Insect Molecular Biology (2001) **10**(3), 191–196

Cytochrome P450s *CYP6D3* and *CYP6D1* are part of a P450 gene cluster on autosome 1 in the house fly

S. Kasai and J. G. Scott

Department of Entomology, Comstock Hall, Cornell University, Ithaca, New York, USA

Abstract

The P450 monooxygenases of insects are important in the metabolism of numerous endogenous and exogenous compounds. However, identity of the P450 isoform(s) involved in these reactions is rarely known. A critical first step in the identification of important P450s is the cloning and sequencing of their genes. Toward this goal we report the genomic sequence of a new cytochrome P450, termed *CYP6D3*, from the house fly, *Musca domestica*. *CYP6D3* is part of a P450 gene cluster located on chromosome 1 and is located upstream of a related gene, *CYP6D1*. The similar genetic structures of *CYP6D3* and *CYP6D1* (5 exons and 4 introns of similar length) suggest one of these genes may have been the result of a duplication event. The CYP6D3 deduced amino acid sequence indicates a protein with 518 amino acids and a molecular weight of 59.3 kDa. The CYP6D3 protein is most similar to house fly CYP6D1 (78%) and Cyp6D2 (56%) from *Drosophila melanogaster*. The deduced amino acid sequences of CYP6D3 and CYP6D1 are identical at the Helix I and heme binding regions.

Keywords: *Musca domestica*, P450 monooxygenases, gene cluster, Insecta.

Introduction

Cytochromes P450 (P450s) comprise a large gene superfamily with each cytochrome P450 designated CYP followed by a family, subfamily and isoform number (Nelson *et al*., 1996). The cytochrome P450 monooxygenases have important roles in the metabolism of endogenous and exogenous compounds (Mansuy, 1998). The large number of substrates metabolized by monooxygenases is due to the presence of multiple P450 isoforms in each species (e.g. eighty-six in *Drosophila melanogaster* (Adams *et al*., 2000)), and the fact that each P450 may have several substrates (Rendic & Di Carlo, 1997). Because the P450s in any one species may have overlapping substrate specificity, it remains difficult to identify the functions of individual P450s. The diversity of P450s within (and usually between) species is quite remarkable. For example, the deduced amino acid sequences of the eighty-six *Drosophila* predicted P450s show no two sequences that are more than 84% identical (Cyp6A17 and Cyp6A23; calculations not shown). For insects the importance of monooxygenases in the metabolism of many substrates is known, however the P450 isoform(s) involved have rarely been identified. A critical first step in the identification of important P450s is the cloning and sequencing of their genes.

CYP6D1 is a cytochrome P450 isolated from house fly, which metabolizes pyrethroid insecticides and other xenobiotics (Scott, 1996). In the pyrethroid resistant LPR strain there is enhanced transcription of *CYP6D1* (Liu & Scott, 1998), leading to increased expression of the protein and thus increased detoxification (and resistance) (Scott, 1999), however the mechanism of enhanced *CYP6D1* transcription in the LPR strain is not known. As a first step towards better understanding this mechanism we sought to obtain the 5′ flanking sequence of *CYP6D1*, beyond the 800 bp that were already known (Scott *et al*., 1999).

We screened a house fly genomic library and obtained a 14 kb clone that contained the 5′ flanking sequence of *CYP6D1 and* a new P450, *CYP6D3*. The relatedness of this new gene, in terms of its intron/exon structure and deduced amino acid sequence, to other P450s is discussed.

Results and discussion

Using the 5′-flanking sequence of *CYP6D1* as a probe, a positive clone was isolated from the house fly genomic library. The entire sequence of the 14 kb insert from this clone was determined (Fig. 1). The sequence of this clone contains a new cytochrome P450 gene (named *CYP6D3*

Received 11 August 2000; accepted after revision 7 December 2000. Correspondence: Dr Jeffrey G. Scott, Department of Entomology, Comstock Hall, Cornell University, Ithaca, NY 14853–0901 USA. Tel.: (607) 255-7340; fax: (607) 255 0939; e-mail: jgs5@cornell.edu

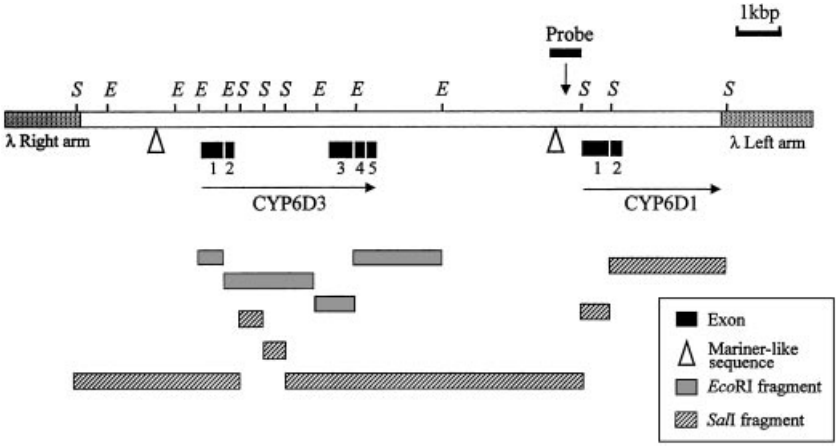

Figure 1. Diagram of the house fly genomic clone (on chromosome 1) containing *CYP6D3*. Restriction fragments used to determine the sequence of the clone are indicated. The accession number for the entire sequence is AF200191.

by the P450 nomenclature committee, Fig. 2) as well as a partial sequence of *CYP6D1* (first 2 exons and introns, Fig. 1). The sequence of *CYP6D1* obtained from this clone matches the sequence previously determined (Scott *et al.*, 1999). Given that a previous study (using allele specific PCR) found that *CYP6D1* was on autosome 1 (Liu *et al.*, 1995) and that *CYP6D1* appears to be a single copy gene (Tomita *et al.*, 1995), we conclude that this clone represents a region of house fly chromosome 1.

The predicted length of the *CYP6D3* open reading frame (ORF) is 1557 bp (Fig. 2) with a nucleotide sequence that is 77.1% similar to *CYP6D1v1* (Tomita & Scott, 1995) (Table 1). These two genes have other features in common as well. Both *CYP6D1* and *CYP6D3* have 5 exons and 4 introns. Furthermore, the lengths of each exon and intron are very similar between these genes (Table 1). For example, lengths of the exons are exactly the same between *CYP6D1* and *CYP6D3*, except for the last exon (*CYP6D3* is 6 bp longer than *CYP6D1*). The similarities of exons between *CYP6D3* and *CYP6D1* are relatively high and ranged from 71.8 to 78.6%, while similarities of the introns are relatively low (between 43.3 and 63.3%). The transcription initiation site (TIS) of *CYP6D1* was previously identified at −86 nt from the translation start site (Scott *et al.*, 1999) within a conserved arthropod promoter element (Cherbas & Cherbas, 1993). An identical sequence (TCAGT) was identified at −97 to −93 from the translation start site of *CYP6D3* (Fig. 2) suggesting this may be its TIS. Expression of *CYP6D3* appears to be developmentally regulated because preliminary Northern blots indicated the presence of a ~2.0 kb transcript that was present in adults, but not in eggs (Kasai & Scott, 2001).

The CYP6D3 deduced amino acid sequence predicts a protein with 518 amino acids and a molecular weight of 59.3 kDa. The deduced amino acid sequences of CYP6D3 and CYP6D1 are 79% identical (Table 2). Furthermore, sequence of the putative Helix I (aa 318–323; AGSETT) and heme binding (aa 454–466; FGEGPRHCIAQRM) regions were completely identical. Relative to CYP6D3, the four most similar P450s belong to family 6: CYP6D1, Cyp6D2, Cyp6D4 and Cyp6D5. *Cyp6D2*, *Cyp6D4* and *Cyp6D5* have been identified from the genome sequence of *D. melanogaster* (Adams *et al.*, 2000; Nelson, 2000). Cyp6D2 shows the highest amino acid similarity to CYP6D1, and CYP6D3, with per cent identities of 60 and 57.4%, respectively (Table 2).

The high degree of similarity, between *CYP6D1* and *CYP6D3* suggests that one of these genes arose relatively recently from a gene duplication event. In addition, there are fragments of mariner-like elements in both the 5′- and 3′-flanking sequence of *CYP6D3* (and the 5′-flanking sequence of *CYP6D1*) (Figs 1 and 2). Mariner elements are a type of transposable element found in *D. mauritiana* that are also capable of transformation in *D. melanogaster* and *Lucilia cuprina* (O'Brochta & Atkinson, 1996). It is possible that a *CYP6D1/D3* duplication event was facilitated by means of such elements. The cluster (i.e. not separated by other genes) of *CYP6D3* and *CYP6D1* is not unusual, because P450 gene clusters are known in *Drosophila* (Dunkov *et al.,* 1996; Maitra *et al.*, 1996; Adams *et al.*, 2000; Nelson, 2000) and on autosome 5 of house fly (Cohen & Feyereisen, 1995). Given that *CYP6D1* and *CYP6D3* are more closely related to each other than to any other known P450s, this suggests that the putative gene duplication event occurred relatively recently (i.e. well after the divergence of the calyptratae and acalyptratae).

Comparison of seventeen CYP6 genomic sequences reveals a limited number of insertion sites within this family (Fig. 3). There may be no introns (*Cyp6A13*) or as many as four (*CYP6D1*, *CYP6D3* and *Cyp6D5*). The length of some introns varies considerably (intron 2 of *CYP6D1*) while others are more similar (e.g. intron 4 of *CYP6D1*). Regardless of whether introns are ancestral or derived (so called 'intron early' and 'intron late' hypotheses), it has been suggested that introns may be useful for phylogenetic analysis (Venkatesh *et al.*, 1999). There is a high degree of similarity

Figure 2. The genomic and deduced amino acid sequences of *CYP6D3*. The sequence of the 2nd intron has been abbreviated. Mariner-like sequences are indicated in bold. Barbie box-like sequences are boxed. Putative polyadenylation sequences are underlined.

Table 1. Comparison of *CYP6D1* and *CYP6D3* from the LPR and Edinburgh strains of house fly, respectively

Sequence	Length (bp) *CYP6D1v1*	*CYP6D3*	% similarity
cDNA	1551	1557	77.1
5′-upstream	100	100	54.3
5′-upstream	300	300	50.4
5′-upstream	500	500	46.8
Exon 1	527	527	78.6
Exon 2	162	162	71.8
Exon 3	449	449	78.4
Exon 4	249	249	76.0
Exon 5	164	170	75.8
Intron 1	74	62	55.6
Intron 2	2377	2117	43.3
Intron 3	66	59	63.3
Intron 4	64	65	50.8
3′-downstream	157	157	56.3

Per cent similarities are based on an analysis using the DNASTAR MegAlign program using the Jotun Hein method.

Table 2. Per cent identity of the amino acid sequences of P450s in subfamily 6D

	CYP6D1v1	CYP6D1v2	Cyp6D2	CYP6D3	Cyp6D4
CYP6D1v2	98.5				
Cyp6D2	59.8	59.6			
CYP6D3	78.6	79.0	57.4		
Cyp6D4	52.2	51.8	50.5	51.6	
Cyp6D5	51.7	51.3	49.4	51.7	52.4

Amino acid sequences were previously published or obtained from Dr David Nelson's P450 Web site (http://drnelson.utmem.edu/CytochromeP450.html). Per cent identities were calculated using DNAStar software via the Jotun Hein method of alignment.

in the position of introns between *CYP6A*, *CYP6B* and *CYP6D* subfamilies, suggesting that intron position and deduced amino acid sequences may give similar phylogenetic patterns. Conversely, relatedness of the CYP6D sequences was slightly different if amino acid (Table 2) or intron positions (Fig. 3) are considered. Unfortunately, a complete phylogenetic analysis would require many additional sequences and is beyond the scope of this paper. A more detailed analysis, utilizing the P450s identified from genome sequencing projects could offer interesting insights into the relatedness of P450s.

CYP6D1 cDNA has been sequenced from four pyrethroid susceptible strains of house flies and the pyrethroid resistant LPR strain (Tomita *et al*., 1995; Tomita & Scott,

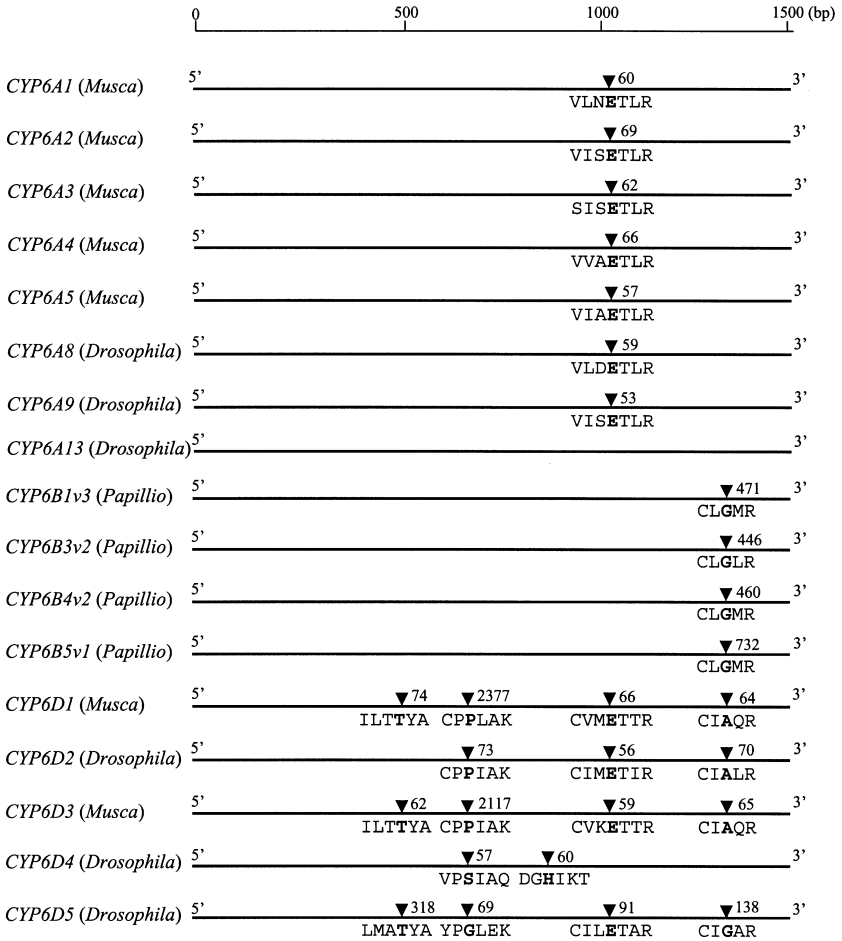

Figure 3. Comparison of the intron positions of seventeen family 6 P450s. The number of nucleotides in the intron is indicated above the arrow and the deduced amino acid sequence is given below the line with the amino acid in bold indicating the precise intron site.

1995). Comparison of the five *CYP6D1* alleles reveals that the deduced amino acid sequence from the LPR allele differs from that of the CS, aabys, ISK and Rutgers (strain not homozygous) alleles by 8, 11, 7 and 6–7 amino acids, respectively (Tomita *et al.*, 1995). Among them, 5 amino acids are the same in CS, aabys, ISK and Rutgers, but are different from LPR: Asp_{150} to Ala, Ile_{153} to Leu, Thr_{165} to Ser, Glu_{218} to Gln and Met_{227} to Ile (Tomita *et al.*, 1995). The observed amino acid substitutions occur at two highly variable regions among cytochromes P450 in family 6, and the changes at residues 218 and 227 are close to a proposed substrate binding region (Gotoh & Fujii-Kuriyama, 1989). The *CYP6D1* partial sequence found in our clone allows us to determine almost half of the CYP6D1 deduced amino acid sequence from the Edinburgh susceptible strain. The first 229 amino acids have seven differences from LPR, including the same amino acid substitutions found for the other four pyrethroid susceptible strains (Tomita *et al.*, 1995). Thus, pyrethroid susceptible strains of house fly from Japan (aabys), Europe (Edinburgh) and North America (others) have *CYP6D1* gene sequences that are more closely related to each other than any are to LPR. This suggests there is a selective advantage for LPR in having this particular allele.

Previous studies have suggested a link between the presence of a 15 bp insert close to the TIS (nt –15 to –29) in the *CYP6D1* 5′ flanking sequence, and higher levels of expression of this gene in the LPR strain (Scott *et al.*, 1999). Comparison of the *CYP6D1* 5′ flanking sequence from the Edinburgh strain supports this link, because the 15 bp insert is missing from this strain as it is from all other pyrethroid susceptible strains examined. The possible role that this insert plays in *CYP6D1* expression has been previously described (Scott *et al.*, 1999).

Barbie boxes are the DNA regulatory elements that may (He & Fulco, 1991), or may not (Shaw *et al.*, 1998) be involved in phenobarbital induction in bacteria. Phenobarbital induction in mammals appears to be controlled by other factors (Trottier *et al.*, 1995; Park *et al.*, 1996; Honkakoski & Negishi, 1997). Although phenobarbital induction has been mapped to chromosome 2 in house flies (Liu & Scott, 1997) and has been widely studied in insects, the role of Barbie boxes in this phenomenon remains unknown. We searched for putative Barbie boxes by identifying regions containing both the consensus core sequence (aaag) and a minimum of eight nucleotides (out of fifteen total) matching the overall consensus sequence (ATCAAAAGCTGGAGG) (Liang *et al.*, 1995). We found six Barbie box like sequences in the 2.7 kb upstream of *CYP6D3* (Fig. 2) and seven in the 3.5 kb upstream sequence of *CYP6D1* (not shown). *CYP6D1* is phenobarbital inducible in wild-type house flies (Scott, 1996; Liu & Scott, 1997). Due to the structural similarities of *CYP6D1* and *CYP6D3* it will be interesting to compare the expression of these two genes, including their responsiveness to phenobarbital treatment. Comparison of the expression of these two genes and of their respective promotor regions should help to identify the factors involved in their regulation.

Experimental procedures

Screening a house fly genomic library

A house fly (Edinburgh strain) genomic library (Tortiglione & Bownes, 1997) was kindly provided by Dr M. Bownes (Edinburgh University). Recombinant phages were infected into XL1-Blue MRA (P2) *E. coli* cells. The 5′ portion of *CYP6D1* (788 bp) was generated by PCR using genomic DNA as template with primers S34 (ctccatagtatcgtggagggt) and A11 (ggtatggtggccttatccga), labelled with [α-^{32}P]dCTP and used as a probe to screen the library. Approximately 9×10^4 plaques were screened under the following conditions: QuickHyb solution (Stratagene, La Jolla, California) at 68 °C for 16 h and a final wash (0.2 × SSC, 0.1%SDS for 30 min at 65 °C). Membranes were exposed to film (Kodak BioMax®) with an intensifying screen (DuPont Cronex®) for 16 h at –80 °C to identify positive clones.

To facilitate sequencing of the large insert found in the positive clone, phage DNA was cut with *Eco*RI or *Sal*I and each restriction fragment subcloned into pUC19 (Fig. 1). Plasmid DNA from putative positive clones was isolated, and the sequences of the inserts were determined by automated sequencing (Cornell Biotechnology Facility). Each sequence was determined until all ambiguities could be resolved.

Sequence analysis

Sequences were aligned and compared using the MegAlign (3.06b) program (Jotun Hein method) from DNASTAR (Madison, WI). Similar sequences were identified using a Blast search of GenBank.

Acknowledgements

We thank J. Ewer and Z. Wen for critical reviews of the manuscript, M. Bownes for the house fly genomic library and D. Nelson for naming the P450. This work was supported by a grant from the National Institutes of Health (GM47835) and Hatch project 414.

References

Adams, M.D., Celniker, S.E., Holt, R.A., *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.

Cherbas, L. and Cherbas, P. (1993) The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochem Mol Biol* **23**: 81–90.

Cohen, M.B. and Feyereisen, R. (1995) A cluster of cytochrome P450 genes of the CYP6 family in the house fly. *DNA Cell Biol* **14**: 73–82.

Dunkov, B.C., Rodriguez-Arnaiz, R., Pittendrigh, B., ffrench-Constant, R.H. and Feyereisen, R. (1996) Cytochrome P450 gene clusters in *Drosophila melanogaster*. *Mol Gen Genet* **251**: 290–297.

Gotoh, O. and Fujii-Kuriyama, Y. (1989) Evolution, structure, and gene regulation of cytochrome P-450. *Frontiers in Biotransformation* (Ruckpaul, K. and Rein, H. eds), pp. 195–243. Taylor & Francis, New York.

He, J. and Fulco, A.J. (1991) A barbiturate-regulated protein binding to a common sequence in the cytochrome P450 genes of rodents and bacteria. *J Biol Chem* **266**: 7864–7869.

Honkakoski, P. and Negishi, M. (1997) Characterization of a phenobarbital-responsive enhancer module in mouse P450 *Cyp2b10* gene. *J Biol Chem* **272**: 14943–14949.

Kasai, S. and Scott, J.G. (2001) Expression and regulation of *CYP6D3* in house fly, *Musca domestica* (L). *Insect Biochem Mol Biol* in press.

Liang, W., He, J. and Fulco, A. (1995) The role of Barbie box sequences as *cis*-acting elements involved in the barbiturate-mediated induction of cytochromes $P450_{BM-1}$ and $P450_{BM-3}$ in *Bacillus megaterium*. *J Biol Chem* **270**: 4438–4450.

Liu, N. and Scott, J.G. (1997) Phenobarbital induction of CYP6D1 is due to a *trans* acting factor on autosome 2 in house flies, *Musca domestica*. *Insect Mol Biol* **6**: 77–81.

Liu, N. and Scott, J.G. (1998) Increased transcription of CYP6D1 causes cytochrome P450-mediated insecticide resistance in house fly. *Insect Biochem Mol Biol* **28**: 531–535.

Liu, N., Tomita, T. and Scott, J.G. (1995) Allele-specific PCR reveals that the cytochrome $P450_{lpr}$ gene is on chromosome 1 in the house fly, *Musca domestica*. *Experientia* **51**: 164–167.

Maitra, S., Dombrowski, S., Waters, L. and Ganguly, R. (1996) Three second chromosome-linked clustered *Cyp6* genes show differential constitutive and barbital-induced expression in DDT-resistant and susceptible strains of *Drosophila melanogaster*. *Gene* **180**: 165–171.

Mansuy, D. (1998) The great diversity of reactions catalyzed by cytochromes P450. *Comp Biochem Physiol* **121C**: 5–14.

Nelson, D.R. (2000) http://drnelsonutmemedu/nelsonhomepagehtml.

Nelson, D.R., Koymans, L., Kamataki, T., Stegeman, J.J., Feyereisen, R., Waxman, D.J., Waterman, M.R., Gotoh, O., Coon, M.J., Estabrook, R.W., Gunsalus, I.C. and Nebert, D.W. (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* **6**: 1–42.

O'Brochta, D. and Atkinson, P. (1996) Transposable elements and gene transformation in non-Drosophilid insects. *Insect Biochem Mol Biol* **26**: 739–753.

Park, Y., Li, H. and Kemper, B. (1996) Phenobarbital induction mediated by a distal CYP2B2 sequence in rat liver transiently transfected *in situ*. *J Biol Chem* **271**: 23725–23728.

Rendic, S. and Di Carlo, F.J. (1997) Human cytochrome P450 enzymes: a status report summarizing their reactions, substrates, inducers, and inhibitors. *Drug Metab Rev* **29**: 413–580.

Scott, J.G. (1996) Cytochrome P450 monooxygenase-mediated resistance to insecticides. *J Pestic Sci* **21**: 241–245.

Scott, J.G. (1999) Molecular basis of insecticide resistance: cytochromes P450. *Insect Biochem Mol Biol* **29**: 757–777.

Scott, J.G., Liu, N., Wen, Z., Smith, F.F., Kasai, S. and Horak, C.E. (1999) House fly cytochrome P450 *CYP6D1*: 5 prime flanking sequences and comparison of alleles. *Gene* **226**: 347–353.

Shaw, G.-C., Sung, C.-C., Liu, C.-H. and Lin, C.-H. (1998) Evidence against the BM1P1 protein as a positive transcription factor for barbiturate-mediated induction of cytochrome $P450_{BM-1}$ in *Bacillus megaterium*. *J Biol Chem* **273**: 7996–8002.

Tomita, T., Liu, N., Smith, F.F. and Sridhar P. and Scott, J.G. (1995) Molecular mechanisms involved in increased expression of a cytochrome P450 responsible for pyrethroid resistance in the housefly, *Musca domestica*. *Insect Mol Biol* **4**: 135–140.

Tomita, T. and Scott, J.G. (1995) cDNA and deduced protein sequence of *CYP6D1*: the putative gene for a cytochrome P450 responsible for pyrethroid resistance in house fly. *Insect Biochem Mol Biol* **25**: 275–283.

Tortiglione, C. and Bownes, M. (1997) Conservation and divergence in the control of yolk protein genes in dipteran insects. *Devel Genes Evol* **207**: 264–281.

Trottier, E., Belzil, A., Stoltz, C. and Anderson, A. (1995) Localization of a phenobarbital-responsive element (PBRE) in the 5 prime-flanking region of the rat *CYP2B2* gene. *Gene* **158**: 263–268.

Venkatesh, B., Ning, Y. and Brenner, S. (1999) Late changes in spliceosomal introns define clades in vertebrate evolution. *Proc Natl Acad Sci USA* **96**: 10267–10271.