Expression of Cytochrome P450 Genes of the CYP4 Family in Midgut and Fat Body of the Tobacco Hornworm, *Manduca sexta* 1

Mark J. Snyder, Jennitte L. Stevens, John F. Andersen, and René Feyereisen 2

Center for Insect Science and Department of Entomology, University of Arizona, Tucson, Arizona 85721

Received November 17, 1994, and in revised form May 10, 1995

Two conserved regions in the alignment of cytochrome P450 family 4 (CYP4) proteins served as guide to the synthesis of degenerate oligonucleotide primers. The primers were used in PCR from a midgut cDNA library and RT-PCR from fat body mRNA, both from last instar larvae of the tobacco hornworm, *Manduca sexta*. The PCR products of 443–449 bp were cloned and sequenced. Nine P450 clones representing four new genes were obtained from the midgut. Fifteen P450 clones representing three new genes were obtained from the fat body. Two genes were expressed in both tissues. A number of putative allelic variants were also observed for three of the P450 genes. The resulting sequences of 130–132 amino acids were aligned to generate a parsimony analysis of CYP4 P450 proteins. Two new subfamilies of CYP4 were designated from *M. sexta* by these procedures, CYP4L and CYP4M. The sequence of a full-length cDNA clone for CYP4M2 (41.2% identity to CYP4C1) confirmed that the PCR products obtained by this method were P450s belonging to the CYP4 family. The developmental expression of the CYP4 genes appeared to be coordinately regulated in both fat body and midgut. In the fat body, CYP4 mRNA levels declined after the first day of the final larval instar, peaked during the wandering stage, and fell again until the prepupal molt. Midgut CYP4 mRNA levels were higher during the active feeding, midwandering, prepupal, and pupal stages. Addition of 2-tridecanone or 2-undecanone to the diet induced several P450s in the midgut and in the fat body. Phenobarbital induced CYP4M1 in the fat body and dietary clofibrate induced the mRNA levels of CYP4M1 and CYP4M3 in the midgut. The results indicate that at least four CYP4 genes are expressed in single tissues of a Lepidopteran insect. Several of these P450 may be involved in tissue responses to xenobiotics. © 1995 Academic Press, Inc.

Key Words: insect P450; CYP4 family; induction; aliphatic ketones; phenobarbital; clofibrate.

Cytochrome P450 monoxygenase enzymes comprise an ancient and widely distributed protein superfamily. To date, more than 230 sequences representing 36 gene families have been described (1). With few exceptions, members of the same family share >40% amino acid sequence identity and P450 proteins within the same subfamily are >55% identical. P450 proteins are found in a very diverse array of organisms, bacteria, plants, fungi, and animals. Phylogenetic analysis of this diversity suggests that there was a common ancestor to all present day P450 forms that existed prior to the divergence of eukaryotes and prokaryotes (2).

There are a multitude of functions attributed to P450 proteins including the metabolism of a wide variety of both endogenous substrates, such as hormones and lipids, and xenobiotics (3). In insects, P450 proteins also metabolize hormones and pheromones (4) but have been studied in most detail for their roles in insecticide resistance (5). The overexpression of two P450 genes, CYP6A1 and CYP6A2, is associated with insecticide resistance in the housefly, *Musca domestica*, and in *Drosophila melanogaster* (6–8).

In addition to the roles described above, P450-dependent metabolism has been associated with the adaptation of insect and vertebrate herbivores to host plant chemicals. The adaptation of herbivores to plant chemicals has been proposed to be one of the driving forces in P450 diversification (9). Krieger et al. (10) first proposed that insect midgut microsomal P450s arose to defend against natural insecticides present in the lar-
val food plants. Increased expression of the P450 gene CYP6B1 was shown to allow larvae of a specialist insect, the black swallowtail butterfly Papilio polyxenes, to feed on host plants (e.g., parsnip) high in furanocoumarins, which are toxic plant defense compounds (11, 12).

Although P450s have many demonstrated functions, little is known about their diversity in insects. Biochemical methods have been used to demonstrate the existence of multiple forms of insect P450 (review in Ref. 5). Such methods (solubilization, chromatography) have revealed for instance the existence of seven or more P450 forms in the housefly (13). Molecular techniques now confirm such multiplicity, for instance, by revealing the existence of a cluster of more than five active P450 genes in the same species (14). We report here that 5 CYP4 genes are expressed in tissues of the tobacco hornworm, Manduca sexta. Previous studies have identified CYP4 genes from the cockroach, Blaberus discoidalis (CYP4Cl, 15) and from D. melanogaster (CYP4D1, 16; CYP4E1, 17; CYP4E2, 18; CYP4D2, 19). We show here that M. sexta has at least two new subfamilies of CYP4 designated CYP4L and CYP4M with three or four genes expressed in fat body and midgut, respectively. Northern hybridization experiments demonstrate that mRNA levels for several of these CYP4 genes change during development and in response to the addition of plant chemicals and xenobiotics to the diet.

MATERIALS AND METHODS

Library construction and mRNA isolation. M. sexta larvae were maintained as previously described (20). A midgut cDNA library was constructed from 36-h-old fifth instar M. sexta larvae from poly(A)^+ mRNA (Fast Track kit, Invitrogen) in the pcDNA II vector (The Librarian kit, Invitrogen). Fat body mRNA was isolated using the same procedure.

Primers for gene cloning. The degenerate (29-mer) forward primer 1, GA(C/T)ACITT(CT)ATGTT(CT/G)AGA/AGGCACGT)CA/CT/GA/CT/AC, was designed around the 1 helix region of the cockroach CYP4C1 (DTFMFEGHDT, residues 309–318; 15). The degenerate (23-mer) reverse primer 2, GC/AG/AT/CT/TT/CT/TC/AGC/TCC/AGT/ATT/AC/AGA/GA/AT/TT, was designed around the heme binding region (NCIGQKFA, residues 451–458 of cockroach CYP4C1; 15). Reverse transcription of fat body mRNA was done using SuperScript enzyme at 42°C according to the manufacturer’s directions (GIBCO BRL). The PCR cycles for both the midgut cDNA and the fat body cDNA were 94°C, 1 min; 54°C, 1 min; 72°C, 1 min, for 35 cycles and 2.5 mM Mg^2+ in the PCR buffer. The PCR products were purified by gel electrophoresis and cloned into the pCR II vector (Invitrogen). Sequence analysis was done using the NCBI Blast Network Server. Clones were further classified, along with homologous regions of selected CYP4 proteins, by parsimony analysis using PAUP (21).

CYP4M2 cloning. A mixture of random-primed ^32P-labeled CYP4L and CYP4M PCR fragments was used as probe to screen a cDNA library prepared as described above from midguts of 36-h-old fifth instar M. sexta larvae fed 0.15% undecanone in an artificial diet. Dual filter lifts of 36,000 clones yielded a clone carrying the complete coding sequence of CYP4M2. This clone was completely sequenced on both strands by the dideoxy chain termination method.

Tissue expression. Total RNA, isolated using the RNAsaid kit (BIO 101), was prepared from midguts of Day 4 fifth instar larvae fed for 72 h on artificial diets (Bioserv) with or without 0.75% nicotine, 0.5% tridecane, 0.5% undecanone, 0.1% clofibrate, or 0.5% phenobarbital. Midguts were dissected free of peritrophic membrane prior to RNA isolation. For northern analysis, 5 µg of RNA was loaded per lane, separated by gel electrophoresis, and blotted onto GeneScreen nylon membrane (DuPont NEN). Additionally, dilution dot blots on GeneScreen were prepared with RNA samples covering the developmental pattern from the fifth larval stage to the adult stage. The blots were hybridized at 42°C to one of the random-primed ^32P-labeled M. sexta P450 clone inserts in 50% formamide, 5 × SSPE, pH 7.4, 5 × Denhardt’s solution, and 1% SDS. Final washing was at 65°C for 15 min in 2 × SSPE and 2% SDS. The blots were stripped and reprobed with other random-primed ^32P-labeled P450 probes as well as with an actin probe from D. melanogaster to standardize the mRNA levels in each lane. The exposed films from each blot were scanned on a LKB Ultrascan laser densitometer and the resulting values were corrected for equal loading by comparison to the levels of actin mRNA.

RESULTS

P450 Cloning Strategy and Characterization of the PCR Products

A combination of degenerate primers for two conserved regions of CYP4 proteins was used to amplify cDNAs from a midgut cDNA library and from reverse-transcribed fat body mRNA of the tobacco hornworm. The amplification products were cloned and 17 midgut and 23 fat body clones were sequenced. This effort resulted in 24 clones that were 443–449 bp long and that scored highly with P450 proteins of the CYP4 family by BLAST Network (NCBI) searches (Table 1). Nine of the 17 clones generated from the M. sexta midgut cDNA library were resolved into four new P450s based on nucleotide sequence alignments: CYP4L2, CYP4M1, CYP4M2, and CYP4M3, as well as two allelic variants of CYP4M2. Allelic variants were named v1 and v2 according to the rules of P450 nomenclature. Fifteen of the 23 clones generated from reverse-transcribed fat body mRNA were resolved into three new P450 genes. CYP4L2 is 93.9% identical at the amino acid level to CYP4L1 identifying it as a separate gene, rather than an allelic variant (1). The alignment of the deduced amino acid sequences is shown in Fig. 1. The designation of allelic variants in Table I is simply a reflection of the <3% difference rule of the P450 nomenclature. We cannot exclude the possibility that two of the alleles of CYP4M2 which differ by a single nucleotide may have resulted from PCR error. Further genetic characterization may indicate that some sequences called allelic variants in fact represent distinct genes and conversely that CYP4L1 and CYP4L2 may be allelic variants.

Abbreviations used: PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate.
TABLE I  
CYP4 Gene Fragments Cloned by PCR from M. sexta Midgut and Fat Body

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Allelic variants</th>
<th>Tissue</th>
<th>cDNAs found</th>
<th>Nucleotide changes</th>
<th>Amino acid changes</th>
<th>GenBank Accession No.</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP4L1</td>
<td></td>
<td>Fat body</td>
<td>2</td>
<td>4</td>
<td>(D28E, F106S, K107R)</td>
<td>L38668</td>
</tr>
<tr>
<td>CYP4Llv1</td>
<td></td>
<td>Fat body</td>
<td>3</td>
<td>4</td>
<td>(D28E, F106S, K107R)</td>
<td>L38669</td>
</tr>
<tr>
<td>CYP4Llv2</td>
<td></td>
<td>Fat body</td>
<td>1</td>
<td>4</td>
<td>(D28E, F106S, K107R)</td>
<td>L38670</td>
</tr>
<tr>
<td>CYP4M1</td>
<td></td>
<td>Midgut</td>
<td>2</td>
<td>14</td>
<td>(T34M, G80V, Q81P, D83G, T84Q)</td>
<td>L38671</td>
</tr>
<tr>
<td>CYP4Mlv1</td>
<td></td>
<td>Midgut</td>
<td>3</td>
<td>1</td>
<td>(V73A)</td>
<td>L38672</td>
</tr>
<tr>
<td>CYP4M2</td>
<td></td>
<td>Fat body</td>
<td>4</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP4M2v1</td>
<td></td>
<td>Fat body</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP4M2v2</td>
<td></td>
<td>Fat body</td>
<td>2</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP4M2v3</td>
<td></td>
<td>Midgut</td>
<td>1</td>
<td>1</td>
<td>(V73A)</td>
<td></td>
</tr>
<tr>
<td>CYP4M2v4</td>
<td></td>
<td>Midgut</td>
<td>1</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYP4M3</td>
<td></td>
<td>Midgut</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Name of gene according to P450 nomenclature (Ref. 1).
* Allelic variant defined as <3% difference at the amino acid level (Ref. 1).
* Clones obtained from PCR of midgut cDNA library or RT-PCR of fat body mRNA.
* Number of clones found.

**A Full-Length cDNA for CYP4M2**

A clone was obtained by screening a cDNA library of M. sexta midgut with a mixture of CYP4L and CYP4M PCR products as hybridization probe. Sequence analysis of this cDNA confirmed that the PCR product for CYP4M2 was entirely contained in a typical P450 sequence of 503 amino acids (Fig. 2). Moreover, the sequence flanking the PCR product corresponded exactly to the conserved amino acid sequences used in the design of the degenerate PCR primers. The full-length CYP4M2 sequence was 41.2% identical to the cockroach CYP4Cl sequence, thus validating the assignment of the PCR products to the CYP4 family.

**Relatedness of the P450 Fragments**

The percentage identity of the deduced amino acid sequences of the PCR products is shown in matrix form in Table II. The M. sexta CYP4 fragments were 34.1–93.9% identical to each other. An estimate of the percentage identity of the whole protein is also given in

![FIG. 1](image-url)

Alignment of the amino acid sequence deduced from the sequences of the cloned CYP4 PCR products. The conceptual sequence corresponding to the PCR primers is shown for CYP4L1 to orient the sequences with respect to the conserved Thr and Cys, which are underlined. Amino acid residues common to all sequences are marked with asterisks. Residue changes in the allelic variants are underlined.
Table II. This was calculated as follows (23). A function was derived to relate the percentage identity of known, whole length P450 proteins with the percentage identity of the portion of these proteins corresponding in length and location to the PCR product obtained with our primer set. Because of the universal conservation of helix I and the heme-binding peptide, it was easy to identify the correct P450 fragment to be used in this calculation. With this formula we calculated that the whole CYP4 proteins of *M. sexta* would be 23.9–91.3% identical (Table II).

<table>
<thead>
<tr>
<th>CYP4 Genes of <em>M. sexta</em></th>
<th>CYP4L1</th>
<th>CYP4L2</th>
<th>CYP4M1</th>
<th>CYP4M2</th>
<th>CYP4M3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP4L1</td>
<td>93.9</td>
<td>35.9</td>
<td>38.6</td>
<td>36.4</td>
<td></td>
</tr>
<tr>
<td>CYP4L2</td>
<td>91.3</td>
<td>34.1</td>
<td>38.6</td>
<td>36.4</td>
<td></td>
</tr>
<tr>
<td>CYP4M1</td>
<td>25.6</td>
<td>23.9</td>
<td>60.1</td>
<td>52.6</td>
<td></td>
</tr>
<tr>
<td>CYP4M2</td>
<td>27.7</td>
<td>28.0</td>
<td>41.8</td>
<td>43.4</td>
<td></td>
</tr>
<tr>
<td>CYP4M3</td>
<td>26.2</td>
<td>26.0</td>
<td>41.8</td>
<td>43.4</td>
<td></td>
</tr>
</tbody>
</table>

*Note*. Numbers on top of the diagonal refer to the % identity of the PCR fragments (from the alignment Fig. 1, computed by the DISTANCES program of GCG (threshold 1.5; denom. length of shorter sequence without gaps). Numbers below the diagonal refer to the % identity of the whole P450 protein calculated using the function of Scott et al. (23).
noted that the CYP4 family has been expanded beyond the 40% rule (1) by inclusion of CYP4C1 and CYP4D1.

**Expression of *M. sexta* CYP4 Genes**

The expression of the *M. sexta* CYP4 genes was studied in a series of Northern and dot blot hybridization experiments. It is likely that the hybridization signal for CYP4L2 accounted for the mRNA levels of CYP4L2 and CYP4L1, because of the close relatedness of these two genes (93.9%). The expression of CYP4 genes during larval–adult development of *M. sexta* was monitored in fat body and midgut (Fig. 4). An increase in CYP4 expression (2- to 7.5-fold) was found in the fat body during the wandering stage on Day 6. In the midgut, CYP4 mRNA levels were high during feeding, wandering, prepupal, and pupal stages. CYP4 expression was low in the fat body of adult females. CYP4L2, CYP4M1, and CYP4M2 showed increased expression in midguts of larvae starved from Day 3 to Day 4 relative to normal Day 4 larvae (Fig. 4). CYP4M1 levels were high in the midgut of adult females.

With specific induction regimes, CYP4M1 and CYP4M3 were highly expressed in midgut while CYP4L2 and CYP4M2 were highly expressed in fat body. Northern blots of midgut RNA from *M. sexta* larvae fed artificial diets containing plant chemicals or phenobarbital are shown in Fig. 5. Dilution dot-blot experiments showed that the expression of CYP4M3 was induced by 2-tridecanone (5-fold), 2-undecanone (3-fold), and to a larger extent by clofibrate (15-fold) in the midgut relative to controls fed an artificial diet without inducers. CYP4M1 was also highly inducible by clofibrate (16-fold). CYP4M2 was expressed at lower levels in the midgut and not affected by any of these dietary chemicals (Figs. 5 and 6). In the fat body, the...
expression of CYP4M1 was induced 2-fold by 2-tridecanone, 3-fold by 2-undecanone, and 6-fold by phenobarbital (Fig. 6).

DISCUSSION

Insect P450 Diversity

We have identified five new P450 genes in the tobacco hornworm, M. sexta by sequencing PCR products obtained with degenerate primers. The reverse primer was designed based on the heme binding region conserved in all P450 proteins. The forward primer was designed from a highly conserved region of all known CYP4 proteins from mammals and insects that corresponds to the I helix of P450cam (CYP101). The highly conserved region (residues 309–317 of the cockroach B. discoidalis CYP4C1; 15) places any such sequences within the CYP4 P450 family. This formal exception to the rule of 40% identity of members of a CYP family (1) has had the effect of “lumping” a series of insect P450 proteins into the CYP4 family. Because of the relatively low annealing temperature of the PCR primers, we cannot assume that the sequences represented by the PCR primers are actually represented in the corresponding full-length sequences. However, sequencing of a cDNA clone from M. sexta, CYP4M2, as well as of a cDNA obtained in similar fashion from D. punctata (Andersen et al., unpublished results) has confirmed the presence of the highly conserved helix I sequence in these two P450 proteins. Thus, placing the sequences obtained by PCR in the CYP4 family, while formally violating the 40% identity rule in some cases, is a reasonable compromise that avoids balkanization of the P450 nomenclature. An advantage of the PCR technique is that a wide variety of P450 sequences are sampled and made available for physiological or toxicological studies. Use of a PCR strategy similar to this one has been successful in identifying CYP4 sequences from guinea-pig, human, and mouse (22) as well as from the mosquito Anopheles albimanus (23). The ease with which allelic variants of P450 genes were obtained in M. sexta and in P. polyxenes (11) may greatly facilitate the study of P450 population genetics.

The total number of cytochrome P450 forms in M. sexta or any other insect species is as yet unknown. Current estimates place that number at between 50 and 100 P450 genes in a primate or rodent (24). Clearly, molecular approaches will soon permit us to estimate the number of P450 genes of an insect genome. Nothing is known about the chromosomal location of the five CYP4 genes of M. sexta. A cluster of six CYP6 genes was found on chromosome V of the housefly, M. domestica (14). Four CYP4 genes are linked on chromosome X of Drosophila melanogaster (16, 19, Dunkov et al., unpublished results). A similar finding of glutathione-S-transferase gene clusters in D. melanogaster (25) argues that detoxification enzymes may be closely grouped on insect chromosomes. P450 gene clusters have been identified in various mammalian species (1, 24) and in yeast (26).

Induction of CYP4 Genes

Cytochrome P450 levels in insects are dependent on a number of factors including the host plant or specific inducer added to the diet (27). In M. sexta midgut and fat body, three CYP4 genes were inducible by the aliphatic ketones 2-tridecanone or 2-undecanone commonly found in wild tomato, by phenobarbital, or by the peroxisome proliferator clofibrate (Fig. 6). For instance we observed coordinate induction of CYP4M1 and CYP4M3 by clofibrate in the midgut. Riskallah et al. (28) had reported that 2-tridecanone caused distinct changes in the spectral characteristics of midgut P450 in larvae of the tobacco budworm, Heliothis virescens. The selective induction of certain P450 genes as documented here, rather than a general increase of all P450s would explain these results. It is difficult to ascertain the physiological importance of the percentage induction of a particular P450 gene by xenobiotics until mRNA levels are correlated with protein levels and enzyme activities. Maximal induction of CYP4M1 was 16-fold in the midgut and 6-fold in the fat body (Fig. 6). Midgut microsomal P450 enzymatic activities were
shown to be induced 1.4- to 10.0-fold by nicotine ingestion (20). Tate et al. (29) have argued that xenobiotics do not induce midgut detoxification enzyme activities as extensively in this species as compared to other insects. The relatively low induction of *M. sexta* midgut glutathione S-transferases by dietary xenobiotics supports this view (30).

In the cockroach, *B. discoidalis*, there was a 50-fold induction of CYP4C1 in the fat body following injection of the peptide hypertrhelasoleral hormone (15). Starvation also induced CYP4C1 expression in the cockroach fat body, and we have observed starvation-induced expression of CYP4L2 and CYP4M2 in the midgut (Fig. 4). Carino *et al.* (6) found a massive induction of CYP6A1 by phenobarbital in the housefly, *M. domestica*. Induction of CYP6B1 was observed in larvae of the black swallowtail, *P. polyxenes*, in response to dietary xanthotoxin, a common furanocoumarin from its host plants (11). The largest inductions of *M. sexta* CYP4 genes noted were 6-fold by phenobarbital in fat body and 16-fold by the peroxisome proliferator clofibrate in midgut (Fig. 6). These findings extend to the CYP4 family the previous discoveries of developmental induction of some members of the CYP2, CYP3, CYP6, CYP73, CYP102, and CYP106 families. There is also evidence for CYP4E2 induction by phenobarbital in an insecticide-resistant strain of *Drosophila* (18). It will be of interest to search for 5′ regulatory elements in *M. sexta* CYP4M1 that may be related to similar phenobarbital responsive regions in other P450 genes, in particular the barbie box (31), which has been observed upstream of two phenobarbital inducible CYP6A genes from insects (32). Likewise, regions corresponding to the peroxisome-proliferator regulatory element that regulates induction of CYP4A6 (33) may be associated with CYP4M1 and CYP4M3 genes in *M. sexta*.

**Tissue and Age Differences**

The midgut is generally thought to be the primary detoxification organ of Lepidopteran larvae (5) although Tate *et al.* (29) showed comparable fat body cytochrome P450 and glutathione S-transferase activities in *M. sexta*. Glutathione S-transferase activity was comparable between *M. sexta* larval midgut and fat body and was inducible in fat body by aliphatic ketones and by phenobarbital (30). Lepidopteran midgut and fat body P450 activities are high during active feeding in the middle of larval instars (29, 34). This appears to be true for *M. sexta* midgut, but not for fat body CYP4 expression during the active feeding period, Days 1–4 or 5 of the last instar (Fig. 4). CYP4, gene expression in the midgut remained high during preupal and pupal development. CYP4 expression is low in adult *M. sexta* midgut and fat body except for CYP4M1 (Fig. 4). Little is known about developmental changes of P450 mRNA levels in insects. Housefly CYP6A1 expression is very low in embryos, rises during larval development to a maximum in actively feeding late larvae, drops in pre­pupae and during the pupal stage, and rises again in adults of both sexes (7). CYP4D1 is expressed in a very similar pattern in *Drosophila* with a peak during the late (3rd) larval instar (16). In general, these patterns correspond to observed changes in P450 enzyme activities (5). The fact that the same gene can be expressed with a different developmental pattern in different tissues was unexpected from toxicological or physiological data in insects. A more precise study of such correlations is now possible in *M. sexta*, an insect that can be precisely staged. Recovery of active CYP4 enzymes following insertion of cDNAs into expression systems (35) will be required to delineate the functions of these various new P450 genes.

**ACKNOWLEDGMENTS**

We thank C. Summers and J. Longbottom for technical assistance and D. Nelson (University of Tennessee, Memphis) for help in naming the genes. J. L. Stevens was supported in part by the University of Arizona Undergraduate Biology Research Program. This work was supported by USDA Grant 89-37263-4960 and NIH Grant ES06694.

**REFERENCES**


