FUNCTIONAL AND DNA SEQUENCE DIVERGENCE OF THE CYP71 GENE FAMILY IN HIGHER PLANTS

Rolf E. Christoffersen, Frank W. Percival[1] and Kristin R. Bozak[2]

Division of Molecular, Cellular and Developmental Biology
Department of Biological Sciences, University of California
Santa Barbara, CA 93106, USA

SUMMARY

During ripening of avocado (*Persea americana*), the CYP71A1 mRNA and protein accumulate to relatively high levels. Although the *CYP71A1* gene was the first plant P450 to be cloned and sequenced, the functional role of this P450 remains obscure. Substrate studies have shown that CYP71A1 will metabolize various monoterpenes (nerol and geraniol), although these have not been detected in ripening fruit. Using DNA from a conserved domain of the *CYP71A1* gene, we have explored the scope of the CYP71 (or related) gene family in avocado using low stringency DNA hybridization. This analysis suggests that there are ~ 10-12 genes in the CYP71 family. An alternative approach using PCR gave essentially identical results.

KEY WORDS

P450 gene family, DNA hybridization, polymerase chain reaction, monoterpene hydroxylase, epoxidase

---

[1] Permanent address: Department of Biology, Westmont College, Montecito, CA 93108, USA

[2] Present address: Department of Biological Sciences, California State Polytechnic University, Pomona, CA 91768-4032, USA

INTRODUCTION

The characterization of plant P450 proteins and the corresponding genes has lagged behind those from animal or microbial sources /1-3/. In a few isolated cases, progress has been made because a highly specialized tissue was found that contained a relatively large amount of a single P450 protein. One example is the mesocarp tissue from ripe avocado fruit, which has been shown by a number of investigators to have a relatively high level of spectrally detectable P450 enzyme /4-7/. This has allowed the purification of the homogenous protein and the determination of 40 residues from the *N*-terminal sequence of the purified polypeptide /6/. The avocado fruit P450 was initially characterized as a *p*-chloro-*N*-methylaniline N-demethylase.

A ripening-related cDNA isolated from ripe avocado fruit contained an open reading frame which defined a new P450 protein family designated CYP71 /8,9/. Comparison of the N-terminal sequence of purified avocado P450 protein described above with the N-terminus of the CYP71A1 sequence showed them to be identical. Thus, the CYP71A1 protein was the first plant P450 to be characterized at the molecular and biochemical level. Subsequently, a number of related CYP71 family members have been discovered in a variety of plant species, including eggplant, *Thlaspi arvense*, *Arabidopsis*, maize, and peppermint /10-13/. The precise functional role of any CYP71 family member has yet to be established.

The CYP71A1 protein has been functionally expressed in yeast cells /8/ and the substrate specificity of the expressed enzyme characterized /8,14/. Spectroscopic analysis of microsomes from ripe avocado strongly suggested that nerol or a related monoterpene might be a substrate for the CYP71A1 protein /7/. It was subsequently confirmed by GC/MS analysis that both nerol and geraniol are metabolized by the CYP71A1 enzyme /14/. The products of this reaction were shown to be a mixture of 2,3-epoxy- or 6,7-epoxynerol (or epoxygeraniol). No diepoxides were detected suggesting the monoepoxide products are poor substrates for the CYP71A1 protein. It is likely that nerol epoxidation does not play a role in fruit ripening. Neither nerol nor geraniol (or the epoxidated derivatives) have been detected in ripening avocado fruit /14/ (K. Courtney and R.E. Christoffersen, unpublished work). In addition, the V_{max} for nerol was reported to be ~ 3.0 min^{-1}, a value slower than that observed with *p*-chloro-*N*-methylaniline (~ 8.0 min^{-1}). These relatively low rates, combined with the spectral studies of

Hallahan *et al.* /7/ and the absence of these compounds in the fruit, strongly suggest that nerol and geraniol are not the natural substrates of the CYP71A1 P450 in avocado fruit /14/.

A related P450 activity, geraniol 10-hydroxylase, has been well described /15-17/ and is the first committed step in the biosynthesis of iridoid monoterpenes. Using a DNA probe derived from CYP71A1, a strongly cross-hybridizing DNA fragment was observed in the genome of *Nepeta racemosa* (catmint) /14/. Tissues showing high levels of CYP71A1-related mRNA and antigen had high enzymatic levels of geraniol 10-hydroxylase and detectable nerol epoxidase activity. Therefore the proposal was made that the CYP71A1 orthologous gene in catmint was responsible for the geraniol 10-hydroxylase activity in this tissue. This evidence further strengthens the proposal that the CYP71 family is likely to be involved in the metabolism of terpenoids.

The number of plant metabolic pathways that utilize P450 mediated reactions is likely to be rather large. This is especially true for P450 enzymes involved in terpenoid metabolic pathways. Many of these compounds have important biological or pharmacological properties and have a variety of hydroxylated or epoxidated derivatives /3/. It is unknown whether a different P450 enzyme is involved in each specific reaction in these pathways, although most well characterized plant P450 enzymes show high specificity for their substrates /18/. In the mint family, the pattern of monoterpene hydroxylation is highly specific and varies with different species, suggesting a distinct but potentially related P450 is involved in each reaction /19-21/.

In order to assess the extent of the CYP71 (or closely related) gene family in a single plant, we have used low stringency DNA hybridization to detect genomic DNA fragments carrying CYP71A1-related sequences. Alternatively, we have used a PCR approach with degenerate primers based on conserved motifs of CYP71A1 that confirmed our low stringency hybridization experiments. Both of these different experiments suggest that the CYP71 family in avocado is likely to consist of at least 10-12 related genes.

MATERIALS AND METHODS

DNA hybridization

Genomic DNA was isolated from ripe avocado fruit using the method of /22/ as previously described in detail /23/. The fragment of

DNA designated probe 71ex2 was a 384 bp Bgl II endonuclease restriction fragment recovered from a digest of pSKcyp71A1, a full length cDNA previously described /8/. The source of the other probe used is described in the section below on PCR amplification. The probes for hybridization were radiolabeled by a random priming /24/ using ^{32}P-dCTP (>3000 Ci/mmole). After labeling, probes were purified by gel filtration through Sephadex G-50.

DNA hybridization was carried out by the method of Church and Gilbert /25/. Briefly, the prehybridization and hybridization were done in 0.25 M sodium phosphate, pH 7.2; 7% SDS; 1 mM EDTA; and 1% bovine serum albumin at 55°C. After overnight hybridization, membranes were washed at low stringency twice for 15 min in 2X SSC, 0.2% SDS at 30°C followed by two more washes for 45 min in 0.2X SSC, 0.2% SDS at 50°C. The membranes were then exposed to film at -70°C with an intensify screen. The membranes were then subjected to a high stringency wash that consisted of 1 h in 0.1X SSC, 0.2% SDS at 50°C followed by 2.5 h in 0.1X SSC/0.2% SDS at 60°C. The membranes were then reexposed to film as described above.

PCR amplification of CYP71-related protein domains

Total genomic DNA from avocado was used as a template in a PCR to generate the 430 bp fragment that was used as a probe in the DNA hybridization experiments. The two degenerate oligonucleotides used in the PCR reaction were designated the M primer (5'-ATG TTY WSI GGI GGI ACI GAY ACI AC -3'; protein motif - MFSGGTDTT) and the C primer (5'-RCA ICC ICK ICC IGC ICC RAA-3'; protein motif - FGAGRRGC). The PCR reaction was carried out in 10 mM Tris-HCl, pH 9.0, 50 mM KCl, 0.1 % (w/v) Triton X-100, 1.5 mM $MgCl_2$, 0.1 mg/ml gelatin, 2 µM of each primer, 1 µg of avocado genomic DNA, 0.2 mM of the four deoxyribonucleotides, and 2.5 units of *Taq* polymerase. The thermal cycle was as follows: ramp up to 94°C in 30 sec, hold for 30 sec, ramp to 40°C in 1 min, hold for 30 sec, ramp to 72°C in 2 min, and hold for 1 min, which was repeated 35 times. Following gel electrophoresis, the 430 bp band was extracted from the gel using Prep-a-gene as recommended by the manufacturer (BioRad, Inc., Richmond, CA) and radiolabeled as described above.

DNA sequence analysis

Plant P450 protein sequences were obtained from the GenBank database. Sequence analysis was done using the Wisconsin Sequence Alignment Package /26/ from the Genetics Computer Group, Madison, Wisconsin, U.S.A. The plant P450 protein sequences were aligned using the PILEUP program. The sequence divergence among the aligned P450 sequences was calculated using the DISTANCES program and the tree calculated using the GROWTREE program with the UMGPA algorithm.

RESULTS

Structure of the CYP71A1 gene

The CYP71A1 gene was isolated from an avocado genomic library and the exon regions identified by restriction mapping and DNA sequence analysis (Fig. 1). Complete details describing the characterization of the CYP71A1 gene will be published elsewhere (Bozak, Lin, Percival, and Christoffersen, manuscript in preparation). The

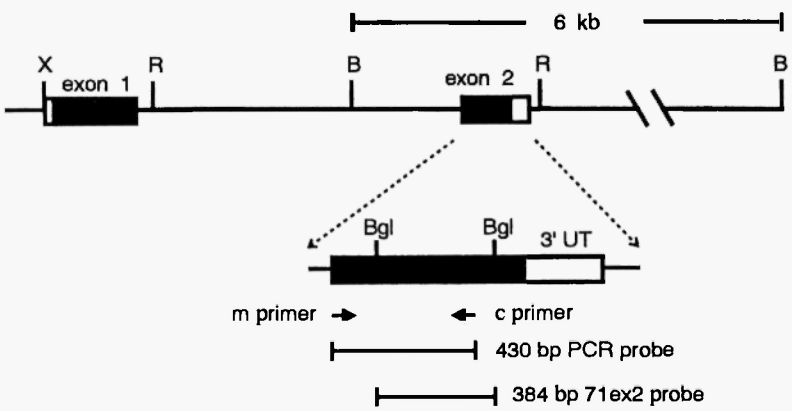

Fig. 1: Restriction map of CYP71A1 gene isolated from a lambda bacteriophage genomic DNA library of *Persea americana* cv Hass. The restriction endonuclease sites are *Bam* HI - B, *Eco* RI - R, *Bgl* II - Bgl. The horizontal arrows indicate the relative position of primers used in the PCR experiments. The bars below indicate the positions of the two DNA probes used in the hybridization experiment described in Fig. 3.

important features of the gene structure pertinent to the experiments described here are briefly presented. The CYP71A1 gene has two exons separated by a single intron of approximately 5 kb. The intron/exon junction occurs between Leu-299 and Asp-300 with exon 2 coding for the conserved heme-binding domain with no interrupting introns. Thus, in contrast to most mammalian P450 genes, the structure of the CYP71A1 gene is relatively simple. Furthermore, the heme-binding domain consists of a compact region approximately 900 bp in length. This size is small enough that the entire region is likely to be contained in a single DNA fragment when the avocado genome is subjected to Southern blot analysis.

PCR amplification of the heme-binding domain from CYP71 related genes

If the intron/exon structure is conserved in CYP71-related genes, it would be theoretically possible to use PCR to amplify either exon 1 or exon 2 using total genomic DNA as a template. Two degenerate primers were synthesized to target conserved protein motifs contained within exon 2 of the CYP71A1 gene. The location of these conserved motifs is indicated in Fig. 1. The 5' sense-strand primer was targeted to the protein motif MFSGGTDTT. This region is at the extreme 5' end of exon 2 and is thought to be part of the I helix based on homology to the sequences of P450cam and BM-3, P450 proteins of known three dimensional structure /27,28/. This region forms part of the oxygen binding pocket on the distal side of the heme. The 3' antisense primer was targeted to a highly conserved P450 motif, FGAGRRGC, that forms the proximal Cys ligand to the heme /29/. In the *CYP71A1* exon 2 sequence, the distance from the 5' end of the M primer to the 3' of the C primer is exactly 428 bp. The PCR products were analyzed by agarose gel electrophoresis and stained with ethidium bromide (Fig. 2). When both primers and template genomic DNA are present a prominent band is observed at 430 bp (lane 2). If either primer is left out or if template DNA is omitted the 430 bp fragment is absent (minus M primer - lane 3; minus F primer - lane 4; no template DNA - not shown). Thus these primers appear to amplify from total genomic DNA a portion of exon 2 from *CYP71A1*. Whether the 430 bp product is homogenous or represents a population of related DNA fragments is explored below.

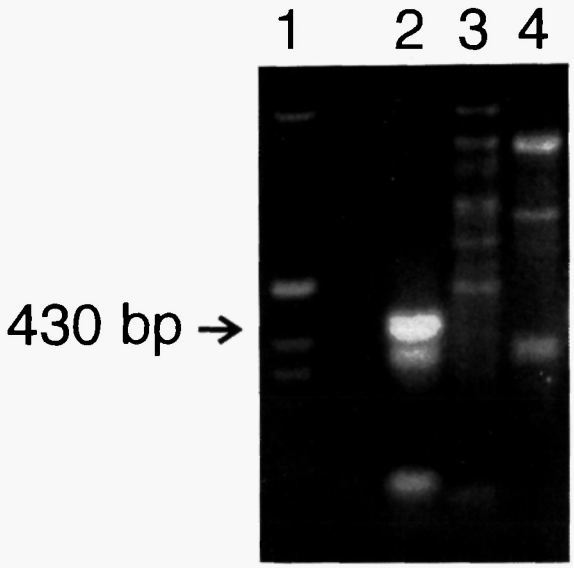

Fig. 2: Gel electophoretic analysis of the products of a PCR amplification using degenerate oligonucleotide primers and total avocado genomic DNA as a template. Molecular size markers are in shown in lane 1. Lane 2 is the complete reaction with both primers and template as described in Materials and Methods. In lanes 3 and 4, either the M primer or the C primer was left out of the PCR reaction. The arrow indicates the 430 bp fragment from lane 2 that was recovered from the gel and used as the PCR probe in hybridization experiments.

Blot hybridization of avocado genomic DNA

Avocado genomic DNA was restricted with Bam HI, separated by agarose gel electrophoresis, and examined for CYP71-related fragments using two different hybridization probes and under different conditions of stringency. As expected, the 71ex2 probe detected the 6 kb Bam HI fragment (Fig. 3 - lanes 1 and 2) predicted from the restriction map of the cloned CYP71A1 gene (Fig. 1). Under conditions of low stringency hybridization, which allows related sequences to cross-hybridize to the probe, at least 10 additional restriction fragments were detected (Fig. 3 - lane 1). Under high stringency washes, these related fragments no longer hybridized to the 71ex2 probe (Fig. 3 - lane 2). When the 430 bp PCR product was used as the probe under high stringency hybridization (Fig. 3 - lane 3), a

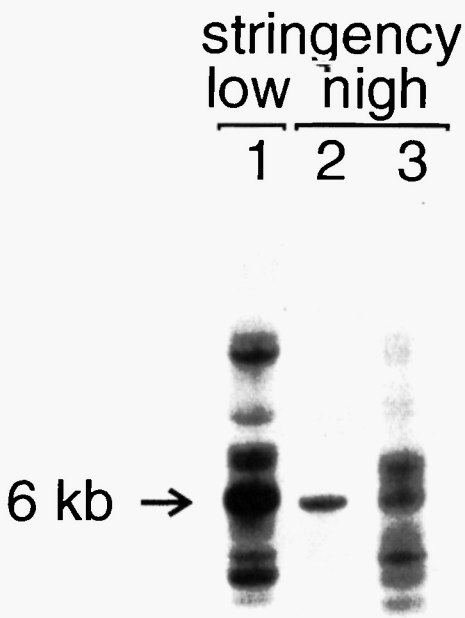

Fig. 3: DNA gel blot analysis of avocado genomic DNA. Each gel lane had 10 μg of *Bam* HI digested avocado DNA. Lanes 1 and 2 were hybridized to the CYP71ex2 probe and lane 3 was hybridized to the 430 bp PCR probe. After hybridization, the membranes were washed under either low or high stringency conditions as indicated above the lanes.

pattern similar to the low stringency hybridization with the 71ex2 probe was observed. This shows that the 430 bp PCR product is likely to be a heterologous mixture of DNA fragments representing a family of genes. These same fragments appear to have high sequence similarity to the regions of the avocado genome that are detected by hybridization to the heme-binding domains of the *CYP71A1* gene.

DISCUSSION

The expression of the *CYP71A1* gene is closely coordinated with the ripening process in avocado fruit. No mRNA is detected until the initiation of ripening which is indicated by autocatalytic ethylene biosynthesis /30/. The level of P450 antigen and chromophore in microsomes from avocado fruit parallels the increase in CYP71A1 mRNA /8/. Microsomes from fully ripe fruit have a relatively high P450 chromophore content of which the overwhelming contribution is from the CYP71A1 protein.

The cloning and sequencing of the CYP71A1 cDNA from avocado has allowed the heterologous expression of the corresponding protein in yeast. Biochemical characterization has shown this P450 to catalyze monoterpene epoxidation. It has been previously suggested that this enzyme may be a lauric acid hydroxylase /5/, but this is less likely due to the low affinity of the CYP71A1 protein for lauric acid and higher affinity for monoterpenes such as nerol and geraniol /7,14/, although, as discussed above, these are not likely to be the natural substrates either. The precise functional role of CYP71A1 in avocado fruit ripening remains unknown. Due to the large number of changes in secondary metabolism that accompany fruit ripening, it seems logical that CYP71A1 is involved in the hydroxylation or epoxidation of a terpenoid. Fruit have evolved to attract animals and thus produce a wide array of compounds that are pleasing to the taste. Chemical defense compounds that would deter herbivores would also have to be catabolized during ripening. Ripening-related P450 enzymes are likely to be involved in the synthesis of flavor attractants and/or the catabolism of chemical deterrants. We have looked for potential natural substrates of the CYP71A1 protein using the type I substrate-binding assay of HPLC-fractionated avocado extracts to identify potential substrate candidates. A major peak of type I substrate binding activity is detected in extracts from unripe fruit (H. George and R.E. Christoffersen, unpublished work). Chemical characterization of this fraction is currently underway.

Higher plants contain a wide variety of terpenes or terpene derivatives /31/ and a number of well characterized steps in terpenoid metabolism are mediated by P450 enzymes /18/. With the wide array of potential terpenoid substrates in the plants, it might be expected that a correspondingly large number of P450 genes have evolved as well.

Only a few of these P450 enzymes have been biochemically characterized to date and even less is known about the genes involved. Intriguingly, a number of CYP71 gene family members have been identified by sequence analysis of cDNAs isolated by screening protocols that targeted P450 genes or, in the case of *Arabidopsis*, by random analysis of a large number of cDNAs. It thus appears that the CYP71 gene family is widely dispersed in the plant kingdom and may play a variety of functional roles unrelated to fruit ripening. Comparison of a number of full-length plant P450 protein sequences (Fig. 4) suggests that the CYP71 family has diverged more than other

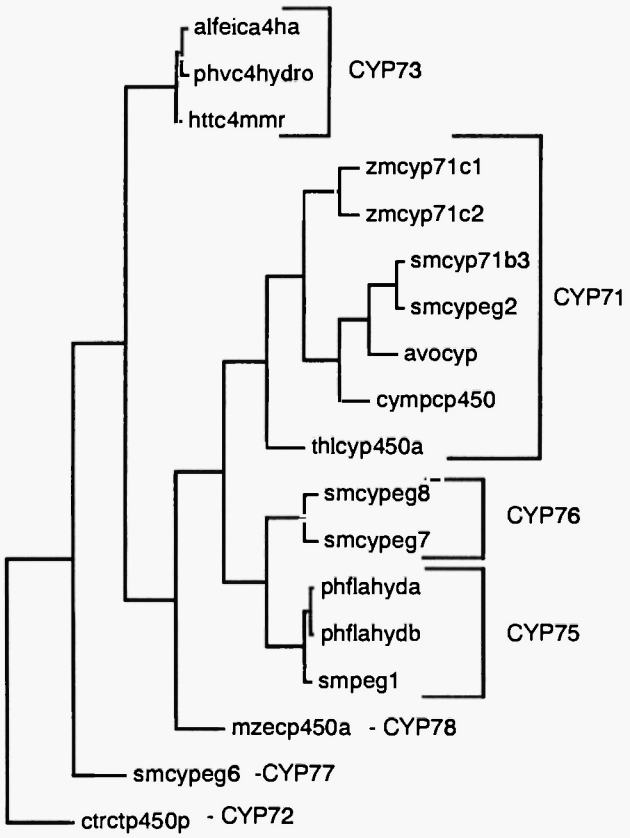

Fig. 4: Sequence relationships among plant P450 protein sequences. Protein sequences were obtained by translation of the corresponding DNA sequence obtained from the GenBank and individual sequences are indicated by their GenBank assigned locus name. The designated P450 family is indicated on the right.

P450 families such as the CYP73 (*t*-cinnamic acid 4-hydroxylase) or the CYP75 (flavonoid 3',5'-hydroxylase). Obviously, this conclusion is based on a very limited amount of sequence data and therefore is subject to revision as more plant P450 sequences become available. The hybridization data presented herein support the notion of a diverse CYP71 family consisting of approximately 10-12 members in avocado, and potentially an equally diverse set of substrate specificities. Under the conditions of hybridization used, it is unlikely that our probes would detect the avocado homologues of highly divergent P450 families such as CYP73 or CYP75.

An important observation made here is that the overall gene structure of the CYP71 family must be conserved. If an intron interrupted the region homologous to exon 2 of *CYP71A1*, the PCR product would have been larger by an unknown amount. A very large intron would prevent successful PCR amplification entirely. It is interesting that the protein domain predicted to be contained in exon 2 represents the heme-binding domain and is generally more conserved among P450 proteins than the region encoded by exon 1 of *CYP71A1*. This bipartite gene structure may have important implications for the evolutionary divergence of this P450 gene family.

ACKNOWLEDGMENTS

This work was funded by a National Science Foundation grant to R.E.C. The authors thank Dr. Laura G. Cass for technical advice, Dr. Dave L. Hallahan for help in primer design, and Dr. Helga George and Katherine Courtney for critically reviewing the manuscript.

REFERENCES

1. Donaldson RP, Luster DG. Multiple forms of plant cytochrome P-450. Plant Physiol 1991; 96: 669-674.
2. Durst F. Biochemistry and physiology of plant cytochrome P-450. In: Ruckpaul K, Rein H, eds. Microbial and Plant Cytochromes P-450: Biochemical Characteristics, Genetic Engineering and Practical Applications. v. 4. London: Taylor and Francis, 1991; 191-232.
3. Durst F, Benveniste I, Salaun JP, Werck-Reichhart D. Function, mechanism and regulation of cytochrome P-450 enzymes in plants. Trans Biochem Soc 1992; 20: 353-357.
4. Dohn DR, Krieger RL. *N*-Demethylation of *p*-chloro-*N*-methylaniline by

subcellular fractions from the avocado pear (*Persea americana*). Arch Biochem Biophys 1984; 231: 416-423.
5. Cottrell S, Hartmant GC, Lewis DFV, Parke DV. Studies on the cytochrome P-450 of avocado (*Persea americana*) mesocarp microsomal fraction. Xenobiotica 1990; 20: 711-726.
6. O'Keefe DP, Leto KJ. Cytochrome P-450 from the mesocarp of avocado (*Persea americana*). Plant Physiol 1989; 89: 1141-1149.
7. Hallahan DL, Nugent JHA, Hallahan BJ, et al. Interactions of avocado (*Persea americana*) cytochrome P-450 with monoterpenoids. Plant Physiol 1992; 98: 1290-1297.
8. Bozak KR, O'Keefe DP, Christoffersen RE. Expression of a ripening-related avocado (*Persea americana*) cytochrome P450 in yeast. Plant Physiol 1992; 100:
9. Nelson DR, Kamataki T, Waxman DJ, et al. The P450 superfamily: Update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. DNA and Cell Biology 1993; 12: 1-51.
10. Umemoto N, Kobayashi O, Ishizaki-Nishizawa O, Toguri T. cDNAs sequences encoding cytochrome P450 (CYP71 family) from eggplant seedlings. FEBS Lett 1993; 330: 169-173.
11. Udvardi MK, Metzger JD, Krishnapillai V, Peacock WJ, Dennis ES. Cloning and sequencing of a full-length cDNA from *Thlaspi arvense* L. that encodes a cytochrome P450. Plant Physiol 1994; 105: 755-756.
12. Newman T, de Bruijn FJ, Green P, et al. Genes Galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. Plant Physiol 1994; 106: 1241-1255.
13. Frey M, Kliem R, Saedler H, Gierl A. Expression of a cytochrome P450 gene family in maize. Mol Gen Genet 1995; in press.
14. Hallahan DL, Lau SMC, Harder PA, et al. Cytochrome p-450-catalysed monoterpenoid oxidation in catmint (*Nepeta racemosa*) and avocado (*Persea americana*) - evidence for related enzymes with different activities. Biochim Biophys Acta 1994; 1201: 94-100.
15. Meehan TD, Coscia CJ. Hydroxylation of geraniol and nerol by a monooxygenase from *Vinca rosea*. Biochem Biophys Res Comm 1973; 53: 1043-1048.
16. Madyastha KM, Meehan TD, Coscia CJ. Characterization of a cytochrome P-450 dependent monoterpene hydroxylase from the higher plant *Vinca rosea*. Biochemistry 1976; 15: 1097-1102.
17. Hallahan DL, Dawson GW, West JM, Wallsgrove RM. Cytochrome P-450 catalysed monoterpene hydroxylation in *Nepeta mussinii*. Plant Physiol Biochem 1992; 30: 435-443.
18. Mihaliak CA, Karp F, Croteau R. Cytochrome P-450 terpene hydroxylases. In: Charlwood BV, Banthorpe DV, eds. Terpenoids Methods in Plant Biochemistry. v. 7. London: Academic Press, 1991: 261-279.
19. Karp F, Mihaliak CA, Harris JL, Croteau R. Monoterpene biosynthesis: specificity of the hydroxylations of (-)-limonene by enzyme preparations from peppermint (*Mentha piperita*), spearmint (*Mentha spicata*), and perilla (*Perilla frutescens*) leaves. Arch Biochem Biophys 1990; 276: 219-226.

20. Croteau R, Karp F, Wagschal KC, Satterwhite DM, Hyatt DC, Skotland CB. Biochemical characterization of a spearmint mutant that resembles peppermint in monoterpene content. Plant Physiol 1991; 96: 744-752.
21. Funk C, Croteau R. Induction and characterization of a cytochrome P-450-dependent camphor hydroxylase in tissue cultures of common sage (*Salvia officinalis*). Plant Physiol 1993; 101: 1231-1237.
22. Jofuku KD, Goldberg RB. Analysis of plant gene structure. In: Shaw CH, ed. Plant Molecular Biology: a Practical Approach. Oxford: IRL Press, 1988: 37-66.
23. Cass LG, Kirven KA, Christoffersen RE. Isolation and characterization of a cellulase gene family member expressed during avocado fruit ripening. Mol Gen Genet 1990; 223: 76-86.
24. Feinberg AP, Vogelstein B. Addendum: A technique for radiolabelling DNA restiction endonuclease fragments to high specific activity. Anal Biochem 1984; 137: 266.
25. Church GM, Gilbert W. Genomic sequencing. Proc Natl Acad Sci USA 1984; 81: 1991-1995.
26. Devereux J, Haeberli P, Smithies O. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Research 1984; 12: 387-395.
27. Poulos TL. Modeling of mammalian P450s on basis of P450cam X-ray structure. Methods Enzymol 1991; 206: 11-30.
28. Ravichandran KG, Boddulpalli SS, Hasemann CA, Peterson JA, Deisenhofer J. Crystal structure of hemoprotein domain of P450BM-3, a prototype for microsomal P450's. Science 1993; 261: 731-736.
29. Nebert DW, Gonzalez FJ. P450 genes: structure, evolution and regulation. Ann Rev Biochem 1987; 56: 945-993.
30. Bozak KR, Yu H, Sirevåg R, Christoffersen RE. Sequence analysis of ripening-related cytochrome P-450 cDNAs from avocado fruit. Proc Natl Acad Sci USA 1990; 87: 3904-3908.
31. Charlwood BV, Banthorpe DV, eds. Terpenoids. Methods in Plant Biochemistry; v. 7. London: Academic Press, 1991.