Diversification of furanocoumarin-metabolizing cytochrome P450 monoxygenases in two papilionids: Specificity and substrate encounter rate

Weimin Li*, Mary A. Schuler†, and May R. Berenbaum*‡

departments of *Entomology and †Cell and Structural Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801

Diversification of cytochrome P450 monoxygenases (P450s) is thought to result from antagonistic interactions between plants and their herbivorous enemies. However, little direct evidence demonstrates the relationship between selection by plant toxins and adaptive changes in herbivore P450s. Here we show that the furanocoumarin-metabolic activity of CYP6B proteins in two species of swallowtail caterpillars is associated with the probability of encountering host plant furanocoumarins. Catalytic activity was compared in two closely related CYP6B4 and CYP6B17 groups in the polyphagous congeners Papilio glaucus and Papilio canadensis. Generally, P450s from P. glaucus, which feeds occasionally on furanocoumarin-containing host plants, display higher activities against furanocoumarins than those from P. canadensis, which normally does not encounter furanocoumarins. These P450s in turn catalyze a larger range of furanocoumarins at lower efficiency than CYP6B1, a P450 from Papilio polyxenes, which feeds exclusively on furanocoumarin-containing host plants. Reconstruction of the ancestral CYP6B sequences using maximum likelihood predictions and comparisons of the sequence and geometry of their active sites to those of contemporary CYP6B proteins indicate that host plant diversity is directly related to P450 activity and inversely related to substrate specificity. These predictions suggest that, along the lineage leading to Papilio P450s, the ancestral, highly versatile CYP6B protein presumed to exist in a polyphagous species evolved through time into a more efficient and specialized CYP6B1-like protein in Papilio species with continual exposure to furanocoumarins. Further diversification of Papilio CYP6B proteins has likely involved interspersed events of positive selection in oligophagous species and relaxation of functional constraints in polyphagous species.

Cytochrome P450 monoxygenases (P450s) comprise a vast superfamily of heme–thiolate enzymes that catalyze the NADPH-associated reductive cleavage of oxygen to produce a functionalized product and water (1). The genes encoding these enzymes constitute one of the largest known gene superfamilies (http://drnelson.utmem.edu/CytochromeP450.html), with the enormous proliferation reflecting the functional versatility of their encoded proteins. Studies in a wide variety of organisms have demonstrated that P450-catalyzed reactions are important for detoxification of exogenous compounds, such as drugs, toxic pollutants, pesticides, and plant allelochemicals, as well as biosynthesis of endogenous agents, such as steroid hormones, pheromones, and defense compounds (2–5).

It has been suggested that, although the earliest P450s in eukaryotes were important in the metabolism of endogenous substrates, coevolution between plants and herbivorous animals, including insects, expedited the diversification of P450 families (6, 7). The earliest eukaryotic P450s in both plants and animals used reactive oxygen to metabolize endogenous compounds, such as steroids and fatty acids. Subsequent reciprocal adaptive selection between plants and herbivorous animals was associated with the rapid diversification of P450s initiating 400 million years ago, concomitant with the colonization of terrestrial habitats by plants and animals. Plants have used P450s to produce defense compounds (allelochemicals), and herbivorous animals, including insects, have used P450s to metabolize the toxins produced by plants. Multiple duplication and divergence events are thought to have allowed xenobiotic-metabolizing P450s, such as CYP2 and CYP3 in mammals and CYP6 in insects, to diversify and acquire new functions. Insect genome projects have revealed tremendous diversity in putative xenobiotic-metabolizing P450 families, with approximately half of the 90 P450s in the Drosophila melanogaster genome belonging to families CYP6 and CYP4 (8). In the evolution of these large gene families, Hughes and Nei (9) and Ota and Nei (10) have proposed that duplication events may be followed by a winnowing process whereby some duplicate genes “die out” because of accumulation of deleterious mutations. This “evolution by the birth-and-death process” allows the number of functional genes within a family to remain stable. The birth-and-death model may be particularly applicable to the diversification of P450s in herbivorous insects, which, during host shifts, encounter different selective forces associated with the biochemical defense profiles of their host plants.

CYP6B family genes and proteins have been characterized in two groups of lepidopterans: the Helicoverpa/Heliothis complex and Papilio species (11). Within the genus Papilio, proliferation of CYP6B genes has occurred within the context of dietary furanocoumarins, a class of secondary compounds that confer protection against herbivores, because, on activation by UV light, they bind covalently to DNA and protein (12, 13). Furanocoumarins occur in two structural configurations, linear and angular, in over a dozen plant families, and are most widely distributed and diverse in Rutaceae and Apiaceae, the preferred hosts for most Papilio species (14).

Despite the toxicity of furanocoumarins, the oligophagous Papilio polyxenes specifically feeds on furanocoumarin-containing Apiaceae (11). Transcripts of at least one gene, CYP6B1, are expressed at elevated levels in response to supplemental furanocoumarins (15). The CYP6B1 protein encoded by this gene displays very high activity against linear furanocoumarins, and considerably less activity against angular furanocoumarins (15). The CYP6B1 protein encoded by this gene displays very high activity against linear furanocoumarins, and considerably less activity against angular furanocoumarins. Surprisingly, two closely related polyphagous Papilio species (14).

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, “Chemical Communication in a Post-Genomic World,” held January 17–19, 2003, at the Arnold and Mabel Beckman Center of the National Academies of Science and Engineering in Irvine, CA.

Abbreviation: P450s, cytochrome P450 monoxygenases.

†To whom correspondence should be addressed. E-mail: maybe@uiuc.edu.

© 2003 by The National Academy of Sciences of the USA
species, *Papilio glaucus*, which occasionally encounters furanocoumarins, and *Papilio canadensis*, which is unlikely to encounter furanocoumarins because of the absence of furanocoumarins in its available host plants, also have inducible metabolisms of furanocoumarins (16). Sixteen highly conserved genes (92–99% protein identity) belonging to two groups, designated the CYP6B4 and CYP6B17 groups, have been isolated from these two species (16, 17). Although all of these P450 transcripts are inducible by furanocoumarins, the induced level of transcripts achieved in *P. glaucus* is generally higher than in *P. canadensis* (16). The initial member of this group to be defined functionally, CYP6B4 from *P. glaucus*, has the demonstrated capacity to metabolize linear furanocoumarins (ref. 18 and Table 1).

To study the evolution of structure and function of P450s within the context of shifts in host plant utilization and concomitant changes in the chemical milieu experienced by the insect, we compared the CYP6B induction profiles and protein functionalities in these two species in the context of host plant furanocoumarin chemistry. For this study, representative P450s from each of the CYP6B4 and CYP6B17 groups in these species were expressed in baculovirus expression systems in conjunction with the insect NADPH P450 reductase needed for full functional activity. Enzymes were tested for their activity and specificities with respect to host plant furanocoumarins and related compounds to determine whether reduced probability of encounter leads to loss-of-function sequence changes consistent with a birth-and-death process of gene diversification. Also, ancestral P450 genes were reconstructed by maximum likelihood methods to chart a theoretical course for this process.

**Methods**

**Construction and Expression of Recombinant Baculovirus.** Three *P. glaucus* sequences, including the CYP6B4 cDNA and the CYP6B17 and CYP6B21 genomic DNAs, and two *P. canadensis* sequences, including the CYP6B25 and CYP6B26 cDNAs were expressed by using the baculovirus expression system. For expression of the CYP6B17 and CYP6B21 sequences, their 5' UTR and introns were removed by using a PCR-based strategy before subcloning them into the pFASTBac1 baculovirus expression vector (Invitrogen). Briefly, this strategy involved amplifying the 5' end 417 bp of the CYP6B17 or CYP6B21 coding sequences with a forward N1 primer (5'-CGGCTCGAGATCATGTTAA-CGATATTTAT-3) that contains the start codon and a reverse C5 primer (5'-CGGCTTAAGTTTTCCTGACGTG-3), and inserting the resulting PCR product into pBluescript SK vector. The remainder of the coding sequence was generated without its intron by amplifying the CYP6B17 and CYP6B21 genomic sequences with the INTFOR1 primer (5'-CTGGCCAGAGAAAATGCCTAGAATCGGTGGATTTGCACC-3') that spans the sequences flanking the intron and a reverse C2 primer (5'-CGGAAGCCTCAATATTATGCTGTTGGACA-3'), and ligating this PCR product with the 5' coding sequence and the pBluescript SK vector. The resulting cDNA sequences, which contain an XhoI site upstream from the translation start, a HindIII site downstream from the translation stop and an EagI site at the junction between the two fragments, were checked for amplification errors by sequencing. All of these CYP6B cDNAs were subcloned into suitable restriction sites of the pFASTBac1 baculovirus expression vector for construction of recombinant viruses and expressed in the Bac-to-Bac expression system (Invitrogen). All procedures of construction and expression of recombinant virus in insect Sf9 cells were performed as described by the manufacturer. *P. polyxenes* CYP6B1 cDNA (L. Pan, Z. Wen, J. Baudry, M.R.B., and M.A.S., unpublished data) was expressed as a positive control to monitor P450 expression quality. Sf9 cell cultures were grown to a density of 0.8–1 × 10^8 cells per ml in SF-900 serum-free medium supplemented with 8–10% FBS, 50 μg/ml streptomycin sulfate, and 50 units/ml penicillin and cotransfected with recombinant P450 virus at a multiplicity of infection (MOI) of 2 and recombinant house fly NADPH P450 virus at an MOI of 20 (ref. 19 and L. Pan, Z. Wen, J. Baudry, M.R.B., and M.A.S., unpublished data). This MOI ratio of 1:10 was intended to supplement the limited electron transfer capacities of Sf9 cells.

<table>
<thead>
<tr>
<th>Specific activity (nmol/min/mmol P450), * means ± SD</th>
<th>P450</th>
<th>CO-diff</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CYP6B4 (pg)</strong></td>
<td>1.906 ± 0.180&lt;sup&gt;a&lt;/sup&gt;</td>
<td>450</td>
</tr>
<tr>
<td><strong>CYP6B17 (pg)</strong></td>
<td>0.060 ± 0.023&lt;sup&gt;a&lt;/sup&gt;</td>
<td>450</td>
</tr>
<tr>
<td><strong>CYP6B21 (pg)</strong></td>
<td>0.518 ± 0.130&lt;sup&gt;a&lt;/sup&gt;</td>
<td>450</td>
</tr>
<tr>
<td><strong>CYP6B25 (pg)</strong></td>
<td>0.372 ± 0.210&lt;sup&gt;a&lt;/sup&gt;</td>
<td>420</td>
</tr>
<tr>
<td><strong>CYP6B26 (pg)</strong></td>
<td>0.640&lt;sup&gt;a&lt;/sup&gt;</td>
<td>420</td>
</tr>
</tbody>
</table>

<sup>a</sup>Specific activities for each furanocoumarin were compared, and significant differences are represented with superscript letters (*P < 0.05* (ANOVA test).

<sup>†</sup>CYP6B1 activities were measured at different ratio of MOIs at 4:1 of P450 vs. P450-NADPH reductase recombinant viruses (28).

<sup>‡</sup>Not tested.

*Specific activities for each furanocoumarin were compared, and significant differences are represented with superscript letters (*P < 0.05* (ANOVA test).
Hem in was added to 2 μg/ml final concentration 24 h after infection. Insect cells were harvested from ~6 plates for each P450 and lysed as described by Chen et al. (20), and protein aliquots were frozen in liquid nitrogen before analysis. Carbon monoxide (CO) difference spectra were measured as in Omura and Sato (21) by using an extinction coefficient for the reduced CO complex of 91 mM⁻¹cm⁻¹.

**P450 Metabolic Activity Assays.** *In vitro* metabolism assays of furanocoumarins were conducted as described in Li et al. (16) except that the substrate concentration in each reaction was lowered to 1 μg/ml (corresponding to final concentrations of 4.6 mM xanthotoxin, 4.6 μM bergapten, 5.4 μM angelicin, or 5.4 μM psoralen) and a different furanocoumarin was added as an internal control after reactions were terminated. Rates of O-deethylation of 7-ethoxycoumarin were determined by measuring the fluorescence of the 7-hydroxycoumarin product (22). All metabolic assays were performed with four replicates and repeated at least twice by using proteins prepared from independent infections. Metabolic activities were compared by analysis of variance as described (16).

**Sequence Alignments, Phylogenetic Analyses, and Reconstruction of Ancestral Sequences.** Amino acid sequences of 22 CYP6B proteins, 12 other CYP6 proteins, and the divergent CYP321A1 protein were aligned to rabbit CYP450s (PDB ID 1DT6) and Bacillus megaterium CYP1012 (PDB ID 2HPD), for which crystal structures are available (23, 24). Multiple sequence alignments were performed by using CLUSTALW or multialignment modules within the MOE program (Chemical Computing Group, Montreal) with the Gonnet weight model and structural alignment enabled. Alignments generated by both methods were compared and modified according to secondary structures of the CYP2C5 and CYP102 proteins. The finalized alignment was analyzed by the MEGA program to construct the phylogeny of the CYP6 family. A maximum parsimony phylogeny of CYP6 sequences was generated by using max–mini branch-and-bound method, and the inferred phylogeny was tested by 500 bootstrap tests. A maximum likelihood method was used to construct ancestral CYP6B sequences from the above multiple alignment of protein sequences on the most parsimonious phylogeny by using PAML (25) with the implementation of Jones amino acid transformation model and iteratively estimated γ shape parameter. To subsequently determine the relationship of ancestral sequences to their descendant P450s, relative amino acid distances between the ancestral sequences and their descendant branches were computed by using the MEGA program based on either complete protein sequences or active site amino acid residues. Active site residues were defined, based on our recent furanocoumarin docking study, as amino acids that are <4.5 Å from the oxo-heme moiety or furanocoumarin substrates docked into 3D models of CYP6B proteins (26).

**Results and Discussion**

**Catalytic Activity of CYP6B Enzymes.** For comparative purposes, five *Papilio* CYP6B enzymes, including CYP6B4 (*P. glaucus*), CYP6B17 (*P. glaucus*), CYP6B21 (*P. glaucus*), CYP6B25 (*P. canadensis*), and CYP6B26 (*P. canadensis*), were coexpressed with house fly P450 reductase in S9 cells as described in Methods. The quality and quantity of the expressed P450 proteins were determined by reduced CO difference analysis (21) using S9 cell lysates prepared from cells cotransfected with recombinant P450 and P450 reductase viruses. In these assays, all of the *P. glaucus* enzymes generated CO-difference maxima at 450 nm (P450 form) with no significant absorbance at 420 nm, indicating that all of the P450 proteins fold and incorporate heme correctly into their apoproteins. Of the two *P. canadensis* proteins, CYP6B25, an in-group homologue to CYP6B4 (17), generated a major CO-difference peak at 450 nm and a minor peak at 420 nm and CYP6B26, an in-group homologue to CYP6B17 and seven amino acids shorter at its C terminus (16), generated a CO-difference peak only at 420 nm (P420 form), indicating that this enzyme is incorrectly folded. Molecular modeling and site-directed mutagenesis presented in Chen et al. (20) have indicated that aromatic residues F116, H117, and F484 positions are required for correct folding of the *P. polyxenes* CYP6B1 enzyme, and that two of these modulate the range of furanocoumarins metabolized. This seems not to be the case for *P. glaucus*/*P. canadensis* enzymes that contain a nonaromatic residue (L484) replacing F484 in the *P. polyxenes* enzyme and nonetheless generate the functional P450. The low stability of the CYP6B25 protein, which also contains L484, is apparently caused by substitutions at other positions because F116, H117, and L484 are conserved between CYP6B4 and CYP6B25. The inability of the CYP6B26 protein to fold into stable P450 indicates the importance of C-terminal residues in determining overall protein folding or configuration of the heme-binding domain.

To determine the influence of furanocoumarin exposure on P450 catalytic activity and selectivity, the metabolic activities of the expressed CYP6B proteins against five linear and angular furanocoumarins were assayed and compared with the metabolic activity of the *P. polyxenes* CYP6B1 protein coexpressed with P450 reductase in S9 cells (Table 1) (27). The substrates tested include methoxylated (bergapten, xanthotoxin), trimethylated (trioxossalen), and unsubstituted (psoralen) linear furanocoumarins and an unsubstituted (angelicin) angular furanocoumarin. These metabolic data indicate that all of the *Papilio* CYP6B enzymes are able to turn over furanocoumarin substrates to some extent, except for the denatured CYP6B26 protein, which did not metabolize any of the tested substrates (Table 1).

The relative abilities of the CYP6B proteins to metabolize linear furanocoumarins were closely associated with the probability of encountering furanocoumarins in host plants. CYP6B1, characterized from the specialist *P. polxenes*, which encounters high levels of furanocoumarins in all of its host plants, turns over linear furanocoumarins at the highest rate (6980 pmol/min/nmol P450 for xanthotoxin) among all of CYP6B enzymes tested. CYP6B4 and CYP6B17, characterized from the generalist *P. glaucus*, which occasionally encounters furanocoumarins, turn over linear furanocoumarin substrates at rates lower than does CYP6B1 but significantly higher than do CYP6B25 and CYP6B26, characterized from the generalist *P. canadensis*, which never encounters furanocoumarins naturally. Further comparisons of P450 substrate specificities have indicated that CYP6B1 is more selective than the CYP6B proteins in polyphagous *P. glaucus*/*P. canadensis*. CYP6B1 exhibits very high activity toward the methoxylated linear furanocoumarins xanthotoxin and bergapten, lower activity toward unsubstituted and other linear furanocoumarins, and much lower activity toward the angular furanocoumarin, angelicin (ratio of activity for xanthotoxin/psoralen/angelicin is 1.0/0.4/0.1) (Table 1) (18, 27). This finding suggests that linear and angular furanocoumarins in *P. polyxenes* are metabolized by distinct P450s that possess relatively high substrate specificity. The in-group homologues CYP6B4 and CYP6B25 display more uniform activities toward a number of furanocoumarin substrates. Whereas CYP6B4 and CYP6B25 turn over methoxylated furanocoumarins at higher rates than other furanocoumarins, they metabolize psoralen and angelicin at rates higher than those of CYP6B1 (for CYP6B4, xanthotoxin/psoralen/angelicin is 1.0/0.7/0.6; for CYP6B25, it is 1.0/0.5/0.3). Importantly, these ratios demonstrate that these P450s exhibit the same preference for unsubstituted linear and angular furanocoumarins in that they turn over psoralen and angelicin at approximately the same rate. In comparison, CYP6B17, the paralog of CYP6B4 in *P. glaucus*, turns over linear
furanocoumarins at rates lower than CYP6B4 and exhibits no activity against angular furanocoumarin. Together, these data indicate that, among the enzymes tested, furanocoumarin metabolism in *P. glaucus* is mediated primarily by CYP6B4.

Additional assays performed to determine the range of substrates other than furanocoumarins metabolized by CYP6B enzymes in *P. glaucus* and *P. canadensis* demonstrate that at least one substrate, 7-ethoxycoumarin (EC), representing alkoxylated coumarins that are widely present in the host plants of swallowtails (12), is metabolized by all three *P. glaucus* P450s at levels lower than those defined for furanocoumarin substrates. Interestingly, despite its higher activities against furanocoumarins, the 7-ethoxycoumarin O-deethylase (ECOD) activity of CYP6B4 is the lowest of all of the *P. glaucus* P450s tested, with an activity that is 5.1- and 1.4-fold lower than those of CYP6B21 and CYP6B17, respectively (Table 2). In contrast, CYP6B25 and CYP6B26 from *P. canadensis* have no detectable ECOD activity. Recent studies have indicated that the more specialized CYP6B1 from *P. polyxenes* also has no ECOD activity (Z. Wens, personal communication).

The catalytic efficiencies and substrate ranges of these CYP6B proteins map closely onto the range of host plants encountered by the insect species producing these P450s. In the specialist *P. polyxenes*, which has the highest frequency of encountering furanocoumarins, the CYP6B enzyme examined is more specialized and capable of metabolizing furanocoumarins at higher rates than the CYP6B proteins expressed in other *Papilio* species. In the generalist *P. glaucus*/*P. canadensis*, the CYP6B proteins have broader substrate ranges than the CYP6B1 proteins but metabolize this broader range of substrates at lower rates. The *P. canadensis* enzymes have even lower metabolic activities for all of the substrates tested than those of their *P. glaucus* orthologues. These relatively low activities may relate to the unpredictability of the chemical milieu encountered by these insects. Orthologues of these low-activity enzymes probably result from the lack of selection pressure to maintain furanocoumarin metabolism. This is particularly evident in *P. canadensis*, which has no natural exposure to furanocoumarins, and the lower activities of its CYP6B enzymes probably result from the accumulation of many deleterious mutations that have no purifying selection driving their elimination. Within *P. glaucus*, differences in substrate specificities between the CYP6B4 and CYP6B17 paralogs are likely to be the consequence of positive selection for duplicated genes. CYP6B7, CYP6B8, and CYP6B27, which are derived from *Helicoverpa zea*/*Helicoverpa armigera*, two closely related agricultural pests with extremely wide host ranges, are inducible by a wide range of chemicals, including insecticides, plant hormones, and plant allelochemicals, such as xanthotoxin (28–30). Among these, the CYP6B8 protein has a demonstrated ability to metabolize a range of plant allelochemicals and insecticides (31), as befits its highly diverse chemical environment. This finding further confirms the association between P450 versatility and host plant range.

**Phylogeny of CYP6B Genes and Reconstruction of Ancestral CYP6B Sequences.** Given the close relationship between CYP6B catalytic efficiency and substrate specificity and the complexity of the chemical environment experienced by the insect, it is of great interest to explore how this association was established and how it evolved into the present-day CYP6B enzymes. For this analysis, a phylogeny of CYP6B genes was constructed based on the multiple sequence alignments described in Methods and resulting in the most parsimonious phylogeny shown in Fig. 1. In this, CYP6B sequences form a distinct clade with the *H. zea* CYP321A1 sequence that is phylogenetically separate from other CYP6 clades. The close relationship of the CYP321A1 protein to CYP6B proteins is confirmed by similarities in their function; CYP321A1 is the only non-CYP6B insect P450 known to be capable of metabolizing furanocoumarins (32). Within the CYP6B branch, *Papilio* and *Helicoverpa* CYP6B genes form distinct clades, suggesting no major duplication event before separation of these species. However, serial duplications occurred in both *P. glaucus*/*P. canadensis* and *H. zea/H. armigera* branches (Fig. 1), yielding at least 17 closely related CYP6B genes in the *P. glaucus*/*P. canadensis* genomes that do not appear to be pseudogenes (16, 17). After the first duplication, two paralogous groups, designated the CYP6B4-group and the CYP6B17-group (Fig. 1), formed. Although early studies demonstrated distinct induction profiles for each of these groups (16), the studies presented here indicate that members of these groups also display distinct substrate preferences, indicating functional divergence of these P450s after their initial duplication. The “redundancy” of CYP6B genes in these polyphagous species is likely a mechanism for adapting to complex chemical environments, maintaining one functional gene while another evolves with different metabolic capabilities. Duplicated copies of P450s may well provide candidates for subsequent functional divergence to adopt new functions as host plant patterns change.

To characterize the process of functional divergence among CYP6B8s, three ancestral sequences, the ancestor of the whole CYP6B subfamily (Anc1), the ancestor of the *Papilio* branch (Anc2), and the ancestor of the *P. glaucus*/*P. canadensis* branch

**Table 2. Ethoxycoumarin 7-O-deethylation activities of *P. glaucus*/*P. canadensis* CYP6B enzymes**

<table>
<thead>
<tr>
<th>P450s</th>
<th>CYP6B4</th>
<th>CYP6B7</th>
<th>CYP6B21</th>
<th>CYP6B25</th>
<th>CYP6B26</th>
</tr>
</thead>
<tbody>
<tr>
<td>O-deethylation*&lt;sub&gt;2&lt;/sub&gt;</td>
<td>102 ± 5</td>
<td>143 ± 4</td>
<td>518 ± 4</td>
<td>ND&lt;sup&gt;1&lt;/sup&gt;</td>
<td>ND&lt;sup&gt;1&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>1</sup>Values listed are mean ± SD of at least two independent determinations each with four replicates.
<sup>2</sup>No detectable O-deethylation activity found for CYP6B25 and CYP6B26.
were reconstructed and compared with current CYP6B proteins (Figs. 2 and 3). The relative distances (ratios of distance between ancestral sequences and respective descendant branches) of all three ancestral sequences are not greater from unity than those computed based on the complete protein sequences, suggesting that these genes evolve at approximately the same rate from their common ancestors (Fig. 2). Similar computations were based on functionally more important active site residues, which directly interact with substrates; the relative distances of Anc1 to *Papilio* or *Helicoverpa* sequences, and Anc3 to CYP6B4 group or CYP6B17 group of sequences, respectively, are statistically different from unity (*P* < 0.05) (Fig. 2). Anc1 displayed a significantly closer relationship to *Helicoverpa* P450s. The amino acid substitution rate leading to the *Papilio* lineage is 1.5-fold higher than the rate leading to the *Helicoverpa* lineage. It is conceivable that Anc1 may have similar functionality to that of *Helicoverpa* P450s and very likely catalyzed a broader range of substrates with lower metabolic efficiency against furanocoumarins compared with most of the *Papilio* P450s. Anc2 is proposed to have existed in a specialist species with constant exposure to furanocoumarins and to have had relatively higher catalytic activities against furanocoumarins compared with most *Papilio* P450s (17). This is likely to be the case, because the amino acid substitution rate of the CYP6B1/CYP6B3 lineage appears to be 1.3-fold that of the CYP6B4/CYP6B17 lineage, although this difference is not statistically significant because of higher sequence variation between CYP6B1/CYP6B3 pair. Anc3 is more closely related to the CYP6B4 group of P450s than the CYP6B17 group, as indicated by comparing the postduplication amino acid substitution rates leading to these lineages. Among all CYP6B lineages, the CYP6B17 and CYP6B4 branches displayed the most dramatic difference (2.1-fold) in the amino acid replacement rate from the immediate ancestor (Fig. 2). After the duplication that led to diversification of the CYP6B4 and CYP6B17 groups, the CYP6B4 group evolved relatively more slowly, and thus very likely resembles Anc3 in both sequence and function; the CYP6B17 group, however, as the duplicate presumably relieved of purifying selection, evolved faster and adapted to novel functionality.

Further insight into the evolutionary lineage of these P450 proteins has been gathered by comparing the predicted structures of these proteins with those predicted for the more efficient CYP6B1 protein. In this latter protein, an aromatic network that involves residues Phe-116, His-117, Phe-484, and Phe-371 is critical for substrate binding affinity to the CYP6B1 active site (20, 33). Three-dimensional models of the present day CYP6B4 and the ancestral Anc1 and Anc2 proteins that we have con-
structed (26) were compared to evaluate the existence and importance of this aromatic network in defining the stabilities and substrate specificities of these proteins. The catalytic pocket of Anc1 displayed substantial differences from that of the CYP6B1 protein (Fig. 4). Among the four networked residues in the CYP6B1 model, Phe-116, His-117, and Phe-371 are conserved in other Papilio CYP6B proteins. In the Anc1 model, the aromatic Phe-116 and His-117 residues are not oriented perpendicular to one another in positions typical for aromatic–aromatic interactions, and the side chain of Phe-371 projects away from the catalytic pocket in a configuration that may not allow it to directly interact with other residues in the catalytic pocket (Fig. 4a). In Anc1, the catalytic pocket is probably larger in volume and with more flexibility than that defined within the CYP6B1 protein because a nonaromatic Ile-484 replaces the larger Phe-484 found in the CYP6B1 protein. In this regard, Anc1 is very similar to H. zea CYP6B8, which is also predicted to contain a large and flexible catalytic pocket able to accommodate a wide range of substrates (31). The absence and/or weakening of the π-π stacking interactions between Phe-484 and furanocoumarin substrates that facilitate binding in the CYP6B1 catalytic pocket predicts that Anc1 probably catalyzes metabolism of furanocoumarins significantly less efficiently than the CYP6B1 protein. 

In contrast, Anc2 is very likely a CYP6B1-like protein that maintains the ability to metabolize furanocoumarins. Residues involved in formation of the aromatic network, Phe-116, His-117, Phe-371, and Phe-484, are conserved and occupy approximately identical spatial locations to those in the CYP6B1 protein (Fig. 4b), indicating that the Anc2 enzyme has an aromatic network that is similar to that in CYP6B1 protein. This might be presumed to make the catalytic pocket relatively narrow and rigid, indicating similar structural constraints to that in the CYP6B1 protein, presumably resulting from strong selection by constant exposure to furanocoumarins. This finding provides important supporting evidence for our hypothesis that the common ancestor of these Papilio species encountered furanocoumarins with high frequency in its host plants. Two residues (Phe-371 and Phe-484) contributing to this aromatic network are not preserved in the P. glaucus/P. canadensis branches, as suggested in our recent studies (26). The lack of important aromatic interactions in the CYP6B4 active site probably causes lower binding affinity to furanocoumarins, which in turn leads to lower turnover rates for linear furanocoumarins than those in P. polyxenes. The less efficient but more versatile CYP6B4 protein is well suited to the toxicological needs of the polyphagous P. glaucus.

In summary, our studies demonstrated for the first time how shifts in herbivore utilization of host plants are associated with the evolution of the sequence, structure and function of their P450s. It is highly likely that insects possessing CYP6B1 enzymes originated from a broadly polyphagous species with limited capacity for specialized metabolism. As the specialist papilionids diverged from this generalist ancestor in association with their furanocoumarin-containing host plants, the ability to metabolize a broad range of substrates declined as the ability to metabolize the principal host plant defense compounds, the furanocoumarins, reached a level of refinement rarely equaled by other herbivores. This tradeoff between P450 breadth and specificity may explain in part the ubiquity of oligophagy that characterizes lepidopterans today (34).

We thank Dr. Jerome Baudry for extensive training in modeling programs and Dr. Liping Pan for providing recombinant baculovirus containing NADPH P450 reductase. This work has been supported by National Science Foundation Grant IBN 02 12242 (to M.R.B.) and National Institutes of Health Grant R01 GM50007 (to M.A.S.).