapid evolution of insect OBPs, whether they contain 4 or 6 Cys, means that the complete evolutionary history of the 4-Cys isoforms may be very difficult to resolve because they are frequently too dissimilar to enable an accurate assessment of their relatedness.

Insects possess divergent OBPs found in a wide variety of different tissues. These insect OBPs appear to be functionally analogous to the structurally unrelated lipocalins of vertebrates, which are also highly divergent (31). Lipocalins have adopted a wide range of roles from binding compounds such as odorants, pheromones, and fatty acids to enzymatic functions in a wide variety of tissues. Lipocalins are also found in insects, but they appear to be far less abundant as only two lipocalins have been reported in D. melanogaster (32). Therefore, insect OBPs rather than lipocalins may be the major transporters of small hydrophobic compounds in insects, although some may well have adopted unexpected roles. Unfortunately, as of yet the functions of most insect OBPs have not been determined. It is our hypothesis that the beetle THPs are carriers of a number of small hydrophobic compounds that would normally be transported through the hemolymph, and we are currently working toward testing this hypothesis.

Acknowledgments—We thank Micromass (Manchester, United Kingdom) for performing electrospray time-of-flight MS on the unfraccionated THP peak, Sherry Gauthier for assistance with PCR and cDNA cloning, Eva Kwok for assistance with isoform purification, modification, and digestion, and Michael J. Kuiper for protein modeling.

Note Added in Proof—A third OBP structure, that of the alcohol-binding protein LUSH from Drosophila, has recently been solved (see Kruse, S. W., Zhao, R., Smith, D. P., and Jones, D. N. M. (2003) Structure of a specific alcohol-binding site defined by the odorant binding protein LUSH from Drosophila melanogaster. Nat. Struct. Biol. 10, 694–700).

* 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 on-line version of this article (available at http://www.mcponline.org) contains supplementary material.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) AY153772–AY153780.

To whom correspondence should be addressed. Tel.: 1–613-533-2984; Fax: 1–613-533-2497; E-mail: grahamla@post.queensu.ca.

Supported in part by a Killiam Research Fellowship.

REFERENCES

1. Vogt, R. G., and Riddiford, L. M. (1981) Pheromone binding and inactivation by moth antennae. Nature 293, 161–163
2. Pelosi, P. (1995) Perireceptor events in olfaction. J. Neurobiol. 30, 3–19
3. Plettnner, E., Lazar, J., Prestwich, E. G., and Prestwich, G. G. (2000) Discrimination of pheromone enantiomers by two pheromone binding proteins from the gypsy moth, Lymantria dispar. Biochemistry 39, 8953–8962
4. Campanacci, V., Krieger, J., Bette, S., Sturgis, J. N., Larrique, A., Cambilau, C., Breer, H., and Tegoni, M. (2001) Revisiting the specificity of Mamestra brassicae and Antheraea polyphemus pheromone-binding proteins with a fluorescence binding assay. J. Biol. Chem. 276, 20078–20084
5. Briand, L., Nespoilous, C., Huet, J.-C., Takahashi, M., and Pemollet, J.-C. (2001) Ligand binding and physico-chemical properties of ASP2, a recombinant odorant-binding protein from honeybee (Apis mellifera L.). Eur. J. Biochem. 268, 752–760
6. Galindo, K., and Smith, D. P. (2001) A large family of divergent Drosophila odorant-binding proteins expressed in gustatory and olfactory sensilla. Genetics 159, 1059–1072
7. Vogt, R. G., Rogers, M. E., Franco, M. D., and Sun, M. (2002) A comparative study of odorant binding protein genes: differential expression of the PBP1-GOBP2 gene cluster in Manduca sexta (Lepidoptera) and the organization of OBP genes in Drosophila melanogaster (Diptera). J. Exp. Biol. 205, 719–744
8. Graham, L. A., and Davies, P. L. (2002) The odorant-binding proteins of Drosophila melanogaster: annotation and characterization of a divergent gene family. Gene (Amst.) 292, 43–55
9. Kim, M.-S., Repp, A., and Smith, D. P. (1998) Lush odorant-binding protein mediates chemosensory responses to alcohols in Drosophila melanogaster. Genetics 150, 711–721
10. Park, S.-K., Shanbhag, S. R., Wang, Q., Hasan, G., Steinbrecht, R. A., and Pikelney, C. W. (2000) Expression patterns of two putative odorant-binding proteins in the olfactory organs of Drosophila melanogaster have different implications for their functions. Cell Tissue Res. 304, 423–437
11. Christophides, G., Mintzas, A. C., and Komitopoulou, K. (2000) Organization, evolution and expression of the MSSP multigene family encoding putative members of the odorant binding protein family in the medfly Ceratitis capitata. Insect Mol. Biol. 9, 185–195
12. Graham, L. A., Tang, W., Baust, J. G., Liou, Y.-C., Reid, T. S., and Davies, P. L. (2001) Characterization and cloning of a Tenebrio molitor hemo-lymph protein with sequence similarity to insect odorant-binding proteins. Insect Biochem. Mol. Biol. 31, 691–702
13. Kodrik, D., Filippov, V. A., Filippova, M. A., and Sehgal, F. (1995) Sericotropicin: an insect neurohormonal factor affecting RNA transcription. Neth. J. Zool. 45, 68–70
14. Paesen, G. C., and Hopp, G. M. (1995) The B proteins secreted by the tubular accessory sex glands of the male mealworm beetle, Tenebrio molitor, have sequence similarity to moth pheromone-binding proteins. Insect Biochem. Mol. Biol. 25, 401–408
15. Rothemund, S., Liou, Y.-C., Davies, P. L., Krause, E., and Sönntgen, F. D. (1999) A new class of hexahelical insect proteins revealed as putative carriers of small hydrophobic ligands. Structure 7, 1325–1332
16. Sandler, B. H., Nikonova, L., Leal, W. S., and Clardy, J. (2000) Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein-bombykol complex. Chem. Biol. 7, 143–151
17. Horst, R., Damberger, F., Lügibuhli, P., Guntert, P., Peng, G., Nikonova, L., Leal, W. S., and Wüthrich, K. (2001) NMR structure reveals intramolecular regulation mechanism for pheromone binding and release. Proc. Natl. Acad. Sci. U. S. A. 98, 14374–14379
18. Larrique, A., Campanacci, V., Roussel, A., Larsson, A. M., Jones, T. A., Tegoni, M., and Cambilau, C. (2002) X-ray structure and ligand binding study of a moth chemosensory protein. J. Biol. Chem. 277, 32094–32098
19. Blanchet, M. A., Bains, G., Pelosi, P., Pevsner, J., Snyder, S. H., Monaco, H. H., and Anzel, L. M. (1998) The three-dimensional structure of bovine odorant-binding protein and its mechanism of odor recognition. Nat. Struct. Biol. 3, 934–939
20. Graham, L. A., Walker, V. K., and Davies, P. L. (2000) Developmental and environmental regulation of antifeetze proteins in the mealworm beetle Tenebrio molitor. Eur. J. Biochem. 267, 6452–6458
21. Liou, Y.-C., Thibault, P., Walker, V. K., Davies, P. L., and Graham, L. A. (1999) A complex family of highly heterogeneous and internally repetitive hyperactive antifeetze proteins from the beetle Tenebrio molitor. Biochemistry 38, 11415–11424
22. Ferrige, A. G., Seddon, M. J., Green, B. N., Jarvis, S. A., and Skillings, J.
(1992) Disentangling electrospray spectra with maximum entropy. *Rapid Commun. Mass Spectrom.* 6, 707–711

23. Graham, L. A., Bendena, W. G., and Walker, V. K. (1996) Cloning and baculovirus expression of a desiccation stress gene from the beetle, *Tenebrio molitor*. *Insect Biochem.* 26, 127–133

24. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) in *Molecular Cloning: A Laboratory Manual*, 2nd Ed., pp. F4–F5, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

25. Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425

26. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The clustal_x windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882

27. Johnson, R. S., Martin, S. A., and Biemann, K. (1988) Collision-induced fragmentation of $(M + H)^+$ ions of peptides. Side chain specific sequence ions. *Int. J. Mass Spectrom. Ion Process.* 86, 137–154

28. Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* 10, 1–6

29. Bause, E., and Hettkamp, H. (1979) Primary structural requirements for N-glycosylation of peptides in rat liver. *FEBS Lett.* 108, 341–344

30. Gaunt, M. W., and Miles, M. A. (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol. Biol. Evol.* 19, 748–761

31. Ganfornina, M. D., Gutiérrez, G., Bastiani, M., and Sánchez, D. (2000) A phylogenetic analysis of the lipocalin protein family. *Mol. Biol. Evol.* 17, 114–126

32. Sánchez, D., Ganfornina, M. D., Torres-Schumann, S., Speese, S. D., Lora, J. M., and Bastiani, M. J. (2000) Characterization of two novel lipocalins expressed in the *Drosophila* embryonic nervous system. *Int. J. Dev. Biol.* 44, 349–359

33. Vogt, R. G., Rybczynski, R., and Lerner, M. R. (1991) Molecular cloning and sequencing of general odorant-binding proteins GOBP-1 and GOBP-2 from the tobacco hawk moth Manduca sexta: comparisons with other insect OBPs and their signal peptides. *J. Neurosci.* 11, 2972–2984

34. Krieger, J., von Nickisch-Rosenegk, E., Mamel, I., Pelosi, P., and Breer, H. (1996) Binding proteins from the antennae of *Bombyx mori*. *Insect Biochem.* 26, 297–307

35. Wojtasek, H., Picimbon, J.-F., and Leal, W. S. (1999) Identification and cloning of odorant-binding proteins from the scarab beetle *Phyllopertha diversa*. *Biochem. Biophys. Res. Commun.* 263, 832–837

36. Vogt, R. G., Callahan, F. E., Rogers, M. E., and Dickens, J. C. (1999) Odorant binding protein diversity and distribution among the insect orders, as indicated by LAP, an OBP-related protein of the true bug *Lygus lineolaris* (Hemiptera, Heteroptera). *Chem. Senses* 24, 481–495

37. Robertson, H. M., Martos, R., Sears, C. R., Todres, E. Z., Walden, K. K. O., and Nardi, J. B. (1999) Diversity of odorant binding proteins revealed by an expressed sequence tag project on male *Manduca sexta* moth antennae. *Insect Mol. Biol.* 8, 501–518

38. Wojtasek, H., Hansson, B. S., and Leal, W. S. (1998) Attracted or repelled?—a matter of two neurons, one pheromone-binding protein, and a chiral center. *Biochem. Biophys. Res. Commun.* 250, 217–222