Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 1465–1469, March 1989
Biochemistry

Isolation and sequence of cDNA encoding a cytochrome P-450 from an insecticide-resistant strain of the house fly, *Musca domestica*

(insect/cDNA cloning/deduced amino acid sequence/molecular evolution)

RENÉ FEYEREISEN*†, JOSETTE F. KOENER†, DAN E. FARNSWORTH*, AND DANIEL W. NEBERT†

*Department of Entomology, Oregon State University, Corvallis, OR 97331; and †Laboratory of Developmental Pharmacology, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892

Communicated by Robert L. Metcalf, November 28, 1988 (received for review August 5, 1988)

ABSTRACT **A cDNA expression library from phenobarbital-treated house fly (*Musca domestica*) was screened with rabbit antisera directed against partially purified house fly cytochrome P-450. Two overlapping clones with insert lengths of 1.3 and 1.5 kilobases were isolated. The sequence of a 1629-base-pair (bp) cDNA was obtained, with an open reading frame (nucleotides 81–1610) encoding a P-450 protein of 509 residues (M_r = 58,738). The insect P-450 protein contains a hydrophobic NH_2 terminus and a 22-residue cysteine-containing fragment near the COOH terminus that is known to bind the heme; both of these features have been found in the more than five dozen vertebrate P-450 proteins of which the sequences are presently known. Interestingly, the termination codon UAA may be contained in a putative polyadenylylation signal (AAUAAA) of the mRNA. This P-450 protein exhibits the most similarity (27% amino acid positional identity) with mammalian proteins of the P450III family but qualifies as a member of another family, which we propose to designate the P450VI gene family. This cDNA and deduced protein sequence should provide important information in the study of evolution of the P-450 gene superfamily, as well as provide an important probe for studying the regulation of insect P-450 and understanding the molecular genetics of insecticide resistance.**

Cytochrome P-450 monooxygenases are ubiquitous enzymes, involved in the metabolism of endogenous compounds and of xenobiotics, such as pesticides, drugs, and chemical carcinogens (1). Thirteen P-450 gene families comprised of more than five dozen members have been described to date, chiefly from mammals, yeast, and bacteria (2–4).

Microsomal P-450 enzymes from insects have many common features with those from other organisms (5, 6). For example, the electron transfer from NADPH through the flavoprotein NADPH:ferrihemoprotein oxidoreductase (EC 1.6.2.4) to multiple forms of P-450 proteins appears to be identical (5, 6). Insect P-450 monooxygenases are involved in the metabolism of insect hormones (7, 8), secondary plant chemicals (9), and synthetic insecticides (6). Insecticide resistance in insect pests of agricultural or medical importance has been attributed in many cases to a modified cytochrome P-450 system (6, 10). However, little is known about the molecular basis of P-450-linked resistance, because of the difficulty in purifying and characterizing P-450 from insect sources. Higher P-450 levels, as well as differences in P-450 catalytic and spectral characteristics, have been observed in insecticide-resistant strains of *Musca domestica* (11) and other insects (6). Such "qualitative" changes have also been observed in phenobarbital-treated house flies (12) and may result either from the increased expression of specific P-450 genes or from the selection of flies carrying a mutant P-450 gene that would confer a selective advantage during insecticide exposure.

In this paper we report the isolation and sequence of a P-450 cDNA from the house fly, *Musca domestica*. Our strategy was to use an insecticide-resistant strain that is characterized by high levels of P-450 enzymes and to treat the insects with phenobarbital to increase their P-450 content further (12, 13). This maximized the likelihood of obtaining cDNA clones from a cDNA expression library. The properties of the house fly P-450 described here indicate that this gene is a member of a separate P-450 gene family, which we suggest be designated the P450VI gene family.

MATERIALS AND METHODS

Insects. The R-Diazinon (Rutgers) strain of *Musca domestica* was provided by F. W. Plapp, Jr. (Texas A & M University). This strain, maintained under selection pressure with diazinon, is resistant to 2,2-bis(parachlorophenyl)-1,1,1-trichloroethane (DDT), organophosphorus, and carbamate insecticides (14). The flies were reared as described (12), and adults were treated with phenobarbital in the drinking water (0.2% wt/vol) for 3 days.

Enzymes and Chemicals. The kit for synthesis of the first and second cDNA strands was obtained from Boehringer Mannheim; Sequenase and avian myeloblastosis virus reverse transcriptase were purchased from United States Biochemical; restriction enzymes and polynucleotide kinase were bought from Bethesda Research Laboratories, and the cloning vector λZAP was from Stratagene. Oligonucleotides were synthesized on a Biosearch model 8700 DNA synthesizer and provided by J. E. Jones. Radioactive compounds were obtained from Amersham and New England Nuclear.

P-450 Protein Purification and Antibody Preparation. A major form of microsomal P-450 protein was partially purified from the abdomens of phenobarbital-treated female adult house flies. Microsomes were solubilized with 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) (1.25 mg/mg of microsomal protein), and the P-450 protein was then purified by ω-aminooctyl-agarose and DEAE-cellulose chromatography. The fraction eluting at 50 mM potassium phosphate buffer (pH 7.25), containing a major hemoprotein band ($M_r \simeq 57{,}000$) on $NaDodSO_4$/PAGE gels was used as the antigen for immunizing two rabbits. Western immunoblots (15) confirmed that the two antisera obtained [rabbit anti-(house fly)P-450] reacted with a microsomal protein of $M_r \simeq 57{,}000$.

cDNA Library Construction and Screening. From 4 g of abdomens from phenobarbital-treated flies, the RNA was extracted by the guanidinium isothiocyanate method and pelleted through 5.7 M CsCl (16). Poly(A)$^+$-enriched mRNA was isolated by oligo(dT)-cellulose chromatography, and cDNA was synthesized according to Gubler and Hoffman (17). The first-strand cDNA was synthesized with avian myeloblastosis virus reverse transcriptase and oligo(dT)-primed poly(A)$^+$-enriched RNA. This step was followed by

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

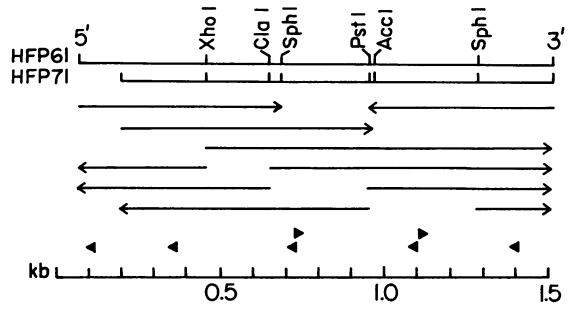

FIG. 1. Restriction map of cDNA clones HFP61 and HFP71 and sequencing strategy. Recognition sites for restriction enzymes were determined experimentally and confirmed by analysis of the sequence. Arrows indicate the direction and length of the sequenced subclones in M13mp18 or M13mp19. Arrowheads indicate the position and direction of synthetic oligonucleotides used as primers for sequencing.

RNase H and DNA polymerase I reactions for second-strand synthesis. The double-stranded cDNA was made blunt-ended with T4 DNA polymerase, methylated with *Eco*RI methylase, and ligated to *Eco*RI linkers. After *Eco*RI digestion, cDNA larger than 600 base pairs (bp) was selected on a Sepharose CL-4B column and ligated to λZAP arms (15). The ligated DNA was packaged *in vitro* with Gigapack gold (Stratagene), and 7×10^5 plaque-forming units were obtained on *Escherichia coli* BB4 cells. Duplicate filter-lifts of the library were screened with the two rabbit anti-(house fly)P-450 antibodies. The immunoscreening was done using horseradish peroxidase-conjugated anti(rabbit)IgG as described (15). The positive phages were purified through two additional rounds of plaque screening and were amplified. The inserts were excised from their λZAP vector and recircularized in the presence of helper phage R408 to form pBluescript phagemids. The recombinant clones were mapped, and appropriate fragments were subcloned in the M13mp18 or M13mp19 plasmid vectors.

DNA Sequencing. We sequenced double-stranded pBluescript clones or M13 subclones by the dideoxynucleotide method of Sanger *et al.* (18) with [α-^{35}S]thiodATP, dITP, dGTP, and Sequenase. The reactions were primed with universal M13 primer or oligonucleotides of 14–17 bases.

Northern (RNA) Blot. Poly(A)$^+$-enriched RNA was denatured with formaldehyde and formamide and separated by electrophoresis on an agarose gel containing formaldehyde; RNA ladders (Bethesda Research Laboratories) were used as markers. The RNA was then transferred to nitrocellulose. Pure P-450 cDNA insert was nick-translated with [α-^{32}P]-dCTP and added for hybridization in 50% formamide (15).

mRNA Sequencing. An oligomer of 24 nucleotides complementary to the 5′ end of the P-450 cDNA clone was synthesized and end-labeled with [γ-^{32}P]ATP via polynucleotide kinase (15). After purification on a Nensorb (New England Nuclear) column, 5 ng of the labeled oligonucleotide was hybridized in siliconized tubes to 12 μg of poly(A)$^+$-enriched RNA from phenobarbital-treated house flies. The mRNA was then sequenced by the dideoxynucleotide method with avian myeloblastosis virus reverse transcriptase (19).

RESULTS

Cloning and Sequencing Strategies. The cDNA expression library that we had prepared from abdomens of phenobarbital-treated house flies was screened with the rabbit polyclonal antibodies against partially purified house fly P-450 protein; this resulted in five positive clones. The two clones that reacted most strongly with both antisera (HFP61 and HFP71) were selected for further analysis. After plaque purification, the inserts were rescued into pBluescript plasmids, and the lengths and restriction maps of these inserts were obtained. Clone HFP61 is 1452 nucleotides long and overlaps at the 3′ end with the 1345-nucleotide clone HFP71. The restriction map (Fig. 1) obtained with the enzymes *Acc* I, *Cla* I, *Pst* I, *Sph* I, and *Xho* I was used to design our subcloning and sequencing strategies. Fragments cloned unidirectionally into plasmids M13mp18 and M13mp19 were sequenced completely.

Nucleotide Sequence. Clone HFP61 had an open reading frame of 1430 nucleotides, a TAA stop codon, and a 3′-

```
                                                                                                                    100
ACCTAAGCTATTCGAGCTAGGTGTGATCGTTATCCACGTGAAAGAAGAGAATAAAAAATTAGCGAAGCAATAAAAAAATTATGGATTTCGGTTCATTTCTTTTGTATGCCCTTGGGGTGTTG
                                                                                                      M  D  F  G  S  F  L  L  Y  A  L  G  V  L
                                                                                                                                  200
GCCTCTTTGGCCTTGTATTTTGTGCGCTGGAATTTCGGCTACTGGAAGAGGCGTGGCATACCCCATGAAGAACCCCATTTGGTGATGGGCAATGTCAAGGGTTTGAGGAGTAAATATCAC
 A  S  L  A  L  Y  F  V  R  W  N  F  G  Y  W  K  R  R  G  I  P  H  E  E  P  H  L  V  M  G  N  V  K  G  L  R  S  K  Y  H
                                                                                300
ATTGGTGAAATTATTGCCGATTATTATCGGAAATTTAAGGGATCGGACCCTTTGCCGGGCATTTTCTTGGGCCATAAACCCGCTGCTGTCGTATTGGACAAGGAGCTAAGAAAACGGGTT
 I  G  E  I  I  A  D  Y  Y  R  K  F  K  G  S  D  P  L  P  G  I  F  L  G  H  K  P  A  A  V  V  L  D  K  E  L  R  K  R  V
                                                                  400
CTCATTAAGGATTTTTCCAATTTTGCCAATCGTGGTCTGTATTACAATGAAAAAGATGATCCGCTGACCGGCCACTTGGTAATGGTGGAGGGTGAAAAATGGCGTAGCCTCAGGACAAAA
 L  I  K  D  F  S  N  F  A  N  R  G  L  Y  Y  N  E  K  D  D  P  L  T  G  H  L  V  M  V  E  G  E  K  W  R  S  L  R  T  K
                                500                                                                                 600
CTCTCACCCACCTTTACGGCTGGAAAAATGAAATATATGTACAACACGGTGCTGGAGGTTGGACAACGCCTGCTCGAGGTGATGTATGAAAAGCTGGAGGTCTCCTCGGAATTGGATATG
 L  S  P  T  F  T  A  G  K  M  K  Y  M  Y  N  T  V  L  E  V  G  Q  R  L  L  E  V  M  Y  E  K  L  E  V  S  S  E  L  D  M
                                                                                                  700
CGTGATATTTTGGCCCGTTTCAATACCGATGTCATTGGCAGTGTGGCCTTTGGTATTGAATGCAATAGCCTGAGGAATCCCCACGATCGTTTTCTGGCCATGGGAAGGAAATCCATTGAA
 R  D  I  L  A  R  F  N  T  D  V  I  G  S  V  A  F  G  I  E  C  N  S  L  R  N  P  H  D  R  F  L  A  M  G  R  K  S  I  E
                                                                                                                  800
GTGCCCAGACATAATGCCCTGATTATGGCCTTCATCGATAGCTTCCCAGAGTTGTCAAGGAAACTGGGCATGCGCGTTTTGCCAGAAGATGTCCATCAATTTTTCATGTCCTCCATCAAG
 V  P  R  H  N  A  L  I  M  A  F  I  D  S  F  P  E  L  S  R  K  L  G  M  R  V  L  P  E  D  V  H  Q  F  F  M  S  S  I  K
                                                                                900
GAAACGGTGGACTATCGGGAGAAGAACAATATTCGGCGCAATGATTTTTTGGATCTCGTCCTGGATTTGAAAAATAATCCGGAGAGTATCTCCAAGCTGGGGGGCTTGACATTCAATGAA
 E  T  V  D  Y  R  E  K  N  N  I  R  R  N  D  F  L  D  L  V  L  D  L  K  N  N  P  E  S  I  S  K  L  G  G  L  T  F  N  E
                                                1000
TTGGCTGCCCAGGTGTTTGTATTCTTCCTGGGGGGCTTTGAGACCTCCTCTTCGACAATGGGGTTTGCCCTCTATGAATTGGCCCAAAATCAGCAGCTGCAGGATCGTCTACGCGAGGAG
 L  A  A  Q  V  F  V  F  F  L  G  G  F  E  T  S  S  S  T  M  G  F  A  L  Y  E  L  A  Q  N  Q  Q  L  Q  D  R  L  R  E  E
                                1100                                                                                1200
GTTAATGAGGTGTTTGATCAGTTCAAGGAGGATAACATTAGCTATGATGCTCTCATGAATATACCCTATTTGGATCAGGTATTGAATGAAACACTCCGCAAATACCCCGTCGGCGTCGGC
 V  N  E  V  F  D  Q  F  K  E  D  N  I  S  Y  D  A  L  M  N  I  P  Y  L  D  Q  V  L  N  E  T  L  R  K  Y  P  V  G  V  G
                                                                                                  1300
AGTGCACTGACCCGCCAGACCTTGAACGACTATGTGGTGCCGCACAATCCCAAATATGTGTTGCCAAAGGGCACCCTGGTCTTCATACCCGTCTTGGGCATACACTATGACCCCGAACTG
 S  A  L  T  R  Q  T  L  N  D  Y  V  V  P  H  N  P  K  Y  V  L  P  K  G  T  L  V  F  I  P  V  L  G  I  H  Y  D  P  E  L
                                                                                1400
TACCCCAATCCCGAAGAATTTGATCCCGAACGTTTCTCCCCTGAAATGGTGAAACAGCGTGACTCCGTCGATTGGCTGGGTTTTGGTGATGGTCCGCGCAATTGTATTGGCATGCGTTTT
 Y  P  N  P  E  E  F  D  P  E  R  F  S  P  E  M  V  K  Q  R  D  S  V  D  W  L  G  F  G  D  G  P  R  N  C  I  G  M  R  F
                                                                1500
GGTAAAATGCAATCACGTTTGGGTTTGGCTTTGGTTATCAGGCATTTCCGTTTTACGGTATGTTCCCGCACCGATATACCAATGCAAATTAATCCCGAATCGCTGGCATGGACACCGAAA
 G  K  M  Q  S  R  L  G  L  A  L  V  I  R  H  F  R  F  T  V  C  S  R  T  D  I  P  M  Q  I  N  P  E  S  L  A  W  T  P  K
                                                1600
AATAATCTGTATTTAAATGTTCAAGCTATAAGAAAGAAAATTAAATAAATCGAAGGATATATGTAAG
 N  N  L  Y  L  N  V  Q  A  I  R  K  K  I  K  *
```

FIG. 2. Nucleotide and deduced amino acid sequence of house fly P-450 cDNA. The first 177 nucleotides were determined by mRNA sequencing. The initiation codon and putative polyadenylylation signal are boxed in black. The invariant residues found in all P-450 proteins (for further discussion, see text) are boxed.

untranslated region of 19 nucleotides (Fig. 2). Interestingly, the stop codon appears to be an integral part of a putative polyadenylylation signal, although no poly(A) was observed. This combination of the termination codon UAA and AAUAAA sequences has been found at the 3′ end of other mRNAs—e.g., human superoxide dismutase (20). Northern blot analysis (data not shown) revealed the presence of a single mRNA band [1.9 kilobases (kb)] hybridizing with nick-translated insert DNA from clone HFP61; we therefore chose to obtain the 5′ sequence and initiation codon by sequencing the mRNA directly with a ^{32}P-labeled synthetic oligonucleotide primer. This procedure provided 80 nucleotides of the 5′-nontranslated region plus 97 bases downstream from the initiation codon. The mRNA sequence was subsequently confirmed by sequencing the 5′ end of an additional cDNA clone that hybridized to clone HFP61. The sequence around the ATG (adenine at −3 and guanine at +4) is found in ≈35% of eukaryotic mRNAs and apparently provides an optimal context for initiation of translation (21).

The complete sequence (Fig. 2) thus contained an open reading frame of 1530 nucleotides, coding for a protein of 509 residues (M_r = 58,738). This agrees with the molecular mass of the partially purified protein as estimated by $NaDodSO_4$/PAGE.

Comparison with Other P-450 Sequences. The sequence in Fig. 2 encodes a P-450 protein as evidenced by a conserved sequence (F--G---C-G) near the COOH terminus that has been found in all P-450 proteins to date and in no other protein in the sequence data bank (3). The cysteine in this conserved sequence is known to participate in heme binding (1, 2). We compared the deduced house fly P-450 amino acid sequence with more than five dozen other P-450 sequences, using Nelson and Strobel's alignment of 34 P-450 sequences (22) as a guide. Although 14 of 22 residues surrounding the heme-binding cysteine were identical in the house fly P-450 and all five members of the P450III gene family established to date, the overall amino acid positional identity was 26.9% (Fig. 3). Similarity with members of other P-450 gene families was even lower. The "ground rules" for establishing a P-450 gene family include characterization of a protein having <37% amino acid positional identity with P-450 proteins encoded by any other gene family (29). We therefore propose this house fly P-450 as a member of a separate P-450 family and suggest family VI. The proposed gene name would thus be *CYP6A1* (30).

The hydropathy profile of house fly P-450 was compared with proteins of the rat and human P450III family and proteins of the human and trout P450I family (Fig. 4). Greatest similarities can be seen between two proteins within the same family irrespective of species—e.g., rat pcn1 and human nf; human P_1 and trout IA1. Nonetheless, the P-450 proteins from the VI and III families have retained a remarkable overall similarity. The hydrophobic membrane-anchor "signal peptide" of >20 amino acids at the NH_2 terminus—which has been found in all vertebrate microsomal P-450 proteins—is present in the insect P450VIA1 protein.

Although the 137 residues shared between the house fly P450VI and vertebrate P450III family are dispersed throughout the enzyme molecule, there appear to be clusters of greater similarity, which may correspond to conserved func-

```
                                              30                                  60
H.fly VIA1   MDFGSFLLYALGVL-A-SL-ALYFVRWNF-GYWKRRGIPHEEPHLVMGNVKGLRSKYHIGEI-IADYYRKFKGSDPLPGIFLGHKPAAVVL
Rat   pcn1   MDLLSALTLETWVLLAVVLVLLYGFGTRTHGLFKKQGIPGPKPLPFFGTV--LN--YYMGLW-KFDVECHKKYGKIW-GLFDGQMPLFAIT
Rat   pcn2   MDLLSALTLETWVLLAVILVLLYRLGTHRHGIFKKQGIPGPKPLPFLGTV--LN--YYKGLG-RFDMECYKKYGKIW-GLFDGQTPVFAIM
Human   nf   MALIPDLAMETWLLLAVSLVLLYLYGTHSHGLFKKLGIPGPTPLPFLGNI--LS--YHKGFC-MFDMECHKKYGKVW-GFYDGQQPVLAIT
Human  HLp   MALIPDLAMETWLLLAVSLVLLYLYGTHSHGLFKKLGIPGPTPLPFLGNI--LS--YHKGFC-MFDMECHKKYGKVW-GFYDGQQPVLAIT
Rabbit  3c   MDLIFSL--ETWVLLAASLVLLYLYGTSTHGLFKKMGIPGPTPLPFIGTI--LE--YRKVINWDFDIECRKKYGKMW-GLFDGRQPLMVIT

               90                                    120                                  150
  VIA1       DKELRKRVLIKD-FSNFANRGLYYNEKDDPLTGHLVMVEGEKWRSLRTKLSPTFTAGKMKYMYNTVLEVGQRLLEVMY-EKLEVSSELDMR
  pcn1       DTEMIKNVLVKECFSVFTNRRDFGPVGIM--GKAVSVAKDEEWKRYRALLSPTFTSGRLKEMFPIIEQYGDILVKYLKQEA-ETGKPVTMK
  pcn2       DTEMIKNVLVKECFSVFTNRRDFGPVGIM--GKAVSVAKDEEWKRYRALLSPTFTSGRLKEMFPIIEQYGDILVKYLKQEA-ETGKPVTMK
  nf         DPDMIKTVLVKECYSVFTNRRPFGPVGFM--KSAISIAEDEEWKRLRSLLSPTFTSGKLKEMVPIIAQYGDVLVRNLRREA-ETGKPVTLK
  HLp        DPDMIKLVLVKECYSVFTNREPFGPVGFM--KSAISIAEDEEWKRLRSLLSPTFTSGKLKEMVPIIAQYGDVLVRNLRRER-ETGKPVTLK
  3c         DPDMIKTVLVKECYSVFTNRRSFGPVGFM--KKAVSISEDEDWKRVRTLLSPTFTSGKLKEMLPIIAQYGDVLVKNLRQEA-EKASPSTLK

               180                                   210                                  240
  VIA1       DILARFNTDVIGSVAFGIECNSLRNPHDRFLAMGRKSIEVPRH--NAL-IMAFIDSFPELSRKLGMRVLPEDVHQFFMSSIKETVDYREK
  pcn1       KVFGAYSMDVITSTSFGVNVDSLNNPKDPFVEKTKKLLRFDFFDPLFLSVVLFPFLTPIYEM-LNICMFPKDSIEFFKKFVYRMKETRLD
  pcn2       KVFGAYSMDVITSTSFGVNVDSLNNPKDPFVEKTKKLLRFDFFDPLFLSVVLFPFLTPIYEM-LNICMFPKDSIAFFQKFVHRIKETRLD
  nf         DVFGAYSMDVITSTSFGVNIDSLNNPQDPFVENTKKLLRFDFLDPFFLSITVFPFLIPILEV-LNICVFPREVTNFLRKSVKRMKESRLE
  HLp        DVFGAYSMDVITSSSFGVNVDSLNNPQDPFVENTKKLLRFDFLDPFFLSITVFPFLIPILEV-LNICVFPREVTNFLRKAVKRMKESRLE
  3c         EIFGAYSMDVITGTSFGVNIDSLRNPQDPFVKNVRRLLKFSFFDPLLLSITLFPFLTPIFEA-LHISMFPKDVMDFLKTSVEKIKDDRLK

                  270                                  300                                330
  VIA1       -NNIRRNDFLDLVLDLKNNPESISKLGGLTFNELAAQVFVFFLGGFETSSSTMGFALYELAQNQQLQDRLREEVNEVFDQFKEDNISYDA
  pcn1       SVQKHRVDFLQLMMNAHNDSKDKESHTALSDMEITAQSIIFIFAGYEPTSSTLSFVLHSLATHPDTQKKLQEEIDRALPN-KAPPT-YDT
  pcn2       SKHKHRVDFLQLMLNAHNNSKDEVSHKALSDVEIIAQSVIFIFAGYETTSSTLSFVLYFLATHPDIQKKLQEEIDGALPS-KAPPT-YDI
  nf         DTQKHRVDFLQLMIDS-QNSKETESHKALSDLELVAQSIIFIFAGYETTSSVLSFIMYELATHPDVQQKLQEEIDAVLPN-KAPPT-YDT
  HLp        DTQKHRVDFLQLMIDSHKNSKETESHKALSDLELVAQSIIFIFAGYETTSSVLSFIMYELATHPDVQQKLQEEIDAVLPN-KAPPT-YDT
  3c         DKQKRRVDFLQLMINS-QNSKEIDSHKALDDIEVVAQSIIILFAGYETTSSTLSFIMHLLATHPDVQQKLQEEIDTLLPN-KELAT-YDT

                  360                                  390                                420
  VIA1       LMNIPYLDQVLNETLRKYPVGVGSALTRQTLNDYVVPHNPKYVLPKGTLVFIPVLGIHYDPELYPNPEEFDPERFSPEMVKQRDSVDWLG
  pcn1       VMEMEYLDMVLNETLRLYPIG--NRLERVCKKD--VEING-VFMPKGSVVMIPSYALHRDPQHWPEPEEFRPERFSKENKGSIDPYVYLP
  pcn2       VMEMEYLDMVLNETLRLYPIG--NRLERVCKKD--IELDG-LFIPKGSVVTIPTYALHHDPQHWPKPEEFHPERFSKENKGSIHPYVYLP
  nf         VLQMEYLDMVVNETLRLFPIA--MRLERVCKKD--VEING-MFIPKGVVVMIPSYALHRDPKYWTEPEKFLPERFSKKNKDNIDPYIYTP
  HLp        VLQMEYLDMVVNETLRLFPIA--MRLERVCKKD--VEING-MFIPKGWVVMIPSYALHRDPKYWTEPEKFLPERFSKKNKDNIDPYIYTP
  3c         LVKMEYLDMVVNETLRLYPIA--GRLERVCKKD--VDING-TFIPKGTIVM-PTYALHRDPQHWTEPDEFRPERFSKKNKDNINPYIYHP

                  450                                  480                               509
  VIA1       FGDGPRNCIGMRFGKMQSRLGLALVIRHFRFTVCSRTDIPMQINPESLAWTPKNNLYLNVQAIRKKIK
  pcn1       FGNGPRNCIGMRFALMNMKLALTKVLQNFSFQPCKETQIPLKLSRQGLLQ-PTKPIILKVVP-RDEIITGS
  pcn2       FGNGPRNCIGMRFALMNMKLALTKVLQNFSFQPCKETQIPLKLSRQAILE-PEKPIVLKVLP-RDAVINGA
  nf         FGSGPRNCIGMRFALMNMKLALIRVLQNFSFKPCKETQIPLKLSLGGLLQ-PEKPVVLKVES-RDGTVSGA
  HLp        FGSGPRNCIGMRFALMNMKLALIRVLQNFSFKPCKETQIPLKLSLGGLLQ-PEKPVVLKVES-RDGTVSGA
  3c         FGAGPRNCLGMRFALMNIKLALVRLMQNFSFKLCKETQVPLKLGKQGLLQ-PEKPIVLKVVS-ESGIIRGA
```

FIG. 3. Alignment of house fly P450VIA1 with the known P450III sequences. Rat liver P450 pcn1 (23) and P450 pcn2 (24), human liver P450 nf (25, 26) and P450 HLp (27), and rabbit liver P450 3c (28) sequences are taken from the cited references. Numbering on top corresponds to the house fly P-450 residues. A * denotes a residue shared between the VI family and all members of the III family.

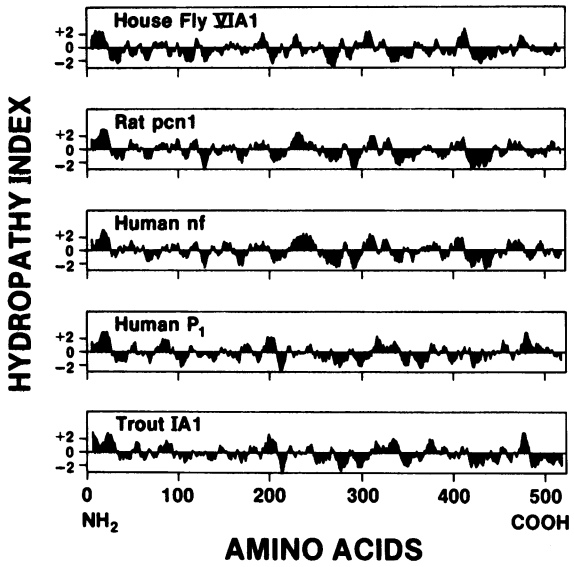

FIG. 4. Comparison of the hydropathy index for the house fly P450VIA1, rat P450 pcn1, human P450 nf, human $P_1$450, and trout P450IA1 proteins. A sliding window of nine amino acids was used (31). The mammalian sequences are taken from published reports (23, 25, 32, and 33, respectively).

tional domains of the protein. These clusters of identical and mostly identical residues between families VI and III include the following sequences: L S P T F T (residues 135–140), S L R N P H D R F (residues 197–205), N E T L R (residues 363–367), P E R F S (residues 423–427), and F G D G P R N C I G M R F (residues 442–454). The Gly-78, Phe-191, Gly-306, Glu-364, Arg-367, Pro-418, Phe-442, Gly-445, Cys-449, Gly-451, and Phe-470 have been found to be invariant across all eukaryotic and prokaryotic P-450 proteins with known sequences in the thirteen gene families (4).

DISCUSSION

The house fly P450VIA1 appears to be a typical microsomal P-450 protein, based on its molecular mass, hydrophobic NH_2 terminus, and conserved heme-binding region.

Alignment of sequences revealed that the house fly P450VIA1 protein is homologous to vertebrate cytochrome P-450 proteins, having the highest similarity (26.9%) with the P450III family. Family III has diverged from other drug-metabolizing P-450 genes by a gene-duplication event about 1100 million yr before present (Mybp) (22). Values for the unit evolutionary period (UEP, time required for a 1% difference in amino acid sequence to occur) are imprecise due to the paucity of P-450 structures available outside the class Mammalia, so it is only possible to say that family VI diverged from family III between 300 and 700 Mybp. Whether family VI arose by a gene-duplication event in a common ancestor or by the divergence of protostomes and deuterostomes thus remains uncertain.

Some aspects of the amino acid structure point to significant differences between family III and family VI. In an active-site model of cytochrome LM2 (P450IIB1), Phe-377 to Tyr-380 has been identified as a tetrapeptide part of the active site situated trans to the thiolate, and the two phenyl rings would accommodate substrates in a sandwich manner (34). This site may be modified in house fly P450VIA1 with a His-389 to Tyr-393 pentapeptide. In the P450III family, neither residue is conserved, which may argue for differences in catalytic activity between family III and VI.

A hydrophobic stretch preceding Thr-309 is found in all P-450 proteins as well as in house fly P450VIA1, but this residue was not found in rat P-450 pcn1 (P450IIIA1; ref. 23), where it is replaced by proline. Analysis of the crystal structure of P450cam (P450CIA1; ref. 35) indicates that this threonine (Thr-252 of P450cam) is part of the distal helix above the heme. The residues situated five and eight positions NH_2-ward were shown to contact the camphor molecule in P450cam and therefore are postulated to play a role in substrate specificity. In house fly P450VIA1, the residues in these positions, Val-301 and Leu-304, are different from their homologues in the P450III family (Ile and Phe) and are not found at that position in any other P-450 proteins.

The conserved region between Asp-184 and Phe-205 may be involved in the interaction of the P-450 protein with the flavoprotein NADPH:ferrihemoprotein oxidoreductase (3). The house fly reductase has been purified, and it has a low but measurable degree of immunoreactivity with the flavoprotein from vertebrate sources (36). The structural requirements of the flavoprotein-cytochrome domains that interact may thus be conserved in all microsomal P-450 proteins.

It should now be possible to utilize this cloned P-450 cDNA as a molecular probe to clarify the chromosomal location of the structural gene and its regulation in insecticide-resistant and susceptible strains. We have confirmed that P450VIA1 mRNA levels are induced by phenobarbital treatment in the R-Diazinon strain (J.F.K. and R.F., unpublished data). The catalytic activity of house fly P450VIA1 and the multiplicity of P-450 genes in insects also remain to be established. We believe that additional members of the P450VI or other P-450 gene families are present in insects. This multiplicity of P-450 genes, best documented in vertebrates (1–4), could prove important in the evolution of insecticide resistance. It is not known whether the selection of a mutation that increases expression of some P-450 genes or a mutation that modifies the catalytic activity of some P-450 proteins is responsible for P-450-associated cases of insecticide resistance. Increased expression of esterase genes in insecticide-resistant *Myzus persicae* and esterase gene amplification in a resistant strain of *Culex quinquefasciatus* have been described (37, 38), but there is no evidence for gene amplification in cases of resistance where P-450 is involved.

In conclusion, the study of insect P-450 structure and multiplicity should improve our understanding of the evolution of these enzymes, and the availability of a molecular probe should lead to important insights into the genetic mechanisms of insecticide resistance.

This work was supported in part by National Institutes of Health Grant GM39014 and by an Intergovernmental Personnel Act Assignment. This is paper no. 8712 from the Oregon Agricultural Experiment Station.

1. Ortiz de Montellano, P. R. (1986) *Cytochrome P-450: Structure, Mechanism, and Biochemistry* (Plenum, New York), p. 556.
2. Nebert, D. W. & Gonzalez, F. J. (1987) *Annu. Rev. Biochem.* **56**, 945–993.
3. Nebert, D. W., Jones, J. E., Owens, J. & Puga, A. (1988) in *Oxidases and Related Redox Systems*, eds. King, T., Mason, H. & Morrison, M. (Liss, New York), pp. 557–578.
4. Nebert, D. W. & Gonzalez, F. J. (1988) in *Frontiers in Biotransformation*, eds. Ruckpaul, K. & Rein, H. (Akademie-Verlag, Berlin), Vol. 3, in press.
5. Hodgson, E. (1985) in *Comprehensive Insect Physiology, Biochemistry and Pharmacology*, eds. Kerkut, G. A. & Gilbert, L. I. (Pergamon, Oxford), Vol. 11, pp. 225–321.
6. Agosin, M. (1985) in *Comprehensive Insect Physiology, Biochemistry and Pharmacology*, eds. Kerkut, G. A. & Gilbert, L. I. (Pergamon, Oxford), Vol. 12, pp. 647–712.
7. Feyereisen, R. & Durst, F. (1978) *Eur. J. Biochem.* **88**, 37–47.
8. Feyereisen, R., Pratt, G. E. & Hamnett, A. F. (1981) *Eur. J. Biochem.* **118**, 231–238.

9. Dowd, P. F., Smith, C. M. & Sparks, T. C. (1983) *Insect Biochem.* **13,** 453–468.
10. Brattsten, L. B., Holyoke, C. W., Leeper, J. R. & Raffa, K. F. (1986) *Science* **231,** 1255–1260.
11. Tate, L. G., Plapp, F. W. & Hodgson, E. (1973) *Chem.-Biol. Interact.* **6,** 237–247.
12. Moldenke, A. F. & Terriere, L. C. (1981) *Pestic. Biochem. Physiol.* **16,** 222–230.
13. Feyereisen, R. (1983) *Pestic. Biochem. Physiol.* **19,** 262–269.
14. Forgash, A. J., Cook, B. J. & Riley, R. C. (1962) *J. Econ. Entomol.* **55,** 544–551.
15. Davis, L. G., Dibner, M. D. & Battey, J. F. (1986) *Basic Methods in Molecular Biology* (Elsevier, New York), p. 388.
16. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
17. Gubler, U. & Hoffman, B. J. (1983) *Gene* **25,** 263–269.
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
19. Geliebter, J. (1987) *Focus* **9,** 5–8.
20. Sherman, L., Dafni, N., Lieman-Hurwitz, J. & Groner, Y. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 5465–5469.
21. Kozak, M. (1986) *Cell* **44,** 283–292.
22. Nelson, D. R. & Strobel, H. W. (1987) *Mol. Biol. Evol.* **4,** 572–593.
23. Gonzalez, F. J., Nebert, D. W., Hardwick, J. P. & Kasper, C. B. (1985) *J. Biol. Chem.* **260,** 7435–7441.
24. Gonzalez, F. J., Song, B. J. & Hardwick, J. P. (1986) *Mol. Cell. Biol.* **6,** 2969–2976.
25. Beaune, P. H., Umbenhauer, D. R., Bork, R. W., Lloyd, R. S. & Guengerich, F. P. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8064–8068.
26. Gonzalez, F. J., Schmid, B. J., Umeno, M., McBride, O. W., Hardwick, J. P., Meyer, U. A., Gelboin, H. V. & Idle, J. R. (1988) *DNA* **7,** 79–86.
27. Molowa, D. T., Schuetz, E. G., Wrighton, S. A., Watkins, P. B., Kremers, P., Mendez-Picon, G., Parker, G. A. & Guzelian, P. S. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5311–5315.
28. Dalet, C., Clair, P., Daujat, M., Fort, P., Blanchard, J. M. & Maurel, P. (1988) *DNA* **7,** 39–46.
29. Nebert, D. W., Adesnik, M., Coon, M. J., Estabrook, R. W., Gonzalez, F. J., Guengerich, F. P., Gunsalus, I. C., Johnson, E. F., Kemper, B., Levin, W., Phillips, I. R., Sato, R. & Waterman, M. R. (1987) *DNA* **6,** 1–11.
30. Nebert, D. W., Nelson, D. R., Adesnik, M., Coon, M. J., Estabrook, R. W., Gonzalez, F. J., Guengerich, F. P., Gunsalus, I. C., Johnson, E. F., Kemper, B., Levin, W., Phillips, I. R., Sato, R. & Waterman, M. R. (1989) *DNA* **8,** in press.
31. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
32. Jaiswal, A. K., Gonzalez, F. J. & Nebert, D. W. (1985) *Science* **228,** 80–83.
33. Heilmann, L. J., Sheen, Y.-Y., Bigelow, S. W. & Nebert, D. W. (1988) *DNA* **7,** 379–387.
34. Schwarze, W., Jaeger, J., Janig, G. R. & Ruckpaul, K. (1988) *Biochem. Biophys. Res. Commun.* **150,** 996–1005.
35. Poulos, T. L., Finzel, B. C., Gunsalus, I. C., Wagner, G. C. & Kraut, J. (1985) *J. Biol. Chem.* **260,** 16122–16130.
36. Feyereisen, R. & Vincent, D. R. (1984) *Insect Biochem.* **14,** 163–168.
37. Devonshire, A. L., Searle, L. M. & Moores, G. D. (1986) *Insect Biochem.* **16,** 659–665.
38. Mouches, C., Pasteur, N., Berge, J. B., Hyrien, O., Raymond, M., Robert de Saint Vincent, B., de Silvestri, M. & Georghiou, G. P. (1986) *Science* **233,** 778–780.