Cytochrome P450 and Uses of Thereof

Joseph Chappell  
*University of Kentucky*, chappell@uky.edu

Lyle F. Ralston

Follow this and additional works at: https://uknowledge.uky.edu/ps_patents

Part of the Pharmacy and Pharmaceutical Sciences Commons

Right click to open a feedback form in a new tab to let us know how this document benefits you.

**Recommended Citation**

https://uknowledge.uky.edu/ps_patents/115

This Patent is brought to you for free and open access by the Pharmaceutical Sciences at UKnowledge. It has been accepted for inclusion in Pharmaceutical Sciences Faculty Patents by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
The invention features isolated cytochrome P450 polypeptides and nucleic acid molecules, as well as expression vectors and transgenic plants containing these molecules. In addition, the invention features uses of such molecules in methods of increasing the level of resistance against a disease caused by a plant pathogen in a transgenic plant, in methods for producing altered compounds, for example, hydroxylated compounds, and in methods for producing isoprenoid compounds.

6 Claims, 8 Drawing Sheets
OTHER PUBLICATIONS


O’Keefe and Leto, “Cytochrome P-450 from the Mesocarp of Avo-


Werck-Reichhart et al., “Cytochromes P450 for Engineering Herbi-


Whitehead et al., “Synthesis of (5)-5-epi-Aristolochene and (5)-


Wüst et al., “Hydroxylation of Limonene Enantiomers and Analogos by Recombinant (-)-Limonene 3- and 6-Hydroxylases from Mint (Mentha) Species: Evidence for Catalysis within Sterically Con-


GenBank Accession No. CA70575 (2004).
* cited by examiner
FIG. 1

farnesyl diphosphate

EAS

5-epi-aristolochene

+ NADPH + O2

1-deoxycapsidiol
(3-hydroxy-5-epi-aristolochene)

3-deoxycapsidiol
(1-hydroxy-5-epi-aristolochene)

capsidiol
FIG. 2

Enzyme activity (% of maximum) vs. Time after elicitation (h)
FIG. 4A

FIG. 4B

KETLRLH-for 5'-AARGARACIYTIMGITYICA-3'
KETLRLY-for 5'-AARGARACIYTIMGITYITA-3'
KETLRLR-for 5'-AARGARACIYTIMGITYIMG-3'
FXPERF -for 5'-TTYIIICCIIGARMGITYTITY'T-3'
FXPERF-rev 5'-ARRAICKYTCIGGIIIRAA-3'
GRRXCP(A/G)-for 5'-GGIMGIMGIITGYCCIGS-3'
PFGXGRR-rev 5'-CKICKICTCIIICCRAAIGG-3'
T7 5'-GTAATACGACTCACTATAGGG-3'
T3 5'-CAATTAACCCTCACTAAAGGG-3'

FIG. 4C
FIG. 5

<table>
<thead>
<tr>
<th>control</th>
<th>0</th>
<th>24</th>
</tr>
</thead>
<tbody>
<tr>
<td>elicitor-treated</td>
<td>0.5</td>
<td>1.5</td>
</tr>
</tbody>
</table>

- CYP71D
- CYP73A
- CYP82E
- CYP92A
- EAS
- Loading control
FIELD OF THE INVENTION

This invention relates to cytochrome P450s and uses thereof.

BACKGROUND OF THE INVENTION

Cytochrome P450s encompass a superfamily of oxidases responsible for the oxidation of numerous endobiotics and thousands of xenobiotics. In addition, in plants, cytochrome P450s play important roles in wound healing, pest resistance, signaling, and anti-microbial and anti-fungal activity.

Capsidiol is a bicyclic, dihydroxylated sesquiterpene produced by many Solanaceous species in response to a variety of environmental stimuli, including exposure to UV (Back et al., Plant Cell. Physiol. 389:899-904, 1998) and infection by microorganisms (Molot et al., Physiol. Plant Pathol. 379-389, 1981; Stolle et al., Phytopathology 78:1193-1197, 1988; Keller et al., Planta. 205:467-476, 1998). It is the primary antioxidant phytoalexin produced in tobacco in response to fungal elicitation, and it is derived from the isoprenoid pathway via its hydrocarbon precursor, 5-epi-aristolochene (Fig. 1). Several of the biosynthetic enzymes leading up to 5-epi-aristolochene formation have been studied (Chappell, Annu. Rev. Plant Physiol. Plant Mol. Biol. 46:521-547, 1995), especially 5-epi-aristolochene synthase (EAS) (Vögeli and Chappell, Plant Physiol. 88:1291-1296, 1988; Back and Chappell, Proc. Natl. Acad. Sci. U.S.A. 93:6841-6845, 1996; Mathis et al., Biochemistry 36:8340-8348, 1997; Starls et al., Science 277:1815-1820, 1997). EAS can commit to sesquiterpene metabolism by catalyzing the cyclization of farnesyl diphosphate (FPP) to 5-epi-aristolochene. However, until the present invention, the enzyme(s) responsible for the conversion of 5-epi-aristolochene to capsidiol has yet to be fully identified and characterized.

Biochemical evidence from previous studies in tobacco (Whitehead et al., Phytochemistry 28:775-779, 1988) and green pepper (Hoshino et al., Phytochemistry 38:609-613, 1995) have suggested that the oxidation of 5-epi-aristolochene to capsidiol occurs in a two-step process with one of the hydroxylation steps being constitutive and the other being mediated by an elicitor-inducible cytochrome P450 (Fig. 1). Because 1-deoxycapsidiol had been isolated from natural sources (Watson et al., Biochem. Soc. Trans. 11:589, 1983), Whitehead et al. (Phytochemistry 28:775-779, 1989), surmised that perhaps the biosynthesis of this intermediate was due to pathogen induction of a corresponding hydroxylase. They therefore prepared synthetic 1-deoxycapsidiol and reported a modest conversion of this compound to capsidiol when fed to control or unelicited tobacco cell cultures.

This was further supported by their observation that radiolabeled 5-epi-aristolochene was only converted to capsidiol when fed to elicitor-induced cell cultures but not control cultures. Whitehead et al. (Phytochemistry 28:775-779, 1989) therefore concluded that the 3-hydroxylase, responsible for hydroxylation of 5-epi-aristolochene at C3 to generate 1-deoxycapsidiol, was pathogen/elicitor inducible, while the 1-hydroxylase, responsible for hydroxylating 1-deoxycapsidiol at the C1 to generate capsidiol, was constitutive. Hoshino et al. (Phytochemistry 38:609-613, 1995) added to the observations of Whitehead et al. (Phytochemistry 28:775-779, 1989) by directly measuring 3-hydroxylase-activity in microsomal preparations of arachidonic acid-elicited Capsicum annuum fruits and seedlings. These assays consisted of incubating 5-epi-aristolochene with microsome preparations and subsequently determining the amount of 1-deoxycapsidiol generated by a combination of thin-layer chromatography (TLC) separations and gas chromatography (GC). Their evidence demonstrated that the conversion of 5-epi-aristolochene to 1-deoxycapsidiol was dependent on both NADPH and O2, and that 1-deoxycapsidiol accumulation in vitro was arrested by the P450 antagonists carbon monoxide (Omura and Sato, J. Biol. Chem. 239:2370-2378, 1964), aneyimid (Coolbaugh et al., Plant Physiol. 62:571-576, 1978), and ketomomazol (Rademacher, Annu. Rev. Plant Physiol. Plant Mol. Biol. 51:501-531, 2000).

Recent results suggest that the hydroxylation of 5-epi-aristolochene is an important regulated step in capsidiol biosynthesis. In studies to evaluate the effectiveness of methyljasmonate as an inducer of capsidiol biosynthesis in tobacco cell cultures, Mandujano-Chávez et al. (Arch. Biochem. Biophys. 381:285-294, 2000), reported that the modest accumulation of this phytalexin was accomplished by a strong induction of EAS. This result implied that steps before or after the sesquiterpen cyclase reaction were limiting. Using an in vivo assay measuring the conversion rate of radiolabeled 5-epi-aristolochene to capsidiol, a very limited induction of the hydroxylase activities was observed in cells treated with methyl jasmonate relative to that in fungal elicitor-treated cells. This result pointed to the hydroxylase reactions as a potentially limiting step in capsidiol biosynthesis.

SUMMARY OF THE INVENTION

In one aspect, the invention features several isolated cytochrome P450 polypeptides (such as CYP711D20, CYP711D21, CYP73A27, CYP73A28, and CYP92A5, and P450s having substantial identity to these polypeptides), as well as isolated nucleic acid molecules that encode these P450s.

In related aspects, the invention features a vector (such as an expression vector) including an isolated nucleic acid molecule of the invention and a cell (for example, a prokaryotic cell, such as Agrobacterium or E. coli, or a eukaryotic cell, such as a mammalian, insect, yeast, or plant cell) including the isolated nucleic acid molecule or vector.

In yet another aspect, the invention features a transgenic plant or transgenic plant component including a nucleic acid molecule of the invention, wherein the nucleic acid molecule is expressed in the transgenic plant or the transgenic plant component. Preferably, the transgenic plant or transgenic plant component is an angiosperm (for example, a monocot or dicot). In preferred embodiments, the transgenic plant or transgenic plant component is a solanaceous, maize, rice, or cruciferous plant or a component thereof. The invention further includes a seed produced by the transgenic plant or transgenic plant component, or progeny thereof.

In another aspect, the invention features a method of providing an increased level of resistance against a disease caused by a plant pathogen in a transgenic plant. The method involves: (a) producing a transgenic plant cell including the nucleic acid molecule of the invention integrated into the genome of the transgenic plant cell and positioned for expression in the plant cell; and (b) growing a transgenic plant from the plant cell wherein the nucleic acid molecule is expressed.
in the transgenic plant and the transgenic plant is thereby provided with an increased level of resistance against a disease caused by a plant pathogen.

In another aspect, the invention features a method for producing an altered compound, the method including the steps of contacting the compound with one or more of the isolated polypeptides disclosed herein under conditions allowing for the hydroxylation, oxidation, demethylation, or methylation of the compound and recovering the altered compound.

In still another aspect, the invention features a hydroxylating agent including any of the isolated polypeptides disclosed herein.

In yet another embodiment, the invention features an isolated nucleic acid molecule that specifically hybridizes under highly stringent conditions to the complement of any one of the sequences described in SEQ ID NO:2 (CYP71D20), SEQ ID NO:4 (CYP71D21), SEQ ID NO:6 (CYP73A27), SEQ ID NO:8 (CYP73A28), or SEQ ID NO:12 (CYP92A5), wherein such a nucleic acid molecule encodes a cytochrome P450 polypeptide.

In another aspect, the invention features a host cell expressing a recombinant isoprenoid synthase and a recombinant cytochrome P450. In preferred embodiments, the host cell further expresses, independently or in combination, a recombinant acetyltransferase, methyltransferase, or fatty acyltransferase. In other preferred embodiments, the host expresses an endogenous or recombinant cytochrome reductase. Preferably, the host cell is a yeast cell, a bacterial cell, an insect cell, or a plant cell.

In a related aspect, the invention features a method for producing an isoprenoid compound, the method including the steps of: (a) culturing a cell that expresses a recombinant isoprenoid synthase and a recombinant cytochrome P450 under conditions wherein the isoprenoid synthase and the cytochrome P450 are expressed and catalyze the formation of an isoprenoid compound not normally produced by the cell; and (b) recovering the isoprenoid compound. In preferred embodiments, the host cell further expresses a recombinant acetyltransferase, a recombinant methyltransferase, or a recombinant fatty acyltransferase. In other preferred embodiments, the host cell expresses an endogenous or recombinant cytochrome reductase. Preferably, the host cell is a yeast cell, a bacterial cell, an insect cell, or a plant cell.

In yet another aspect, the invention features an isoprenoid compound produced according to the above-mentioned methods.

By “P450 polypeptide,” “cytochrome P450,” or “P450” is meant a polypeptide that contains a heme-binding domain and shows a CO absorption spectra peak at 450 nm according to standard methods, for example, those described herein. Such P450s may also include, without limitation, hydroxylase activity, dual hydroxylase activity, demethylase activity, or oxidase activity. Such enzymatic activities are determined using methods well known in the art.

By “polypeptide” is meant any chain of amino acids, regardless of length or post-translational modification (for example, glycosylation or phosphorylation).

By “substantially identical” is meant a polypeptide or nucleic acid exhibiting at least 80 or 85%, preferably 90%, more preferably 95%, and most preferably 97%, or even 98% identity to a reference amino acid sequence (for example, the amino acid sequence shown in SEQ ID NO:1, 3, 5, 7, and 11) or nucleic acid sequence (for example, the nucleic acid sequences shown in SEQ ID NO:2, 4, 6, 8, and 12, respectively). For polypeptides, the length of comparison sequences will generally be at least 16 amino acids, preferably at least 20 amino acids, and most preferably at least 25 amino acids, and most preferably at least 35 amino acids. For nucleic acids, the length of comparison sequences will generally be at least 50 nucleotides, preferably at least 60 nucleotides, more preferably at least 75 nucleotides, and most preferably at least 110 nucleotides.

Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine.

By an “isolated polypeptide” is meant a P450 polypeptide (for example, a CYP71D20 SEQ ID NO:1, CYP71D21 SEQ ID NO:3, CYP73A27 SEQ ID NO:5, CYP73A28 SEQ ID NO:7, or CYP92A5 SEQ ID NO:11 polypeptide) that has been separated from components that naturally accompany it. Typically, the polypeptide is isolated when it is at least 60%, by weight, free from the proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably at least 90%, and most preferably at least 99%, by weight, a P450 polypeptide. An isolated P450 polypeptide may be obtained, for example, by extraction from a natural source (for example, a plant cell); by expression of a recombinant nucleic acid encoding a P450 polypeptide; or by chemically synthesizing the protein. Purity can be measured by any appropriate method, for example, column chromatography, polyacrylamide gel electrophoresis, or by HPLC analysis.

By “derived from” or “obtained from” is meant isolated from or having the sequence of a naturally-occurring sequence (e.g., cDNA, genomic DNA, synthetic, or combination thereof).

By “isolated nucleic acid molecule” is meant a nucleic acid molecule, e.g., a DNA molecule, that is free of the nucleic acid sequence(s) which, in the naturally-occurring genome of the organism from which the nucleic acid molecule of the invention is derived, flank the nucleic acid molecule. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector; into an autonomously replicating plasmid or virus; or into the genomic DNA of a prokaryote or eukaryote; or that exists as a separate molecule (for example, a cDNA or a genomic or cDNA fragment produced by PCR or restriction endonuclease digestion) independent of other sequences. The term “isolated nucleic acid molecule” also includes a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequence.

By “specifically hybridizes” is meant that a nucleic acid sequence is capable of hybridizing to a DNA sequence at least under low stringency conditions, and preferably under high stringency conditions. For example, high stringency conditions may include hybridization at approximately 42°C in about 50% formamide, 0.1 mg/ml sheared salmon sperm DNA, 1% SDS, 2xSSC, 10% Dextran sulfate, a first wash at approximately 65°C in about 2xSSC, 1% SDS, followed by a second wash at approximately 65°C in about 0.1xSSC. Alternatively high stringency conditions may include hybridization at approximately 42°C in about 50% formamide, 0.1 mg/ml sheared salmon sperm DNA, 0.5% SDS, 5x SSPE, 1x Denhardt’s, followed by two washes at room temperature in 2xSSC, 0.1% SDS, and two washes at between 55-60°C in
0.2xSSC, 0.1% SDS. Reducing the stringency of the hybridization conditions may involve lowering the wash temperature and/or washing at a higher concentration of salt. For example, low stringency conditions may include washing in 2xSSC, 0.1% SDS at 40°C.

By "transformed cell" is meant a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a DNA molecule encoding (as used herein) a P450 polypeptide.

By "positioned for expression" is meant that the DNA molecule is positioned adjacent to a DNA sequence which directs transcription and translation of the sequence (i.e., facilitates the production of, for example, a P450 polypeptide, a recombinant protein, or an RNA molecule).

By "reporter gene" is meant a gene whose expression may be assayed; such genes include, without limitation, beta-glucuronidase (GUS), luciferase, chloramphenicol transacylase (CAT), green fluorescent protein (GFP), beta-galactosidase, herbicide resistant genes, and antibiotic resistance genes.

By "expression control region" is meant any minimal sequence sufficient to direct transcription. Included in the invention are promoter elements that are sufficient to render promoter-dependent gene expression controllable for cell-, tissue-, or organ-specific gene expression, or elements that are inducible by external signals or agents (for example, light-, pathogen-, wound-, stress-, or hormone-inducible elements or chemical inducers such as salicylic acid (SA) or 2,2-dichloroisonicotinic acid (INA)); such elements may be located in the 5' or 3' regions of the native gene or engineered into a transgenic construct.

By "operably linked" is meant that a gene and a regulatory sequence(s) are connected in such a way as to permit gene expression when the appropriate molecules (for example, transcriptional activator proteins) are bound to the regulatory sequence(s).

By "plant cell" is meant any self-propagating cell bounded by a semi-permeable membrane and typically is one containing a plastid. Such a cell also requires a cell wall if further propagation is desired. Plant cell, as used herein includes, without limitation, algae, cyanobacteria, seeds, suspension cultures, embryos, meristematic regions, callus tissue, leaves, roots, shoots, gametophytes, sporophytes, pollen, and microspores.

By "plant component" is meant a part, segment, or organ obtained from an intact plant or plant cell. Exemplary plant components include, without limitation, somatic embryos, leaves, stems, roots, flowers, tendrils, fruits, scions, and rootstocks.

By "transgene" is meant any piece of DNA which is inserted by artifice into a cell and typically becomes part of the genome, for example, the nuclear or plastidic genome, of the organism which develops from that cell. Such a transgene may include a gene which is partly or entirely heterologous (i.e., foreign) to the transgenic organism, or may represent a gene homologous to an endogenous gene of the organism.

By "transgenic" is meant any cell which includes a DNA sequence which is inserted by artifice into a cell and becomes part of the genome of the organism which develops from that cell. As used herein, the transgenic organisms are generally transgenic plants and the DNA (transgene) is inserted by artifice into the nuclear or plastidic genome. A transgenic plant according to the invention may contain one or more engineered traits.

By "pathogen" is meant an organism whose infection of viable plant tissue elicits a disease response in the plant tissue. Such pathogens include, without limitation, bacteria, myco-plasmas, fungi, insects, nematodes, viruses, and viroids. Plant diseases caused by these pathogens are described in Chapters 11-16 of Agrios, Plant Pathology, 3rd ed., Academic Press, Inc., New York, 1988.

By "increased level of resistance" is meant a greater level of resistance to a disease-causing pathogen in a transgenic plant (or cell or seed thereof) of the invention than the level of resistance relative to a control plant (for example, a non-transgenic plant). In preferred embodiments, the level of resistance in a transgenic plant of the invention is at least 20% (and preferably 30% or 40%) greater than the resistance of a control plant. In other preferred embodiments, the level of resistance to a disease-causing pathogen is 50% greater, 60% greater, and more preferably even 75% or 90% greater than a control plant; with up to 100% above the level of resistance as compared to a control plant being most preferred. The level of resistance is measured using conventional methods. For example, the level of resistance to a pathogen may be determined by comparing physical features and characteristics (for example, plant height and weight, or by comparing disease symptoms, for example, delayed lesion development, reduced lesion size, leaf wilting and curling, water-soaked spots, and discoloration of cells) of transgenic plants.

By "purified antibody" is meant antibody which is at least 60%, by weight, free from proteins and naturally-occurring organic molecules with which it is naturally associated. Preferably, the preparation is at least 75%, more preferably 90%, and most preferably at least 99%, by weight, antibody, for example, an acquired resistance polypeptide-specific antibody. A purified P450 antibody may be obtained, for example, by affinity chromatography using a recombinantly-produced P450 polypeptide and standard techniques.

By "specifically binds" is meant an antibody which recognizes and binds a P450 protein but which does not substantially recognize and bind other molecules in a sample, for example, a biological sample, which naturally includes a P450 protein such as CYP71D20, CYP71D21, CYP73A27, CYP73A28, or CYP92A5.

Other features and advantages of the invention will be apparent from the following description of the preferred embodiments thereof, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic diagram of a proposed alternative pathway for the biosynthesis of capsidiol in elicitor-treated Nicotiana tabacum cells. 5-epi-aristolochene is synthesized from FPP by the action of a sesquiterpene cyclase, 5-epi-aristolochene synthase (EAS), and is subsequently hydroxylated at C1 and C3 to form capsidiol.

FIG. 2 is a graph showing an induction time course for sesquiterpene cyclase enzyme activity and sesquiterpene hydroxylase activity in cellulase-treated cell cultures. Sesquiterpene cyclase (5-epi-aristolochene synthase, EAS) enzyme activity was determined in extracts prepared from control (open squares) and elicitor-treated (closed squares) cells collected at the indicated time points. Sesquiterpene hydroxylase activity was determined using an indirect assay for control (open circles) and elicitor-treated (closed circles) cells. Cell cultures were incubated with [3H]-5-epi-aristolochene for 3 hours ending at the indicated time points before quantifying the incorporation of radioactivity into extracellular capsidiol, a dihydroxylated form of aristolochene (Mandujano-Chávez et al., Arch. Biochem. Biophys. 381:285-294, 2000).

FIG. 3 is a series of graphs showing the dose dependent inhibition of 5-epi-aristolochene hydroxylase activity by
anecymidol and ketoconazole. Cell cultures were incubated in the presence of cellulase (0.5 µg/mL) plus the indicated concentrations of anecymidol (A) or ketoconazole (B) for 12 hours prior to measuring the in vivo 5-epi-astilbene hydroxylase activity in the cell suspension cultures (squares), or the EAS enzyme activity in extracts prepared from the collected cells (triangles). The in vitro activity of a purified EAS preparation (Hackett and Chappell, J. Biol. Chem. 270:7375-7381, 1995) was also measured at the indicated inhibitor concentrations as an additional test for non-specific effects of these inhibitors (circles).

**FIG. 4A** is a schematic diagram of the primary structure of a generalized cytochrome P450 with conserved domains used for the design of PCR primers highlighted (SEQ ID NO:26-29).

**FIG. 4B** is a list of the degenerate P450-specific primers (SEQ ID NO:30-36) that were used in various combinations with vector specific primers in the amplification of cytochrome P450 cDNA fragments.

**FIG. 4C** is a scanned image of an ethidium bromide-stained agarose gel showing the PCR products amplified from a directional cDNA library prepared with mRNA isolated from elicitor-treated cells using the degenerate primer GRRXCP (A/G)=or (SEQ ID NO:35) and the T vector-specific primer (SEQ ID NO:37). The T3 vector-specific primer is also shown (SEQ ID NO:38).

**FIG. 5** is a series of Northern blots showing the induction time course for CYP71D1, CYP73A, CYP82E, CYP92A, and EAS transcript accumulation in elicitor treated cells. Total RNA was extracted from tobacco suspension cultures incubated with the cellulase elicitor for the indicated durations, size fractionated by agarose gel electrophoresis under denaturing conditions, and transferred to a nylon membrane before probing with the respective full-length cDNAs. The uniformity of sample loading was verified by ethidium bromide staining of ribosomal RNA (loading control).

**FIG. 6** is a series of graphs showing carbon monoxide (CO) difference spectra of the microsomal fraction isolated from yeast expressing the CYP92A5 (A) and CYP71D20 (B) cDNAs. Expression of the respective plasmid constructs engineered into the yeast (WAT11) cells was induced by a galactose treatment, followed by isolation of microsomal preparations. The difference absorption spectra of microsomes incubated in the presence (solid lines) and absence (broken lines) of carbon monoxide was determined.

**FIG. 7** is a series of gas chromatograms of the reaction products formed upon incubation of microsomes isolated from WAT11 yeast cells containing the CYP71D20 expression construct (A and C) or vector control DNA (B and D) with sesquiterpene substrates. Microsomes isolated from the indicated yeast lines were incubated with 5-epi-astilbene (A and B) or 1-deoxycapsidol (C and D) in the presence (solid lines) or absence (dashed lines) of NADPH. The identities of 5-epi-astilbene, 1-deoxycapsidol, and capsidiol were verified by mass spectrometry.

**FIG. 8** is a sequence comparison of the amino acid sequence of *Nicotiana tabacum* 5-epi-astilbene (sesquiterpene) hydroxylase NcCYP71D20 (SEQ ID NO:1) with other plant terpene hydroxylases (SEQ ID NO:39-43). NcCYP71A5v1 (GenBank accession number CA1970575) catalyzes the mono-hydroxylation of nerol and geraniol, linear monoterpenes, while PaCYP71A1 (A35867) catalyzes the epoxidation of these substrates (Hahlo et al., Biochem. Biophys. Acta. 1201:94-100, 1994). McCYP71D18 (AAD44150) and MpCYP71D13 (AAD44151) catalyze the mono-hydroxylation at C6 and C3 of limonene, a cyclic monoterpene, respectively (Lupien et al., Arch. Biochem.

Biophys. 368:181-192, 1999). AtCYP701A3 (AAC39505) encodes for kaurene oxidase, which catalyzes a 3-step reaction including a hydroxylation followed by oxidation of a diene (Hellwell et al., Plant Physiol. 119:507-510, 1999). Shown are sequences from *Mentha piperita* (McCYP71D13; SEQ ID NO:39), *Mentha spicata* (McCYP71D18; SEQ ID NO:40), *Neepa racemosa* (NCY71A5vb; SEQ ID NO:41), *Nicotiana tabacum* (NCY71D20; SEQ ID NO:42), *Persica americana* (PaCYP71A1; SEQ ID NO:42), and *Arabidopsis thaliana* (CYP701A3; SEQ ID NO:43). Conserved residues are shaded.

**DETAILED DESCRIPTION**

Capsidiol is a bicyclic, dihydroxylated sesquiterpene produced by several Solanaceous species in response to a variety of environmental stimuli. It is the primary antimicrobial compound produced by *Nicotiana tabacum* in response to fungal elicitation, and it is formed via the isoprenoid pathway from 5-epi-astilbene. Much of the biosynthetic pathway for the formation of this compound has been elucidated, except for the enzyme(s) responsible for the conversion of the allylic sesquiterpene 5-epi-astilbene to its dihydroxylated form, capsidiol.

Accordingly, an in vivo assay for 5-epi-astilbene hydroxylase activity was developed and used to demonstrate a dose dependent inhibition of activity by anecymidol and ketoconazole, two well-characterized inhibitors of cytochrome P450 enzymes. Using degenerate oligonucleotide primers designed to the well-conserved domains found within most P450 enzymes, including the heme binding domain, cDNA fragments representing four distinct P450 families (CYP71, CYP73, CYP82, and CYP92) were amplified from a cDNA library prepared against mRNA from elicitor-treated cells using PCR. The PCR fragments were subsequently used to isolate full-length cDNAs (CYP71D20 (SEQ ID NO:2) and D21 (SEQ ID NO:4), CYP73A27 (SEQ ID NO:6) and A28 (SEQ ID NO:8), CYP82E1 (SEQ ID NO:10), and CYP92A5 (SEQ ID NO:12)), and these in turn were used to demonstrate that the corresponding mRNAs were all induced in elicitor-treated cells, albeit with different induction patterns.

There now follows a description of the cloning of several P450s from *Nicotiana tabacum*. These examples are provided for the purpose of illustrating the invention, and are not to be considered as limiting.

**Inhibition of the 5-epi-astilbene to Capsidiol Conversion by P450 Antagonists**

Using an indirect assay, a detailed induction time course of 5EAH activity in elicitor-induced cell cultures was determined relative to that of EAS activity (FIG. 2), the well-characterized sesquiterpene cyclase activity that catalyzes the formation of 5-epi-astilbene from FPP (FIG. 1). Using assays for EAS and 5EAH, EAS activity is not detectable in control cell cultures, but is induced significantly within 3 hours and reaches its maximal level within 15 to 18 hours of elicitor-treatment. Similar to the EAS enzyme activity, 5EAH activity was negligible in control cell cultures. Nonetheless, after an apparent lag phase of 8 hours, a rapid induction of hydroxylase activity was observed 10 to 15 hours post elicitor addition to the cell cultures, reaching a maximum by 18 hours followed by a rather gradual decline of 10% to 20% over the next 8 hours.

Tobacco cell suspension cultures treated with cellulase plus varying concentrations of anecymidol or ketoconazole
were pre-incubated for 12 hours before measuring the cells' ability to convert exogenous supplied [3H] labeled 5-epi-aristolochene to radioabeled capsid during a subsequent 3 hour incubation period (FIG. 3). Apparent activity of SEAH was inhibited in a dose-dependent manner with approximately 50% inhibition by either 25 μM anemomol or ketocanazole, and more than 80% by 75 μM anemomol and 95% by 100 μM ketocanazole (FIGS. 3A and B). Importantly, neither the in vitro activity of recombinant EAS nor the induction of EAS in the elicitor-treated cell cultures was significantly affected by anemomol at concentrations as high as 100 μM (FIG. 3A). Ketocanazole also does not appear to affect the in vitro activity of EAS. However, the indubibility of cyclase activity in elicitor-treated cell extracts was inhibited by ketocanazole at concentrations above 50 μM (FIG. 3B). Therefore, the specificity of ketocanazole as an inhibitor of P450 type reactions should be assessed at or below a concentration of 50 μM under these experimental conditions.

Isolation of Elicitor-inducible Cytochrome P450 CDNs

A two-step approach for the isolation of candidate P450 cDNAs was followed. A PCR strategy was first employed using a directional CDN library prepared against mRNA isolated from elicitor-induced cells as the template and degenerate PCR primers (FIG. 4). Sequence alignments of cytochrome P450s from multiple families across kingdoms were used to identify conserved regions to which a series of degenerate primers were prepared (FIGS. 4A and B). In cloning experiments, 450 to 550 bp products were expected from reactions utilizing the primer prepared to the heme-binding domain (GRRXCP(A/G)) (SEQ ID NO:27 and 28) and the 17 vector primer (FIG. 4C). The mixtures of reaction products were shotgun cloned, and approximately 100 of the cloned PCR fragments were sequenced. About half of the sequenced DNAs contained signature sequences typical of P450 enzymes as revealed by BlastX database searches, and these corresponded to typical plant P450 family members of the CYP71, CYP73, CYP92, and CYP82 classes. Each of these PCR fragments was isolated multiple times in separate experiments. In addition, we isolated full-length cDNAs for these P450 family members. Table 1 compares the similarity and identity of the full-length cDNAs of P450 family members with those of their nearest family member in the GenBank database. In addition, FIG. 8 shows a amino acid alignment of several terrpen cytochrome P450s. Alignments were performed using the algorithm of the MACVECTOR software suite.

### Table 1

<table>
<thead>
<tr>
<th>Cytochrome P450</th>
<th>Nearest relative/ accession number</th>
<th>% identity</th>
<th>% similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYP71D20</td>
<td>CYP71D7 (S. chamaejasme) EMBL U84835</td>
<td>76.5</td>
<td>88.8</td>
</tr>
<tr>
<td>CYP71D21</td>
<td>CYP71D7 (S. chamaejasme) GenEMBL U848435</td>
<td>76.3</td>
<td>88.8</td>
</tr>
<tr>
<td>CYP73A27</td>
<td>CYP73A15 (E. vulgare) GenEMBL X904947</td>
<td>79.4</td>
<td>92.6</td>
</tr>
<tr>
<td>CYP73A28</td>
<td>CYP73A15 (F. vulgaris) GenEMBL Y094447</td>
<td>70.2</td>
<td>92.4</td>
</tr>
<tr>
<td>CYP82E1</td>
<td>CYP82E1 (N. tabacum) GenEMBL AB015762</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>CYP92A5</td>
<td>CYP92A3 (N. tabacum) GenEMBL X96784</td>
<td>95.5</td>
<td>98.6</td>
</tr>
</tbody>
</table>

The cloned fragments were used in a second step to isolate full-length clones from the CDN library. Screening the CDN library by hybridization with the CYP71 and CYP73 gene fragments yielded four full-length cDNAs, two CYP71D1s and two CYP73A3s. The former clones were designated CYP71D20 and CYP71D21, and the latter were designated CYP73A27 and CYP73A28. The other two CDN fragments corresponded to tobacco CDNAs already found in the GenBank database, CYP82E1 and CYP92A3. These two CDNAs were cloned using specific primers designed with the help of the available sequence information to amplify the full-length CDNAs.

**Induction of Cytochrome P450 mRNAs in Elicitor-treated Cells**

To correlate a biochemical role for P450s in saquinverpine metabolism, RNA blot analyses were used to determine the steady-state levels of the mRNAs coding for all four of the cytochrome P450 clones and EAS in control and elicitor-treated cells (FIG. 5). The mRNAs for all four of the P450s were rapidly and transiently induced with slightly different time courses relative to one another and to the EAS mRNA. CYP73A27 mRNA, for instance, displayed an induction pattern similar to that of EAS with the maximum mRNA level occurring 9 to 12 hours after elicitation. While the EAS mRNA remained high throughout the duration of the experiment, the CYP73A27 mRNA was negligible in cells 24 hours after elicitor-treatment. In contrast, the CYP71D3 mRNA was more rapidly induced than the EAS mRNA, reaching its maximum 6 to 9 hours after elicitation, and was declining by 12 hours when the EAS mRNA level was still very high.

**Functional Identification of CYP71D20 as 5-epi-aristolochene Hydroxylase**

To ascribe functional identity to the various P450 cDNAs, full-length cDNAs for CYP71D20, CYP82E1 and CYP92A5 were inserted into the yeast expression vector pYeDP60 (Urban et al., Biochimie 72:463-472, 1990; Pompon et al., Methods Enzymol. 272:51-64, 1996) and the expression of each in WAT11, a yeast line containing an integrated Arabidopsis thaliana cytochrome reductase gene (Pompon et al., Methods Enzymol. 272:51-64, 1996; Urban et al., J. Biol. Chem. 272: 19176-19186, 1997), was determined. Engineering the CYP73A27 CDN required an extra modification because of an unusually long N-terminus with several hydrophilic residues that may interfere with proper intracellular targeting (Nedelkina et al., Plant Mol, Biol. 39:1079-1090, 1999). This unusual leader sequence therefore was replaced with the membrane anchoring sequence of CYP73A1, a cytochrome P450 that previously demonstrated to express well in yeast (Fahrenhorst and Dixon, Arch. Biochem. Biophys. 305: 509-515, 1993; Pompon et al., Methods Enzymol. 272:51-64, 1996). Expression of all these CDNAs was under the control of the glucose-repressible, galactose-inducible GAL10-CYC1 promoter (Guarente et al., Proc. Natl. Acad. Sci. U.S.A. 79:7410-7414, 1982), and expression was compared to yeast transformed with the parent pYeDP60 vector (control) alone.

After induction with galactose for approximately 16 hours, control cells and cells containing the various P450 constructs were collected, and microsomes prepared from each were analyzed for general P450 expression by CO-difference spectroscopy (Omura and Sato, J. Biol. Chem. 239:2370-2378, 1964). Microsomes prepared from cells containing the CYP71D20 (FIG. 6A) and CYP92A5 (FIG. 6B) constructs both showed characteristic CO difference spectra with peaks at 450 nm, indicating that the encoded proteins were assembling properly with their heme cofactor. Using the extinction coefficient of 91 mM⁻¹ cm⁻¹ for heme binding proteins (Omura and Sato, J. Biol. Chem. 239:2370-2378, 1964), it
was determined that approximately 107 pmol of CYP71D20 and 268 pmol of CYP92A5 were expressed in the yeast cells per milligram of total yeast protein.

Both 5-epi- aristolochene and 1-deoxy capsidiol were metabolized to only one product with the same retention time as capsidiol. Obvious by its absence, no reaction product having a retention time similar to deoxy capsidiol was detectable in the 5-epi- aristolochene incubations (FIG. 7). Co-injection of authentic capsidiol with the respective reaction products resulted in a single GC peak having a 16.2 minute retention time, identical to capsidiol. Mass spectrometric data for the separate reaction products were identical to that for the capsidiol standard (EIMS m/z 236, 221, 203, 185, 175, 163, 157, 133, 121, 107, 93, 79, 67, 65, 53, 41).

The in vivo assay data presented in FIGS. 2 and 3 of the current work indicate that the conversion of 5-epi-aristolochene is catalyzed by at least one inducible cytochrome P450 mediated reaction.

Furthermore, any of the cytochrome p450 polypeptides described herein may include one or more hydroxylase activities which can incorporate hydroxyl groups into at least two distinct sites on an isoprenoid compound. The addition of these hydroxyl groups may occur, for example, sequentially, by adding a hydroxyl group first to one site and then the other, in either order. Moreover, such hydroxylations may be mutated to limit their ability to hydroxylate a substrate at only one site, or, alternatively, to provide stereochemical specificity to their hydroxylating activity.

The above-described experiments were performed using the following materials and methods.

Chemicals

Standard laboratory reagents were purchased from Becton Dickinson Microbiology Systems (Spark, Md.), FisherBiotek (Fair Lawn, N.J.) and Sigma Chemical Company (St. Louis, Mo.).

Biological Materials and Induction Treatments

Nicotiana tabacum cv. KY14 plants and cell suspension cultures were used. Cell suspension cultures were maintained in modified Murashige-Skoog (Vögeli and Chappell, Plant Physiol. 88:1291-1296, 1988) medium in their rapid phase of growth (3 days old) were used for all experiments. At the indicated times, cells were collected and separated from media by vacuum filtration and stored at −80°C.

Induction treatments were performed by the addition of the fungal elicitors, cellutase (Trichoderma viride, Type RS, Onozuka) or paractitcinein (O'Donohue et al., Plant Mol. Biol. 27:577-586, 1995) at the indicated concentrations. Paractitcinein was purified from E. coli cells overexpressing a recombinant paractitcinein protein containing a carboxy-terminal hisidine purification tag.

In Vivo 5-epi-aristolochene Hydroxylase Assay and Inhibition Studies

5-epi-aristolochene hydroxylase activity was measured as the incorporation of [1H]-5-epi-aristolochene into extracellular capsidiol by intact cells. [1H]-5-epi-aristolochene was produced by incubating an excess of [12H] farnesyli diphosphate (1 μM, 20.5 Ci/mmole) with recombinant [1H]-epi-aristolochene synthase (Buck et al., Arch. Biochem. Biophys. 315:527-532, 1994; Rising et al., J. Am. Chem. Soc. 122: 1861-1866, 2000). The hexane extractable radioactivity from reactions was treated with a small amount of silicone to remove any farnesoic or residual FPP before quantifying the yield of radioactive 5-epi-aristolochene by liquid scintillation counting. The hexane solvent was removed under a gentle stream of N2 gas, and the dried residue was re-dissolved in acetone.

Control and elicitor-treated cells were then incubated with [1H]-5-epi-aristolochene (approximately 100,000 dpm at 2.5 μM) for 3 hour periods at various points during an induction time course before collecting the cell and media samples. Detection and quantification of capsidiol in the extracellular culture media was performed as reported previously (Chappell et al., Phytochemistry 26:2259-2260, 1987), and the amount of radioactivity incorporated into capsidiol was determined. For these determinations, samples were separated by TLC, and the zones corresponding to capsidiol were scraped from the plate for scintillation counting.

Inhibition studies were performed by the addition of the P450 inhibitors ancydimol (Coolbaugh et al., Plant Physiol. 62:571-576, 1978; Hoshino et al., Phytochemistry 38:609-613, 1995) and ketoconazole (Hoshino et al., Phytochemistry 38:609-613, 1995; Rademaker, Annu. Rev. Plant Physiol. Plant Mol. Biol. 51:501-531, 2000) directly to the cell cultures or enzyme assay mix. Cell cultures were incubated in the presence of cellulase (0.5 μg/ml) and indicated concentrations of ancydimol or ketoconazole for 12 hours prior to the addition of [1H]-5-epi-aristolochene. After a further 3 hour incubation period, the cells and media were collected. The amount of radioactivity incorporated into extracellular capsidiol was determined as described above. To evaluate secondary effects of these inhibitors, the level of inducible sesquiterpene cyclase activity in the collected cells was determined according to (Vögeli et al., Plant Physiol. 93:182-187, 1990), as well as in vitro assays with purified recombinant EAS (Back et al., Arch. Biochem. Biophys. 315:527-532, 1994) incubated with the indicated concentrations of ancydimol and ketoconazole.

All experiments were replicated in several independent trials. While the absolute values presented may have varied between experiments by as much as 50%, the trends and time courses were consistent throughout.

Construction of an Elicitor-induced cDNA Library

Cell cultures were incubated with fungal elicitor (0.5 μg cellulase/ml) for 6 hours before collecting the cells by filtration. The cells were kept frozen at −80°C until total RNA was extracted from them using Trizol (Life Technologies, Rockville, Md.) according to the manufacturer’s instructions. Poly (A)+ RNA was purified by two rounds of oligo (dT) cellulose column chromatography (Life Technologies, Rockville, Md.). cDNA synthesis and library construction were subsequently carried out using the UNI-ZAP XR library kit (Stratagene, La Jolla, Calif.), according to manufacturer’s instructions.

PCR Cloning Strategy

Cytochrome P450 cDNA fragments were amplified from the elicitor-induced cDNA library using various combinations of degenerate forward and reverse primers with the vector-specific T3 and T7 primers. The template DNA was prepared from a 500 μl aliquot of the elicitor-induced cDNA library (3x108 plfu/μl) by heat denaturation at 70°C for 10 minutes, followed by phenol/chloroform extraction, ethanol precipitation and re-suspension in 500 μl of sterile, deionized water. Amplification reactions were performed in 50 μl volumes containing 50 mM KCl; 10 mM Tris-HCl, pH 8.8; 1.5 mM MgCl2; 200 μM of each dNTP; 2 μl template DNA; 20 pmol each of forward and reverse primer; and 1 unit Taq Polymerase (Life Technologies, Rockville, Md.). Reactions were preheated at 94°C for 2 minutes, followed by thirty-five cycles of denaturing at 94°C for 1 minute, annealing at 50°C for 1 minute 30 seconds, and polymerization at 72°C for 2 minutes. The reactions were completed by a 10-minute extension at 72°C. Aliquots of the reaction products were
examined for DNA products by agarose gel fractionation, and ligated directly into the pGEM-T Easy vector (Promega, Madison, Wis.). Resulting recombinant plasmids containing insert DNAs within the expected size range were sequenced using T7 and Sp6 primers.

DNA Sequencing

All the DNA sequencing reactions were performed using the BIGDY™ Terminator Cycle sequencing kit (Perkin-Elmer, Wellesley, Mass.) with the sequences being read on an automated ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, Calif.). Computer assessment of the DNA sequence information was performed using the MACVECTOR (Oxford Molecular, Madison, Wis.) software package.

cDNA Library Screening

The cDNA library was screened with digoxigenin labeled probes. A 258 bp DNA fragment amplified from the pGEM-deg.4 sequence using gene-specific primer (5'-GGCG-GGAAATTTCCTGGAAATGATCTTGGTATG-3' (SEQ ID NO: 13)) and reverse (5'-GTCTCAATATCTTTATATT-3' (SEQ ID NO: 14)); and 374 bp DNA fragment amplified from the pBKCS-CYVPB3.843 clone with forward (5'-GCGGTTGTGTAAGCTCATG-3' (SEQ ID NO: 15)) and reverse (5'-TTATGAGACAAATGG-GCCTGAAAGCA-3' (SEQ ID NO: 16)) primers, were used to screen for CYP71D1s. The probes were labeled with digoxigenin-11-dUTP using the PCR DIG Labeling Mix (Roche Molecular Biochemicals, Indianapolis, Ind.), hybridized to plaque lifts of the cDNA library plated at approximately 10,000 PFUs per 150 mm plate, and was hybridization detected with the DIG detection system according to the manufacturer’s instructions (Roche Molecular Biochemicals, Indianapolis, Ind.). Plaques exhibiting strong hybridization were plaque purified, auto-subcloned to their plasmid forms according to the manufacturer’s recommendations (Stratagene, La Jolla, Calif.), and then subjected to DNA sequencing as described above.

RNA Analysis

RNA gel blots analysis was carried out using 10 µg aliquots of total RNA. RNA samples were heat-denatured at 70°C for 15 minutes in sample buffer (1x MOPS, 5% formaldehyde, 1% bromophenol blue, 30% glycerol, and 3% ethidium bromide), and size fractionated on a 1.2% agarose gel containing 1x MOPS and 18.1% formaldehyde. Uniformity of sample loading was determined by visual inspection of the gel for RNA bands. The RNAs were then transferred to a Zeta Probe nylon membrane (Bio-Rad Laboratories, Hercules, Calif.) and hybridized according to the manufacturer’s recommendations. Full-length cDNAs were probed with [32P]dCTP (PRIME-IT Kit, Stratagene, La Jolla, Calif.) prior to hybridization. After hybridization, the membranes were washed in 2xSSC/0.1% SDS once at room temperature followed by sequential washes in 0.2xSSC/0.1% SDS at 42°C. and 65°C. Hybridization was detected with a Phosphoimager (Molecular Dynamics, model 445 SI).

Construction of Yeast Expression Vectors.

The coding regions of the P450 cDNAs were cloned into the pYeDP60 expression vector (Urban et al., J. Biol. Chem. 272:19176-19186, 1990; Pompon et al., Methods Enzymol. 272:51-64, 1996). Appropriate BamHI, EcoRI, and SstI restriction sites (underlined) were introduced via PCR primers containing these sequences either upstream of the translation start site (ATG) or downstream of the stop codon (TAA or TGA). The primers used to amplify the CYP71D20 cDNA were 5'-GGGCGATCCATGCAGTTTCGACCCTTCTGATTCTCC-3' (SEQ ID NO: 17) and 5'-GGGGAATTCCTTACTCTCGAGAAGGGTTGATAGG-3' (SEQ ID NO: 18); for the CYP82E1 cDNA 5'-CCCGGGATCCGTAGTATCATGTC-3' (SEQ ID NO: 20); and for the CYP92A5 cDNA, 5'-GGCGGATTCCAGCTTCTGTGTGTCCAGCCCGCGCAAACTGTCGCTCGATTCCGCCTGCGAATACTA-3' (SEQ ID NO: 24) coding for the N-terminal sequence of CYP73A1 (GenEMBL Z 17369) up to the hinge region were used for the modification of the membrane anchoring segment of CYP73A27 to avoid possible problems with intracellular targeting due to the unusual N-terminus

[SNP information can be added here]

Yeast Expression Studies

Veriﬁed pYeDP60-P450 cDNA constructs were introduced into the yeast WAT11 line, a derivative of the W303-1B strain (MAT a; ade 2-1; his 3-11; leu 2-3, -112; ura 3-1; can 6; cyr 4), provided by Dr. P. Urban (Centre de Génétique Moléculaire, CNRS, Gif-sur-Yvette, France). The endogenous NADPH-cytochrome P450 reductase (CPR1) locus has been replaced with ATR1, a NADPH-cytochrome P450 reductase from Arabidopsis thaliana (Pompon et al., Methods Enzymol. 272:51-64, 1996; Urban et al., J. Biol. Chem. 272:19176-19186, 1997), in the WAT11 line. Yeast was grown overnight in a 30°C shaker in YPAD (1 g/l yeast extract; 1 g/l peptone; 20 g/l glucose; 200 mg/l adenine) liquid media. Cultures were harvested in an A600 between 0.5 and 1.5. Cells were collected by centrifugation at 2,500 g for 5 minutes at 4°C, and resuspended in ice-cold, sterile dH2O. Cells were pelleted again as above and resuspended in 1 M sorbitol. Forty µl of yeast suspension was mixed with 0.5 to 1 µg plasmid DNA (in 5 µl dH2O) in a pre-chilled 0.5 ml tube, and transferred to a chilled cuvette with a 0.2 cm electrode gap. One pulse at 1.5 kV, 25 µF, and 200 Ohms was applied by an Eppendorf Electroporator (model 2510). A mixture of 500 µl of YPAD/1 M sorbitol was immediately added to the electroporated cells. Cells were allowed to recover at 30°C for 1 hour, then spread onto SGA plates (1 g/l bactoagar; 7 g/l yeast nitrogen base; 20 g/l glucose; 20 mg/l tryptophan; and 20 g/l agar). Transformed colonies appeared after 3 to 6 days of incubation at 30°C. Recombinant plasmids were conﬁrmed by PCR assays performed directly on randomly selected yeast colonies.
For expression studies, one colony was added to SGI media (1 g/l bacto casamino acids; 7 g/l yeast nitrogen base; 20 g/l glucose; and 20 mg/l tryptophan) and grown at 30°C for approximately 24 hours. An aliquot of this culture was diluted 1:50 into 250 ml of YPGE (10 g/l bactopeptone; 10 g/l yeast extract; 5 g/l glucose; and 3% ethanol by volume) and the cells were grown until all glucose was consumed. The absence of glucose was determined by placing a 200 μl aliquot of culture into a 1.5 ml tube, inserting a DIASTIX urinanalysis reagent strip (Bayer, Elkhart, Ind.) for 30 seconds, and observing colorimetric changes indicating glucose levels. Induction was initiated by the addition of 5 grams of galactose (final concentration of 2%). The cultures were maintained at 30°C for an additional 16 hours before collecting the cells by centrifugation at 7,000 g for 10 minutes. The pelleted cells were washed with 100 ml of TESS buffer (50 mM Tris-Cl pH 7.5; 1 mM EDTA; 0.6 M sorbitol). The cells were centrifuged as above, resuspended in 100 ml of TES-M (TES supplemented with 10 mM 2-mercaptoethanol), and allowed to incubate at room temperature for 10 minutes. The yeast cells were centrifuged again at 7,000 g for 10 minutes, and the pellet was resuspended in 2.5 ml extraction buffer (1% bovine serum albumin, fraction V; 2 mM 2-mercaptoethanol; 1 mM phenylmethylsulfonyl fluoride, all dissolved in TES). Glass beads (0.5 mm in diameter, Bioprec Products, Inc., Bartlesville, Okla.) were added until skimming the surface of the cell suspension. Cell walls were disrupted manually by hand shaking in a cold room for 10 min at 30 second intervals separated by 30 second intervals on ice. Cell extracts were transferred to a 50 ml centrifuge tube, the glass beads were washed three times with 5 ml of extraction buffer, and the washes were pooled with the original cell extracts. Microsomes were prepared by differential centrifugation at 10,000 g for 10 minutes at 4°C to remove cellular debris, followed by centrifugation at 100,000 g for 70 minutes at 4°C, and microsomal pellets were resuspended in 1.5 ml TEG-M buffer (50 mM Tris-Cl pH 7.5; 1 mM EDTA; 20% glycerol; and 1.5 mM 2-mercaptoethanol) and stored frozen at −80°C until further assayed.

CO Difference Spectra

Fe(CO) vs. Fe(S) difference spectroscopy (Omura and Sato, J. Biol. Chem. 239:2370-2373, 1964) was performed using 0.4 ml of microsomes suspended in 1.6 ml of 50 mM Tris-Cl pH 7.5; 1 mM EDTA; and 20% glycerol. A small amount of the reducing agent, sodium dithionite, was added, and the mixture was distributed between two cuvettes. A baseline was recorded between 400 and 500 nm on a Perkin Elmer Lambda 18 UV/visible spectrophotometer. CO was then bubbled into the sample cuvette for 1 minute, and the difference spectrum recorded again. The amount of functional P450 was estimated based on an absorbance coefficient of 91 M⁻¹ cm⁻¹.

5-epi-aristolochene-1, 3-hydroxylase Assays

5-epi-aristolochene-1, 3-hydroxylase assays were performed in 0.5 ml polyethylene tubes in 100 μl volumes. 5-epi-aristolochene or 1-deoxyxycapsidol dissolved in hexane was added to the tube, and the organic solvent was removed by incubation of the open tube at 30°C. 5-epi-aristolochene and 1-deoxyxycapsidol were resuspended in 2 μl dimethyl sulfoxide before adding the reaction mixture. Reactions were carried out in 100 mM Tris-Cl pH 7.5, to which microsomal protein was added to a final concentration of 1 mg/ml. Reactions were initiated by the addition of 2 mM NADPH. The final concentration of 5-epi-aristolochene and 1-deoxyxycapsidol in these assays varied from 20 to 50 μM. After incubations for variable lengths of time at 30°C, the reactions were extracted with two volumes of ethyl acetate. The organic extracts were concentrated and evaluated by GC and GC-MS along with standards of 5-epi-aristolochene (Whitehead et al., Phytochemistry 28:775-779, 1989; Rising et al., J. Am. Chem. Soc. 122:1861-1866, 2000), 1-deoxyxycapsidol (Whitehead et al., Phytochemistry 29:479-182, 1990), and capsidol (Whitehead et al., Phytochemistry 26:1367-1369, 1987; Milat et al., Phytochemistry 30:2171-2173, 1991). GC analysis was routinely performed with a Hewlett-Packard HP-5 capillary column (30 m × 0.25 mm, 0.25 μm phase thickness) and FID as described previously (Rising et al., J. Am. Chem. Soc. 122:1861-1866, 2000). GC-MS analysis was performed at the University of Kentucky Mass Spectrometry Facility using a Varian 3400 gas chromatograph and a Finnigan INCOS 50 quadrupole mass selective detector. The GC was equipped with a J&W DB-5 ms capillary column (15 m × 0.25 mm, 0.25 μm phase thickness) and run with He as the carrier gas (10 psi). Splitless injections were done at an injection port temperature of 280°C. The column temperature was maintained at 40°C for 1 minute and then increased to 280°C at 10°C per minute. Following separation by the GC column, samples were introduced directly into the electron impact ionization source. Mass spectra were acquired at 70 eV, scanning from 40-440 Da in 1 second.

Production of Cytochrome P450s

Using the standard molecular techniques described herein, the isolation of additional cytochrome P450 coding sequences is readily accomplished. For example, using all or a portion of the amino acid sequence of any of the disclosed P450s, one may readily design P450-specific oligonucleotide probes, including P450 degenerate oligonucleotide probes (i.e., a mixture of all possible coding sequences for a given amino acid sequence). These oligonucleotides may be based upon the sequence of either DNA strand and any appropriate portion of the P450 nucleotide sequence. General methods for designing and preparing such probes are provided, for example, in Ausubel et al., 2000, Current Protocols in Molecular Biology, Wiley Interscience, New York, and Berger and Kimmel, Guide to Molecular Cloning Techniques, 1987, Academic Press, New York. These oligonucleotides are useful for P450 gene isolation, either through their use as probes capable of hybridizing to a P450 complementary sequence, or as primers for various amplification techniques, for example, polymerase chain reaction (PCR) cloning strategies.

Hybridization techniques and screening procedures are well known to those skilled in the art and are described, for example, in Ausubel et al. (supra); Berger and Kimmel (supra); Chen et al., Arch. Biochem. Biophys. 324:255, 1995; and Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York. If desired, a combination of different oligonucleotide probes may be used for the screening of a recombinant DNA library. The oligonucleotides may be detectably-labeled using methods known in the art and used to probe filter replicas from a recombinant DNA library. Recombinant DNA libraries are prepared according to methods well known in the art, for example, as described in Ausubel et al. (supra), or they may be obtained from commercial sources.

As discussed above, P450 oligonucleotides may also be used as primers in a polymerase chain reaction (PCR) amplification cloning strategy. PCR methods are well known in the art and are described, for example, in PCR Technology, Erlich, ed., Stockton Press, London, 1989; PCR Protocols: A Guide to Methods and Applications, Inns et al., eds., Aca-
Additional methods for identifying sequences encoding P450 s are provided in Maughan et al (Arch. Biochem. Biophys. 341:104-111, 1997) and Clark et al. (Plant Mol. Biol. 33:875-885, 1997).

Useful P450 sequences may be isolated from any appropriate organism. Confirmation of a sequence’s relatedness to a P450 polypeptide disclosed herein may be accomplished by a variety of conventional methods, for example, by comparing the sequence with a known P450 sequence found in a database. In addition, the activity of any P450 may be evaluated according to any of the techniques described herein.

P450 Polypeptide Expression

P450 polypeptides may be produced by transformation of a suitable host cell with all or part of a P450 DNA (for example, any one of the P450 cDNAs described herein) in a suitable expression vehicle or with a plasmid construct engineered for increasing the expression of a P450 polypeptide in vivo.

Those skilled in the field of molecular biology will appreciate that any of a wide variety of expression systems may be used to provide the recombinant protein. The precise host cell used is not critical to the invention. The P450 protein may be produced in a prokaryotic host, for example, E. coli TB1, or in a eukaryotic host, for example, Saccharomyces cerevisiae, insect cells, mammalian cells (for example, COS 1 or NIH 3T3 cells), or any of a number of plant cells including, without limitation, algae, tree species, ornamental species, temperate fruit species, tropical fruit species, vegetable species, legume species, monocots, dicots, or in any plant of commercial or agricultural significance. Particular examples of suitable plant hosts include, but are not limited to, Conifers, Petunia, Tomato, Potato, Tobacco, Grape, Arabidopsis, Lettuce, Sunflower, Oilseed rape, Flax, Cotton, Sugarbeet, Celery, Soybean, Alfalfa, Medicago, Lotus, Vigna, Cucumder, Carrot, Eggplant, Cauliflower, Horseradish, Morning Glory, Poplar, Walnut, Apple, Asparagus, Grape, Rice, Maize, Millet, Onion, Barley, Orchard grass, Oat, Rye, Tobacco, and Wheat.


For prokaryotic expression, DNA encoding a P450 polypeptide is carried on a vector operably linked to control signals capable of effecting expression in the prokaryotic host. If desired, the coding sequence may contain, at its 5’ end, a sequence encoding any of the known signal sequences capable of effecting secretion of the expressed protein into the periplasmic space of the host cell, thereby facilitating recovery of the protein and subsequent purification. Prokaryotes most frequently used are various strains of E. coli; however, other microbial strains may also be used. Plasmid vectors are used which contain replication origins, selectable markers, and control sequences derived from a species compatible with the microbial host. Examples of such vectors are found in Pouwels et al. (supra) or Ausubel et al. (supra). Commonly used prokaryotic control sequences (also referred to as “regulatory elements”) are described herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences. Promoters commonly used to direct protein expression include the beta-lactamase (penicillinase), the lactose (lac), the trypthion (trp) (Goeddel et al. Nucil. Acids Res. 8:4087 (1980), and the tac promoter systems, as well as the lambda-derived PsuhL promoter and N-gene ribosome binding site (Simultake et al. Nature 292:128 (1981)).

One particular bacterial expression system for P450 production is the E. coli pET expression system (Novagen). According to this expression system, DNA encoding a P450 is inserted into a pET vector in an orientation designed to allow expression. Since the P450 gene is under the control of the T7 regulatory signals, P450 expression is dependent on inducing the expression of T7 RNA polymerase in the host cell. This is typically achieved using host strains which express T7 RNA polymerase in response to IPTG induction. Once produced, recombinant P450 is then isolated according to standard methods known in the art, for example, those described herein.

Another bacterial expression system for P450 production is the pGEX expression system (Pharmacia). This system employs a GST gene fusion system that is designed for high-level expression of a gene or gene fragment as a fusion protein with rapid purification and recovery of the functional gene product. The P450 of interest is fused to the carboxyl terminus of the glutathione S-transferase protein from Schistosoma japonicum and is readily purified from bacterial lysates by affinity chromatography using Glutathione Sepharose 4B. Fusion proteins can be recovered under mild conditions by elution with glutathione. Cleavage of the glutathione S-transferease domain from the fusion protein is facilitated by the presence of recognition sites for site-specific proteases upstream of this domain. For example, proteins expressed in pGEX-2T plasmids may be cleaved with thrombin; those expressed in pGEX-3X may be cleaved with factor Xa.

Other prokaryotic systems useful for expressing eukaryotic P450s are described by Cooper (Mol. Biol. Evol. 5:455-506).

One preferred eukaryotic expression system is the mouse 3T3 fibroblast host cell transfected with a pMAMneo expression vector (Clontech). pMAMneo provides an RSV-LTR enhancer linked to a dexamethasone-inducible MMTV-LTR promoter, an SV40 origin of replication which allows replication in mammalian systems, a selectable neomycin gene, and SV40 splicing and polyadenylation sites. DNA encoding a P450 is inserted into the pMAMneo vector in an orientation designed to allow expression. The recombinant P450 is then isolated as described below. Other preferable host cells which may be used in conjunction with the pMAMneo expression vector include COS cells and CHO cells (ATCC Accession Nos. CRL 1560 and CCL 61, respectively).

Alternatively, if desired, a P450 is produced by a stably-transfected mammalian cell line. A number of vectors suitable for stable transfection of mammalian cells are available to the public, e.g., see Pouwels et al. (supra); methods for constructing such cell lines are also publicly available, e.g., in Ausubel et al. (supra). In one example, cDNA encoding the P450 is cloned into an expression vector which includes the dihydrofolate reductase (DHFR) gene. Integration of the plasmid and, therefore, the P450-encoding gene into the host cell chromosome is selected for by inclusion of 0.01-300 μM methotrexate in the cell culture medium (as described in Ausubel et al., supra). This dominant selection can be accomplished in most cell types. Recombinant protein expression can be increased by DHFR-mediated amplification of the transfected gene. Methods for selecting cell lines bearing gene amplifications are described in Ausubel et al. (supra); such methods generally involve extended culture in medium containing gradually increasing levels of methotrexate. DHFR-containing expression vectors commonly used for this purpose include pCSEI1-DHFR and pAd26SV(A) (described in Ausubel et al., supra). Any of the host cells described above or, preferably, a DHFR-deficient CHO cell line (for example, CHO DHFR cells, ATCC Accession Number CRL 9096) are among the host cells preferred for DHFR selection of a stably-transfected cell line or DHFR-mediated gene amplification.

A cytochrome P450 may also be produced in insect cells, such cells include, without limitation, Spodoptera frugiperda (SF-21, or Drosophila melanogaster Schneider (SI-2)) cells. For P450 production, insect cells are typically infected with a baculovirus, for example, Autographa californiae Multiple Nuclear Polyhedrosis Virus (AcMNPV) containing an expression cassette for such a protein, e.g., cytochrome P450, at a multiplicity of infection of 1 to 10. The infected cells are generally cultured in a standard insect cell culture medium for 24 to 48 hours prior to recovering the protein using standard molecular biology techniques. If desired, the P450 polypeptide may also be produced in insect cells directly transfected with a DNA construct containing an expression cassette encoding the P450.

Furthermore, any of the cytochrome P450s described herein may be produced in yeast, for example, Pichia pastoris. In order to produce the P450 yeast cells are transformed with an expression cassette containing, for example, a promoter such as the AOX1 or phosphoglycerate kinase gene promoter, the P450 gene to be expressed, and a terminator. Such an expression cassette may contain an origin of replication or it may be integrated into the yeast genomic DNA. The expression cassette is generally introduced by lithium acetate transformation or by the use of spheroplasts. In order to select for successfully transformed cells, the yeast are plated, for example, on minimal media which only allows yeast carrying the introduced expression cassette to grow.

In addition, expression of recombinant proteins in yeast using a Hansenula polymorpha expression system is described in U.S. Pat. Nos. 5,741,674 and 5,672,487.

A P450 may also be produced by a stably-transfected plant cell line or by a transgenic plant. Such genetically-engineered plants are useful for a variety of industrial and agricultural applications as discussed below. Importantly, this invention is applicable to gymnosperms and angiosperms, and will be readily applicable to any new or improved transformation or regeneration method.

A number of vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants are available to the public; such vectors are described in Pouwels et al. (supra), Weissbach and Weissbach (supra), and Gelvin et al. (supra). Methods for constructing such cell lines are described in, e.g., Weissbach and Weissbach (supra), and Gelvin et al. (supra). Typically, plant expression vectors include (1) a cloned P450 gene under the transcriptional control of 5' and 3' regulatory sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (for example, one conferring inducible or constitutive expression, or environmentally- or developmentally-regulated, or pathogen- or wound-inducible, or cell- or tissue-specific expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

The P450 DNA sequence of the invention may, if desired, be combined with other DNA sequences in a variety of ways. The P450 DNA sequence of the invention may be employed with all or part of the gene sequences normally associated with a P450. In its component parts, a DNA sequence encoding a P450 is combined in a DNA construct having a transcription initiation control region capable of promoting transcription and translation in a host cell. In general, the constructs will involve regulatory regions functional in plants which provide for production of a P450 as discussed herein. The open reading frame coding for the P450, or a functional fragment thereof, will be joined at its 5' end to a transcription initiation regulatory region such as the sequence naturally found in the 5' upstream region of a P450 structural gene, for example, a CYP71D12 (SEQ ID NO:2) or CYP71D12 (SEQ ID NO:4) gene. Numerous other transcription initiation regions are available which provide for constitutive or inducible regulation.

For applications when developmental, cell, tissue, hormonal, environmental, or pathogen-inducible expression are desired, appropriate 5' upstream non-coding regions are obtained from other genes; for example, from genes regulated during seed development, embryo development, leaf development, or in response to a pathogen.

Regulatory transcription termination regions may also be provided in DNA constructs of this invention as well. Transcription termination regions may be provided by the DNA sequence encoding a P450 or any convenient transcription termination region derived from a different gene source. The transcript
termination region will contain preferably at least 1-3 kb of sequence 3' to the structural gene from which the termination region is derived.

An example of a useful plant promoter according to the invention is a caulimovirus promoter, such as a cauliflower mosaic virus (CaMV) promoter. These promoters confer high levels of expression in most plant tissues, and the activity of these promoters is not dependent on virally encoded proteins. CaMV is a source for both the 35S and 19S promoters. In most tissues of transgenic plants, the CaMV 35S promoter is a strong promoter (see, e.g., Odell et al., Nature 313:810 (1985)). The CaMV promoter is also highly active in monocots (see, e.g., Dekeyser et al., Plant Cell 2:591 (1990); Terada and Shimamoto, Mol. Gen. Genet. 220:389 (1990)). Moreover, activity of this promoter can be further increased (i.e., between 2-10 fold) by duplication of the CaMV 35S promoter (see, e.g., Kay et al., Science 236:1290 (1987); Ow et al., Proc. Natl. Acad. Sci. U.S.A. 84:4870 (1987); and Fang et al., Plant Cell 1:141 (1989)). Other useful plant promoters include, without limitation, the nopaline synthase promoter (An et al., Plant Physiol. 88:547 (1988)) and the octopine synthase promoter (Fromm et al., Plant Cell 1:977 (1989)).

For certain applications, it may be desirable to produce the P450 gene product in an appropriate tissue, at an appropriate level, or at an appropriate developmental time. For this purpose, there is an assortment of gene promoters, each with its own distinct characteristics embodied in its regulatory sequences, which have been shown to be regulated in response to the environment, hormones, and/or developmental cues. These include gene promoters that are responsible for heat-regulated gene expression (see, e.g., Callis et al., Plant Physiol. 88:965 (1988); Takahashi and Komeda, Mol. Gen. Genet. 219:365 (1989); and Takahashi et al., Plant J. 2:751 (1992)), light-regulated gene expression (e.g., the pea rbcS-3A described by Kuhlmeier et al. (plant Cell 1:471 (1989); the maize rbcS promoter described by Schaffner and Sheen, (Plant Cell 3:997 (1991); or the chlorophyll a/b-binding protein gene found in pea described by Simpson et al. (EMBO J. 4:2723 (1985)), hormone-regulated gene expression (for example, the abscisic acid (ABA) responsive sequences from the Em gene of wheat described by Marcotte et al. (Plant Cell 1:969 (1989); the ABA-inducible HVA1 and HVA22, and the ca92A promoters described for barley and Arabidopsis by Stauba et al. (Plant Cell 6:617 (1994); Shen et al. (Plant Cell 7:295 (1994)), and wound-induced gene expression (for example, of wunl described by Siebertz et al. (Plant Cell 1:961 (1989)), or organ-specific gene expression (for example, of the tuber-specific storage protein gene described by Roshal et al. (EMBO J. 6:1155 (1987); the 23-kDa zein gene from maize described by Schemtander et al. (EMBO J. 7:1249 (1988); or the French bean beta-phaseolin gene described by Busotos et al., (Plant Cell 1:389 (1989)) of pathogen-inducible gene expression described by Chappell et al. in U.S. Ser. Nos. 08/471,983, 08/443,639, and 08/577,483, hereby incorporated by reference.

Plant expression vectors may also optionally include RNA processing signals, for example, introns, which have been shown to be important for efficient RNA synthesis and accumulation (Callis et al., Genes and Dev. 1:1183 (1987)). The location of the RNA splice sequences can dramatically influence the level of transgene expression in plants. In view of this fact, an intron may be positioned upstream or downstream of a P450-encoding sequence in the transgene to modulate levels of gene expression.

In addition to the aforementioned 5' regulatory control sequences, the expression vectors may also include regulatory control regions which are generally present in the 3' regions of plant genes (Thornburg et al., Proc. Natl. Acad. Sci. U.S.A. 84:744 (1987); An et al., Plant Cell 1:115 (1989)). For example, the 3' terminator region may be included in the expression vector to increase stability of the mRNA. One such terminator region may be derived from the PI-II terminator region of potato. In addition, other commonly used terminators are derived from the octopine or nopaline synthase signals.

The plant expression vector also typically contains a dominant selectable marker gene used to identify those cells that have become transformed. Useful selectable genes for plant systems include genes encoding antibiotic resistance genes, for example, those encoding resistance to hygromycin, kanamycin, bleomycin, G418, streptomycin, or spectinomycin. Genes required for photosynthesis may also be used as selectable markers in photosynthetic-deficient strains. Alternatively, the green-fluorescent protein from the jellyfish Aequorea victoria may be used as a selectable marker (Sheen et al., Plant J. 8:777, 1995; Chiu et al., Current Biology 6:525 (1996)). Finally, genes encoding herbicide resistance may be used as selectable markers; useful herbicide resistance genes include the bar gene encoding the enzyme phosphinothricin acetyltransferase and conferring resistance to the broad-spectrum herbicide BASTA (Hoechst AG, Frankfurt, Germany).

Efficient use of selectable markers is facilitated by a determination of the expression level of a transgene in a plant cell to a particular selectable agent and a determination of the concentration of this agent which effectively kills most, if not all, of the transformed cells. Some useful concentrations of antibiotics for tobacco transformation include, e.g., 75-100 μg/ml (kanamycin), 20-50 μg/ml (hygromycin), or 5-10 μg/ml (bleomycin). A useful strategy for selection of transformants for herbicide resistance is described, e.g., by Vasil et al., supra.

It should be readily apparent to one skilled in the art of molecular biology, especially in the field of plant molecular biology, that the level of gene expression is dependent, not only on the combination of promoters, RNA processing signals, and terminator elements, but also on how these elements are used to increase the levels of selectable marker gene expression.

Plant Transformation

Upon construction of the plant expression vector, several standard methods are available for introduction of the vector into a plant host, thereby generating a transgenic plant. These methods include (1) Agrobacterium-mediated transformation (A. tumefaciens or A. rhizogenes) (see, e.g., Lichtenstein and Fuller, In: Genetic Engineering, vol. 6, PWJ Rigby, ed, London, Academic Press, 1987; and Lichtenstein, C. P., and Draper, J., In: DNA Cloning, Vol II, D. M. Glover, ed, Oxford, IRL Press, 1985), (2) the particle delivery system (see, e.g., Gordon-Kamm et al., Plant Cell 2:603 (1990); or BioRad Technical Bulletin 1687, supra), (3) microinjection protocols (see, e.g., Green et al., supra), (4) polyethylene glycol (PEG) procedures (see, e.g., Draper et al., Plant Cell Physiol. 23:451 (1982); or e.g., Zhang and Wu, Theor. Appl. Genet. 76:835 (1988)), (5) liposome-mediated DNA uptake (see, e.g., Freeman et al., Plant Cell Physiol. 25:1335 (1984)), (6) electroporation protocols (see, e.g., Gelvin et al., supra; Dekeyser et al., supra; Fromm et al., Nature 319:791 (1986); Sheen, Plant Cell 2:1027 (1990); or Jang and Sheen, Plant Cell 6:1665 (1994)), and (7) the vortexing method (see, e.g., Kindle, supra). The method of transformation is not critical to the present invention. Any method which provides for efficient transformation may be employed. As newer methods are available to transform crops or other host cells, they may be directly applied.
The following is an example outlining one particular technique, an Agrobacterium-mediated plant transformation. By this technique, the general process for manipulating genes to be transferred into the genome of plant cells is carried out in two phases. First, cloning and DNA modification steps are carried out in *E. coli*, and the plasmid containing the gene construct of interest is transferred by conjugation or electroporation into Agrobacterium. Second, the resulting Agrobacterium strain is used to transform plant cells. Thus, for the generalized plant expression vector, the plasmid contains an origin of replication that allows it to replicate in Agrobacterium and a high copy number origin of replication functional in *E. coli*. This permits facile production and testing of transgenes in *E. coli* prior to transfer to Agrobacterium for subsequent introduction into plants. Resistance genes can be carried on the vector, one for selection in bacteria, for example, streptomycin, and another that will function in plants, for example, a gene encoding kanamycin resistance or herbicide resistance. Also present on the vector are restriction endonuclease sites for the addition of one or more transgenes and directional T-DNA border sequences which, when recognized by the transfer functions of Agrobacterium, delimit the DNA region that will be transferred to the plant.

In another example, plant cells may be transformed by shooting into the cell tungsten microprojectiles on which cloned DNA is precipitated. In the Biolistic Apparatus (Bio-Rad) used for the shooting, a gunpowder charge (22 caliber Power Piston Tool Charge) or an air-driven blast drives a plastic macroprojectile through a gun barrel. An aliquot of a suspension of tungsten particles on which DNA has been precipitated is placed on the front of the plastic macroprojectile. The latter is fired at an acrylic stopping plate that has a hole through it that is too small for the macroprojectile to pass through. As a result, the plastic macroprojectile smashes against the stopping plate, and the tungsten microprojectiles continue toward their target through the hole in the plate. For the present invention, the target can be any plant cell, tissue, seed, or embryo. The DNA introduced into the cell on the microprojectiles becomes integrated into either the nucleus or the chloroplast.

In general, transfer and expression of transgenes in plant cells are now routine practices to those skilled in the art, and have become major tools to carry out gene expression studies in plants and to produce improved plant varieties of agricultural or commercial interest.

Transgenic Plant Regeneration

Plants cells transformed with plant expression vectors can be regenerated, for example, from single cells, callus tissue, or leaf discs according to standard plant tissue culture techniques. It is well known in the art that various cells, tissues, and organs from almost any plant can be successfully cultured to regenerate an entire plant; such techniques are described, e.g., in Vasil supra; Green et al., supra; Weissbach and Weissbach, supra; and Gelvin et al., supra.

In one particular example, a cloned P450, under the control of the EAS4 promoter and the nopaline synthase terminator and carrying a selectable marker (for example, kanamycin resistance), is transformed into Agrobacterium. Transformation of leaf discs (for example, of tobacco leaf discs), with vector-containing Agrobacterium is carried out as described by Horsch et al. (Science 227:1229 (1985)). Putative transformants are selected after a few weeks (for example, 3 to 5 weeks) on plant tissue culture media containing kanamycin (e.g., 100 µg/ml). Kanamycin-resistant shoots are then placed on plant tissue culture media without hormones for root initiation. Kanamycin-resistant plants are then selected for greenhouse growth. If desired, seeds from self-fertilized transgenic plants can then be sowed in soil-less medium and grown in a greenhouse. Kanamycin-resistant progeny are selected by sowing surface sterilized seeds on hormone-free kanamycin-containing media. Analysis for the integration of the transgene is accomplished by standard techniques (see, for example, Ausubel et al., supra; Gelvin et al., supra).

Transgenic plants expressing the selectable marker are then screened for transmission of the transgene DNA by standard immunoblot and DNA detection techniques. Each positive transgenic plant and its transgenic progeny is unique in comparison to other transgenic plants established with the same transgene. Integration of the transgene DNA into the plant genomic DNA is in most cases random, and the site of integration can profoundly affect the levels and the tissue and developmental patterns of transgene expression. Consequently, a number of transgenic lines are usually screened for each transgene to identify and select plants with the most appropriate expression profiles.

Transgenic lines are generally evaluated for levels of transgene expression. Expression at the DNA level is determined initially to identify and quantify expression-positive plants. Standard techniques for RNA analysis are employed and include PCR amplification assays using oligonucleotide primers designed to amplify only transgene RNA templates and solution hybridization assays using transgene-specific probes (see, e.g., Ausubel et al., supra). The RNA-positive plants are then analyzed for protein expression by Western immunoblot analysis using specific antibodies to the P450 (see, e.g., Ausubel et al., supra). In addition, in situ hybridization and immunocytochemistry according to standard protocols can be done using transgene-specific nucleotide probes and antibodies, respectively, to localize sites of expression within transgenic tissue.

Once the recombinant P450 is expressed in any cell or in a transgenic plant (for example, as described above), it may be isolated, e.g., using affinity chromatography. In one example, an anti-P450 antibody (e.g., produced as described in Ausubel et al., supra, or by any standard technique) may be attached to a column and used to isolate the polypeptide. Lysis and fractionation of P450-producing cells prior to affinity chromatography may be performed by standard methods (see, e.g., Ausubel et al., supra). Once isolated, the recombinant protein can, if desired, be further purified, for example, by high performance liquid chromatography (see, e.g., Fisher, Laboratory Techniques in Biochemistry and Molecular Biology, eds., Work and Burdon, Elsevier, 1980).

These general techniques of polypeptide expression and purification can also be used to produce and isolate useful P450 fragments or analogs.

Use

The aforementioned cytochrome P450 polypeptides of the invention are useful in the biosynthesis of hormones, lipids, and secondary metabolites, and may also help plants tolerate potentially harmful exogenous chemicals such as herbicides, pesticides, and pollutants. In addition, such cytochrome P450 polypeptides are useful in the chemical defense of plants against insects, as well as against bacterial, viral, and fungal infection.

Engineering Plant Disease Resistance

Plasmid constructs designed for the expression of a P450 gene product are useful, for example, for activating plant defense pathways that confer anti-pathogenic properties to a transgenic plant, for example, the production of phytoalexins. P450 genes that are isolated from a host plant (e.g., Nicotiana) may be engineered for expression in the same plant, a closely
related species, or a distantly related plant species. For example, a P450 gene may be engineered for constitutive low-level expression and then transformed into a Nicotiana host plant. Alternatively, the P450 gene may be engineered for expression in other solanaceous plants, including, but not limited to, potato and tomato. To achieve pathogen resistance, it is important to express a P450 protein at an effective level. Evaluation of the level of pathogen protection conferred to a plant by ectopic expression of the P450 gene is determined according to conventional methods and assays.

INDUSTRIAL APPLICATIONS

The invention also includes engineering host cells to include novel isoprenoid metabolic pathways useful in the production of new isoprenoid compounds. By introducing genes encoding an isoprenoid synthase (as disclosed in U.S. Pat. No. 5,824,774 and WO 00/17327) and a cytochrome P450, an acetyltransferase, a methyl transferase, a fatty acyltransferase, or a combination thereof, various isoprenoid reaction products may be modified, controlled, or manipulated, resulting in enhancement of production of numerous isoprenoid reaction products, for example, the production of novel monoterpenes, diterpenes, and sesquiterpenes. Such compounds are useful as phytalexins, insecticides, perfumes, and pharmaceuticals such as anti-bacterial and fungal agents.

In one working example, an isoprenoid synthase or a chimeric isoprenoid synthase (as disclosed in U.S. Pat. No. 5,824,774 and WO 00/17327) and a P450 gene are introduced into yeast, for example, using any of the procedures described herein. If desired, such cells may also express, either independently or in combination, an acetyltransferase (see, for example, Walker et al., Proc. Natl. Acad. Sci. USA 18;583-587, 2000), a methylase transferase (see, for example, Diener et al., Plant Cell 12;853-870, 2000) gene, or a fatty acyltransferase gene, as well as a cytochrome reductase. Cells are then cultured under standard conditions and the production of isoprenoid compounds is assayed according to methods known in the art. Isoprenoid compounds are further purified according to methods well known in the art. Cells expressing novel isoprenoid compounds are taken as useful in the invention.

Such methods provide a unique approach for producing novel isoprenoid starting materials and end products. Either prokaryotic or eukaryotic cells transformed with any of the aforementioned enzymes (or combinations thereof) may be used. Moreover, isoprenoid compounds may be produced in any number of ways known in the art including an in vitro combination of purified enzymes with an appropriate substrate or direct fermentation using a host cell which expresses any combination of the aforementioned enzymes and the appropriate substrates sufficient to drive production of isoprenoid compounds.

The invention is also useful for the production of insect attractants and deterrents, which may either deter insect pests or attract insect predators. In addition, the invention is also useful for generating novel flavorings and perfumes.

OTHER EMBODIMENTS

From the foregoing description, one skilled in the art can easily ascertain the essential characteristics of this invention, and can make various changes and modifications of the invention to adapt it to various usages and conditions. Thus, other embodiments are also within the claims.

All publications and patents mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent was specifically and individually indicated to be incorporated by reference.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 43
<210> SEQ ID NO 1
<211> LENGTH: 473
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 1

Met Gln Phe Phe Ser Leu Val Ser Ile Phe Leu Phe Leu Ala Phe Leu Phe Leu
1 5 10 15

Phe Leu Leu Arg Lys Trp Lys Asn Ser Asn Ser Gln Ser Lys Lys Leu 20 25 30

Pro Pro Gly Pro Trp Lys Ile Pro Ile Leu Gly Ser Met Leu His Met 35 40 45

Ile Gly Gly Glu Pro His His Val Leu Arg Asp Leu Ala Lys Lys Tyr 50 55 60

Gly Pro Leu Met His Leu Gln Leu Gly Glu Ile Ser Ala Val Val 65 70 75 80

Thr Ser Arg Asp Met Ala Lys Glu Val Leu Lys Thr His Asp Val 90 95

Phe Ala Ser Arg Pro Lys Ile Val Ala Met Asp Ile Ile Cys Tyr Asn
100 105 110
Gln Ser Asp Ile Ala Phe Ser Pro Tyr Gly Asp His Trp Arg Glu Met 115 120 125
Arg Lys Ile Cys Val Met Glu Leu Leu Amin Ala Lys Amin Val Arg Ser 130 135 140
Phe Ser Ser Ile Arg Arg Asp Glu Val Val Arg Leu Ile Asp Ser Ile 145 150 155 160
Arg Ser Asp Ser Ser Ser Gly Glu Leu Val Amin Phe Thr Glu Arg Arg 165 170 175
Ile Trp Phe Ala Ser Ser Met Thr Cys Arg Ser Ala Phe Gly Glu Val 180 185 190
Leu Lys Gly Glu Asp Ile Phe Ala Lys Ile Arg Glu Val Ile Gly 195 200 205
Leu Ala Glu Gly Phe Asp Val Val Asp Ile Phe Pro Thr Tyr Lys Phe 210 215 220
Leu His Val Leu Ser Gly Met Lys Arg Lys Leu Leu Amin Ala His Leu 225 230 235 240
Lys Val Asp Ala Ile Val Glu Asp Val Ile Amin Glu His Lys Amin 245 250 255
Leu Ala Ala Gly Lys Ser Amin Gly Ala Leu Glu Asp Met Phe Ala Ala 260 265 270
Gly Thr Glu Thr Ser Ser Thr Thr Thr Val Trp Ala Met Ala Glu Met 275 280 285
Met Lys Amin Pro Ser Val Phe Thr Lys Ala Gin Ala Glu Val Arg Glu 290 295 300
Ala Phe Arg Asp Lys Val Ser Phe Asp Gin Amin Asp Val Glu Leu 305 310 315 320
Lys Tyr Leu Leu Val Ile Lys Glu Thr Leu Arg Leu His Pro Pro 325 330 335
Ser Pro Leu Leu Val Val Pro Arg Glu Cys Arg Glu Asp Thr Asp Ile Amin 340 345 350
Gly Tyr Thr Ile Pro Ala Lys Thr Lys Val Val Met Val Amin Val Trp Ala 355 360 365
Leu Gly Arg Asp Pro Lys Tyr Trp Asp Asp Ala Glu Ser Phe Lys Pro 370 375 380
Glu Arg Phe Glu Gin Cys Ser Val Asp Phe Phe Gly Amin Amin Phe Glu 385 390 395 400
Phe Leu Pro Phe Gly Gly Arg Arg Ile Cys Pro Gly Met Ser Phe 405 410 415
Gly Leu Ala Amin Leu Tyr Leu Pro Leu Ala Gin Leu Leu Tyr Hist Phe 420 425 430
Asp Thr Lys Leu Pro Thr Gly Ile Met Pro Arg Asp Leu Asp Leu Thr 435 440 445
Glu Leu Ser Gly Ile Thr Ala Arg Lys Glu Asp Tyr Leu Amin 450 455 460
Ala Thr Pro Tyr Glu Pro Ser Arg Glu 465 470

<210> SEQ ID NO 2
<211> LENGTH: 1660
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<400> SEQUENCE: 2

ggatgttcta ataatctccc attatctccc gaaatgcaa ttcttcagct tggtttccat
tttcttctc ctactttccc tatattttgtt gaggaaatgg agaaactcca atagccaaag
caaaaatgt ccaccaggtc ctaggaataa accaacta acaaatctgc ggaagatgct ttcttatgat
tgttggagaa cggcactact tctcttaga tttagccaaa aatatggcac cacttatgca
ccctcagcta gggcagattt ctgcaagtgt ggtcttcctct agggcacttg gaaagagatg
getaaaaact cagcagtcg tttttgctac tggcctaaaa attggagca tggagcattat
tgtttataac cgctcgccaa ttgctcttag gctctatgtgc gcacactgga gaaagagctc
taaattaattg tctcgctagc aagaatttt ctggcttttc gaagtgagct
agttaaatattt cagcaccagga tctactgggt ctagagctcc atgaagcgtga tagcagcat
nttgccagta ctcacgagggc aagacataat tgcacaamaa atacacagag taatagatt
agcagagagc ttgtagtggt tagacatctt coccatacata cagttttcctt atgtttctcg
nttggttaga cgttaaaatct tgaattgcca ccaaattgta gacgtcattt gttgagatgt
catcaacagc ccaccagaaat atcctgcagc tggcacaagtt atgggctcat tagaagcacat
gttgctgccc ggaacagacat ctcctacac aaccacagtt tggctcattg gcctaatagat
gaaaatcca aagtaatttt ccaaacagaa cagcaagctg cgggagcactt ttgggcaaca
agtatttttt gatgaaatgt atgtggagga gctgaasaac ttaaagtgag tcattaaaga
aacctttgaga cttccacocg cttgtccacct tttgtgctccaa agagatacga ggagaagatc
ggtataaac ggtcacaact ttcctgcacaa gacaagacctgct gttgattagtt gttggccatt
gggagagat ccacaaattt ggagaagcgc ggaagacctt aacgccagac gatttgycga
atgtctctgag gatttttttt gtaataatttt tgcattttcc ctccttggcct cctcaattc
ctattaacctt gactcggaaac ttcacaccgag aactactgca aggaaacttt gcctaatcga
atattcagga taactcttagt ctagaaaggg ttgactttctac ttaaatgcca ctcttttatta
accccttctga ggtactatcac atatggtcatt aacacattta aatccctctt atcaacacct
atatgcaca actactctat ctcgctttgtt ataggttaata tgtgacctca aatcgttagt
ttttttat aaaaatgtact acattcatc tagaagcgag aacaacacat taatctgtcc
ataattttttag gttttacttg ttctactttaa aaaaaaaaa

<210> SEQ ID NO 3
<211> LENGTH: 474
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum
<400> SEQUENCE: 3

Met Glu Phe Phe Ser Leu Val Ser Ile Phe Leu Phe Leu Ser Phe Leu
1 5 10 15

Phe Leu Leu Arg Lys Cys Lys Asn Ser Asn Ser Glu Thr Lys Glu Leu
20 25 30

Pro Pro Gly Pro Trp Lys Ile Pro Ile Leu Gly Ser Met Leu His Met
35 40 45

Leu Gly Glu Pro His His Ile Leu Arg Asp Leu Ala Lys Tyr
50 55 60

Gly Pro Ile Met His Leu Lys Phe Gly Glu Ile Ser Ala Val Val Val
65 70 75 80

Thr Ser Arg Glu Met Ala Lys Glu Val Leu Lys Thr His Asp Val Val
<table>
<thead>
<tr>
<th>Phe Ala Ser Arg Pro Lys Ile Val Ala Met Asp Ile Ile Cys Tyr Asn</th>
<th>100</th>
<th>105</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gln Ser Asp Ile Ala Phe Ser Pro Tyr Gly Asp His Trp Arg Gln Met</td>
<td>115</td>
<td>120</td>
<td>125</td>
</tr>
<tr>
<td>Arg Lys Ile Cys Val Met Glu Leu Leu Asn Ala Lys Asn Val Arg Ser</td>
<td>130</td>
<td>135</td>
<td>140</td>
</tr>
<tr>
<td>Phe Ser Ser Ile Arg Arg Arg Glu Val Val Arg Leu Ile Asp Ser Ile</td>
<td>145</td>
<td>150</td>
<td>155</td>
</tr>
<tr>
<td>Arg Ser Asp Ser Ser Gly Glu Leu Val Asn Phe Thr Glu Arg Ile</td>
<td>165</td>
<td>170</td>
<td>175</td>
</tr>
<tr>
<td>Ile Trp Phe Ala Ser Ser Met Thr Cys Arg Ser Ala Phe Gly Glu Val</td>
<td>180</td>
<td>185</td>
<td>190</td>
</tr>
<tr>
<td>Leu Lys Gly Glu Asp Val Phe Ala Lys Ile Arg Glu Val Ile Gly</td>
<td>195</td>
<td>200</td>
<td>205</td>
</tr>
<tr>
<td>Leu Ala Glu Gly Phe Asp Val Ala Asp Ile Phe Pro Ser Tyr Lys Phe</td>
<td>210</td>
<td>215</td>
<td>220</td>
</tr>
<tr>
<td>Leu His Val Leu Ser Gly Met Lys Arg Lys Leu Leu Asn Ala His Leu</td>
<td>225</td>
<td>230</td>
<td>235</td>
</tr>
<tr>
<td>Lys Val Asp Ala Ile Val Glu Asp Val Ile Asn Glu His Lys Lys Asn</td>
<td>245</td>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>Leu Ala Thr Gly Lys Thr Arg Asn Gly Ala Leu Gly Glu Asp Met Phe Ala</td>
<td>260</td>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td>Ala Gly Thr Glu Thr Ser Ser Thr Thr Thr Val Trp Ala Met Ala Glu</td>
<td>275</td>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td>Met Met Lys Asn Pro Asn Val Phe Asn Lys Ala Glu Ala Glu Val Arg</td>
<td>290</td>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>Glu Thr Phe Lys Asp Lys Val Thr Phe Asp Glu Ile Asp Ala Glu Glu</td>
<td>305</td>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>Leu Glu Tyr Leu Lys Leu Val Ile Lys Glu Thr Leu Arg Leu His Pro</td>
<td>325</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>Pro Ser Pro Leu Leu Val Pro Arg Glu Cys Arg Glu Asp Thr Asp Ile</td>
<td>340</td>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td>Asn Gly Tyr Thr Ile Pro Ala Lys Thr Lys Val Met Val Asn Val Trp</td>
<td>355</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>Ala Leu Gly Arg Asp Pro Lys Tyr Trp Asp Asp Ala Glu Ser Phe Lys</td>
<td>370</td>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td>Pro Glu Arg Phe Glu Gln Cys Ser Val Asp Phe Phe Gly Asn Asn Phe</td>
<td>385</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td>Glu Phe Leu Pro Phe Gly Gly Gly Arg Arg Ile Cys Pro Gly Met Ser</td>
<td>405</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>Phe Gly Leu Ala Asn Leu Tyr Leu Pro Leu Ala Glu Leu Leu Tyr His</td>
<td>420</td>
<td>425</td>
<td>430</td>
</tr>
<tr>
<td>Phe Asp Trp Lys Leu Pro Ser Gly Met Met Pro Gly Asp Leu Asp Leu</td>
<td>435</td>
<td>440</td>
<td>445</td>
</tr>
<tr>
<td>Thr Glu Leu Ala Gly Ile Thr Ile Ala Arg Lys Gly Asp Leu Tyr Leu</td>
<td>450</td>
<td>455</td>
<td>460</td>
</tr>
<tr>
<td>Met Ala Thr Pro Tyr Glu Pro Ser Arg Glu</td>
<td>465</td>
<td>470</td>
<td></td>
</tr>
</tbody>
</table>

<seq> SEQ ID NO 4
<seq> LENGTH: 1614
<seq> TYPE: DNA
<org> ORGANISM: Nicotiana tabacum
<400> SEQUENCE: 4

ggatggtcata ataatctcc atatatctc ccaaattgaa ttcttcaagt tgggttctat 60
attccatac etatctttgc attttttgtt aaggaatgt aagaactca atagccaaac 120
caccaaatgg ccttccaggt gcttgggaat atcaaatctc ttctatgtcttgc 180
tggttgagaa ccaacccata tctttaagga tttttagcaca aatatgccac ccaatatgca 240
ccctcagttt ggtcgaaatt cttgccgttt ggtcctctct cgggcagatgg ccaaaaaggt 300
gttaaaaaact atagaagttg ctttttgctac ttaggctaaa atcttggtgca ctgacaattt 360
tctttataa ctgcttgata tgctcttttag ccctcatgga gactactgg cacaactggtg 420
taaattgtgt gcataagtgac ttctttatgc aagaattggt cgggcttttca gcctgtagcag 480
acgtgatgaa gcctggtgcct tcattaagcttc atataagttgctcttgta cggagctgct 540
agttattttt aacgcaaggg taatttttttg cggagatggc ctgaagttgta autagcagc 600
tgggcaagta ttaaagggcc aacgaatttg tttcaaaaaag attaggaag tattaggggt 660
acgagaaggg tttgatgtgg cgcgaatctt cacctttttg gacgtttctttt gcttgctatcg 720
tggaatgag cgttaagttgc tgtgcttccc taactcggtta gtagcatttt cttggagattg 780
cactaagcg caccaagggaa acctttttac tgggaaactc atcttgagcaac ctggaacatc 840
gtttgtgcccc gacagagaa ttctatccac acttactctgct gtagctagtt gtagaatgat 900
gaaagctttc ctttttttta cacaatcgca gacgagagtc cagagaaatc ttaaaagcaca 960
agtaaactttt gatgaaattg atcgcagagga gctggaatcc attttaagttg tattaaagag 1020
aactttttgc cttctattcc ctgctctttc ttttgcttcca agatcagtaa gggaagattc 1080
agatattacc gcgtcactta ttcctgggaag aaccaagtta atgttaattg ttttgccatt 1140
gggaagagac ccaaaaattt ggattgacct gcaagatttt acgcaagagc attttgccga 1200
atgtgtcttg gatttttttg ttaataattt tگtgtctcttt ccttttggcc gtggagagag 1260
aatgtttctg gattgttcgt cttttttctgc taattttttctg gcatctgctt ctcatttgct 1320
atataaactt cattttttttt ccgagaggg cttgatggcc aatgagccttt gcggagcatttt 1380
attactgtgga atccaatctt ctagaaggg cctgctttcact ttaatggtcag ctccttatca 1440
aactctcttg gataatctta atggcaccag tttttttatt tcaattttcc aaccttacttt 1500
ttggagacttt tgcagttttc cccttattgc atataatgag tttttttatt 1560
aggtatata cctgcttgcag agagagcagc ttaattagcttt gttttttttttt 1614

<210> SEQ ID NO 5
<211> LENGTH: 537
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 5

Met Lys Asn Met Ala Lys Leu Leu Asn Lys Thr Ile Phe Cys Ile Leu
1    5   10   15
Phe Thr Ile Ala Phe Leu Ser Phe Ala Lys Leu Leu Ser Ser Tyr Leu
20    25   30
Ser Met Pro Phe Pro Leu Lys Tyr Met Ser Leu Ile Val Pro Leu Leu
35    40   45
Pro Leu Ile Ile Asn Phe Leu Tyr Val Lys Pro Gln Asn Asn Leu Pro
50    55   60
Pro Gly Pro Thr Ala Val Pro Ile Phe Gly Asn Thr Leu Gln Val Gly
65    70   75   80
Aasn Asp Leu Aasn His Gln Leu Leu Ala Thr Met Ser Gln Thr Tyr Gly 85 90 95
Pro Ile Phe Leu Leu Lys Leu Gly Ser Lys Aasn Leu Ala Val Val Ser 100 105 110
Aasn Pro Glu Leu Ala Aasn Glu Val Leu His Thr Glu Gly Val Glu Phe 115 120 125
Gly Ser Arg Pro Arg Aasn Val Val Phe Asp Ile Phe Thr Gly Aasn Gly 130 135 140
Gln Aasn Met Val Phe Thr Ile Tyr Gly Asp His Trp Arg Lys Met Arg 145 150 155 160
Arg Ile Met Thr Leu Pro Phe Thr Aasn Lys Val Val His Glu Tyr 165 170 175
Ser Aasn Met Thr Leu Val Aasn Met Aasn Leu Val Val Aasn Asp Leu Lys 180 185 190
Lys Aasn Glu Lys Leu Val Lys Tyr Gly Glu Ile Val Ile Arg Lys Arg Leu 195 200 205
Gln Leu Met Leu Tyr Aasn Ile Met Met Tyr Arg Met Met Phe Asp Ala Lys 210 215 220
Phe Glu Ser Glu Aasn Asp Pro Leu Phe Ile Glu Ala Thr Lys Phe Aasn 225 230 235 240
Ser Glu Arg Ser Arg Leu Ala Gln Ser Phe Asp Tyr Aasn Tyr Gly Aasn 245 250 255
Phe Ile Pro Leu Leu Arg Pro Phe Leu Arg Gly Tyr Leu Aasn Lys Cys 260 265 270
Lys Aasn Leu Gln Thr Arg Arg Leu Ala Phe Phe Aasn Aasn Tyr Phe Val 275 280 285
Glu Lys Arg Arg Lys Ile Met Asp Aasn Gly Glu Lys His Lys Ile 290 295 300
Ser Cys Ala Ile Asp His Ile Ile Asp Ala Glu Met Lys Gly Glu Ile 305 310 315 320
Aasn Glu Aasn Val Leu Tyr Ile Val Glu Aasn Ile Aasn Val Ala Ala 325 330 335
Ile Glu Thr Thr Leu Trp Ser Met Glu Trp Ala Ile Ala Glu Leu Val 340 345 350
Aasn His Pro Ile Val Gln Gln Lys Ile Arg Asp Glu Ile Ser Thr Val 355 360 365
Leu Lys Gly Arg Ser Val Thr Glu Ser Aasn Leu His Glu Leu Pro Tyr 370 375 380
Leu Leu Ala Thr Val Aasn Glu Thr Leu Arg Leu His Thr Pro Ile Pro 385 390 395 400
Leu Leu Val Pro His Met Aasn Leu Glu Ala Lys Leu Gly Gly Tyr 405 410 415
Thr Ile Pro Lys Glu Thr Lys Val Val Aasn Ala Trp Trp Leu Ala 420 425 430
Aasn Asp Pro Ala Trp Tyr Lys Aasn Pro Aasn Glu Phe Arg Pro Glu Arg 435 440 445
Phe Leu Glu Glu Asp Ser Thr Glu Ala Val Ala Gly Gly Lys 450 455 460
Val Asp Phe Arg Tyr Leu Pro Phe Gly Met Gly Arg Arg Ser Cys Pro 465 470 475 480
Gly Ile Ile Leu Ala Leu Pro Ile Leu Gly Leu Val Ile Ala Lys Leu 485 490 495
Val Ser Aasn Phe Glu Met Gln Gly Pro Pro Gly Val Glu Lys Val Asp
<table>
<thead>
<tr>
<th>Thr</th>
<th>Ser</th>
<th>Glu</th>
<th>Arg</th>
<th>Gly</th>
<th>Gly</th>
<th>Gln</th>
<th>Phe</th>
<th>Ser</th>
<th>Leu</th>
<th>His</th>
<th>Ile</th>
<th>Ala</th>
<th>Lys</th>
<th>His</th>
<th>Ser</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>505</td>
<td>510</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Thr Val Val Phe Lys Pro Ile Ala Ala
530

535

<210> SEQ ID NO 6
<211> LENGTH: 1745
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 6

cctctagcta atgaaaaaac tggcacaact ttcacaacag accatctttt gcattctttt 60
tacaattgca tttctttcat ttcgcaagtt actgtctccc tacctatca taacgccct 120
tottaagta agttcaccat ttgctctctt actttctctt atatacaact atctatatgt 180
	taacgcccac aacaacctcc cacctggtcc aacagagttc ccataatattt gtaatttggt 240
tcaagtggg aagttcattga acacacacct ctgtggcacc atgtacaaca cctaacggtc 300
tatatattt cttcaaccttg gttcaaaaaa cctagctgtg gtataaacc cagagctagc 360
tgacacaaag aaggggtcga gtgtgggtctc gttccaccga aagttctgtct 420
cgacatatc actggttaag gacagacaggt ggtgtctcga attatagttg acatttggcg 480
	aaaaattgg cgtttattta gctttctcttt ttctcaacac aaagttggtcg accaatatac 540
tgatattgga gagaagttgaga tggactgtctg tgtatatttgc ggataaagaga atgaaaattg 600
gatattgag gaaatttttt gattggaattc ttggtgtcct cttcatacat acaattgtgta 660
tgaaatttgtc tatagttgtca aattttgacct tgaagaagttc ttcggtgcaac 720
	aaaaattaaa tccaagagaa gcagattggc tcagactttt gactaacaatt atyttgtgtt 780
tatccctttta attgagcact tttgaggg gttacottaac aatgttaaag accataacac 840
	aaaaaggact gccctctctc acaaatattt tggagaagaa agaagggaaaa taattgtgatga 900
	aaaaattgaga aaccaatagaa tatttttggtt cttattatac atattaggtg cagaattga 960
	aaaaagaaaa ataagaaaaa atgtactctta tttgttgag aataatcact tggcaacat 1020
tgaaacaact ctatgtggcca ttgatagggc ctagctgaa atttgaatct atcagattttgt 1080
tcaacagag attatggagt aacatctcaac acggagatcag tcaacagacct 1140
	aaaaatttcaagatgtccct aagcctgaaaa aacaactttaa gaacactaatt acaagccacac 1200
	aaaaatcact atcactgttc cccatagaa ccctggaaga caaaaattgtag ggtgttaaac 1260
	tatctataa gccacattgg tgggtgtgaa tgggtggtgtt ctggcataac acccttcttg 1320
	ggtaaacaa cccagagaaata cttggcagaa caacacttgc gagggttttt gagggtgatgata gtagcacaag 1380

ggacgtgttt ggttagcgca agttagatt cagtagctta cccattcoga taagggagcag 1440
gagtgccccc ggaattaccc tgtcaagtcc aatgtggggtt tttgcatag ccaaaacttgtt 1500
gttcaatatt ggaatagggt gtttacacgg tgtgaaaggg gttgtcataa gttgaaggg 1560
	aggaggattt aggctgcaaca ttgcaaaaca tgttcaaggtt gtcttacact catattgtgct 1620
	atatattattt gagattgtctgttttt tattaattt gttttacg aagcnaaac 1680

tatcaatagta cttcataaag attttcattag tttttcttaaaaaaaaa 1740

aaaaaa 1745

<210> SEQ ID NO 7
<211> LENGTH: 534
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 7

Met Ala Lys Leu Leu Asn Asn Thr Ile Phe Cys Ile Leu Phe Ser Ile

Val Phe Leu Ser Phe Ala Lys Leu Leu Ser Ser Tyr Leu Ser Ser Ile Pro

Phe Pro Leu Glu Tyr Ile Ser Leu Ile Val Leu Leu Leu Pro Leu Ile

Ile Asn Phe Leu Cys Val Lys Pro Gln Asn Asn Leu Pro Pro Gly Pro

Thr Ala Val Pro Ile Phe Gly Asn Trp Leu Gln Val Gly Asn Asp Leu

Asn His Gln Leu Leu Ala Thr Met Ser Gln Thr Tyr Gly Pro Ile Phe

Leu Leu Lys Gly Ser Lys Asn Leu Ala Val Val Ser Asn Pro Glu

Leu Ala Asn Gin Val Leu His Thr Gin Gly Val Glu Phe Gly Ser Arg

Pro Arg Asn Val Val Phe Asp Ile Phe Thr Gly Asn Gly Gin Asp Met

Val Phe Thr Ile Tyr Gly Asp His Trp Arg Lys Met Arg Arg Arg Ile Met

Thr Leu Pro Phe Thr Asn Lys Val Val His Gin Tyr Ser Asp Met

Trp Glu Asn Gin Met Asp Leu Val Val Asp Asp Leu Lys Lys Asn Gin

Lys Val Lys Tyr Asp Gly Ile Val Ile Arg Lys Arg Leu Gin Leu Met

Leu Tyr Asn Ile Met Tyr Arg Met Met Phe Asp Ala Lys Phe Glu Ser

Gln Asp Asp Pro Leu Phe Ile Glu Ala Thr Lys Phe Asn Ser Glu Arg

Ser Arg Leu Ala Gin Ser Phe Asp Tyr Asn Tyr Gly Asp Phe Ile Pro

Leu Leu Arg Pro Phe Leu Lys Gly Tyr Leu Asn Lys Cys Lys Asp Leu

Gln Thr Arg Arg Leu Ala Phe Asn Asn Tyr Phe Val Gly Lys Arg

Arg Lys Ile Met Gly Asn Asn Thr Lys Asp Cys Ala

Ile Asp His Ile Ile Asp Ala Glu Met Lys Gly Glu Ile Ser Glu Gin

Asn Val Leu Tyr Ile Val Glu Asn Ile Asn Val Ala Ala Ile Glu Thr

Thr Leu Trp Ser Met Glu Trp Ala Ile Ala Glu Leu Val Asn His Pro

Ile Val Gin Gin Lys Ile Arg Asp Glu Ile Ser Thr Val Leu Lys Gly

Lys Ser Val Lys Glu Ser Asn Leu His Glu Leu Pro Tyr Leu Leu Ala

Thr Val Asn Glu Thr Leu Arg Leu His Thr Pro Ile Pro Leu Leu Val
Pro His Met Asn Leu Glu Glu Ala Lys Leu Gly Gly Tyr Thr Ile Pro
405 410 415
Lys Glu Thr Lys Val Val Val Asn Ala Trp Trp Leu Ala Asn Asn Pro
420 425 430
Ala Trp Trp Lys Asn Glu Asn Glu Phe Arg Pro Glu Arg Phe Leu Glu
435 440 445
Glu Asp Ser Ser Thr Glu Ala Val Ala Gly Lys Val Asp Phe
450 455 460
Arg Tyr Leu Pro Phe Gly Met Gly Arg Arg Ser Cys Pro Gly Ile Ile
465 470 475 480
Leu Ala Leu Pro Ile Leu Gly Val Ile Ala Lys Leu Val Ser Asn
485 490 495
Phe Glu Met Gln Ala Pro Pro Pro Gly Val Gly Lys Val Asp Thr Ser Glu
500 505 510
Lys Gly Gly Gln Phe Ser Leu His Ile Ala Lys His Ser Thr Val Val
515 520 525
Phe Lys Pro Ile Ala Ala
530

<210> SEQ ID NO 8
<211> LENGTH: 1693
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 8
ccctcattca attaaaaaca ttgccacaac aggctttttt gcattttcttt
60
ttcattcattc tttctttctt ttgcacattc actactctcc tacotttctta tacottttcc
120
ttgtggact attcAtccct ttgttcctttt acctccctta ataAtcaacct ttctttgtgt
180
taagcccaaa aacacaccct caccgtgctc aacacgctg ccacattttg gttaattgtt
240
tcaacctg gccaccttg aacctcaccct cttgcccacc atgtcaccaca cctatgtgctc
300
tttcttttca tccacaaatt ggctaaaatc cctagtcttg tttgataact cttgagtc
360
taccacggt ctatcaccgc agggttgctg cttgctgctc cgcctgccag accgtttgtt
420
tgatattat ccgtgtaatt gagacagact cttgtctcatt atttatgttg accattggcc
480
aaaatagtgc cgtagtattgc cgcctctatt ttcacactac aagctgttgcc accatatacg
540
tgatatgtgc gagaagctga tggagcttgtt ttgataact cttgagaaga atgagaaagt
600
gaaatagac ggaatgtgct ttagaaaaac atgtgcacgt atggatatac accatattata
660
tcgaatgtgc ttgtgagcact ctggtgtgac cttgtctca tggaggccac
720
aagtttatat tcaagagaca gcagattcgc tcagagcttt gatcactaat tgggtattt
780
ttacctccttg cttaggcacat cttgggaagg ggctttatat aaccaccaac agtttattat
840
aaggggcct gcattctttc cattttcttt cttgatgtga cttgtatcag cttgatgatg
900
aatggagac aacacccaa ataatttgtgc atattgtaacct attatatgat ctgaaaatgaa
960
aggagagata aagagagcag acatggtctg gattcaccat caggtgccag cttgattattt
1020
tgaaacact ctattgcctca tggagctggc cttgatgctg cttgataacat atctcaattt
1080
tcaaaggag aatggtgcct caaatctcaac ggtgcaccat ctaagaacg
1140
aagaactacat gatgctgctc attgctgctg acaccataaat gaccatataa gacattcaccac
1200
accacactac tataattgaa ccaatgtggg cttgaaaga atggtgatgc gtggtgatac
1260
tattcactaa aaaaactgggt tgggttggtg tgggtgattc aaccctgctg
1320
gtggaaaaac ccagaacagat tcgggcgcga ggcttttcct cgaggagatga gtagccacaga 1380
gggcagctgt ggtggctgca aggttgatgt caggtacttg ccttctggta tggggagggc 1440
aggctgcgcc ggaatctact ttcgaactgc aatgttgggg cttgcatag ccaaaactgtg 1500
gtcaatatatt gaaatgcagg ctctcaggg tggaggaaa gttgatacaaa gttgagaagg 1560
aggcgagttc aggcgtcgca ttcgaaacca ttccaqagtt gtttccaagc ctatgtgcgc 1620
ataatattac agtttttgtt actctataaa gatttcaagt aatattacag tttttgttaa 1680
aaaaaaaaaa aaaa 1693

<210> SEQ ID NO 9
<211> LENGTH: 519
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 9
Met Tyr His Leu Leu Ser Pro Ile Glu Ala Ile Val Gly Leu Val Thr Val Leu
1    5    10    15
Phe Ala Phe Leu Leu Tyr Leu Leu Trp Thr Lys Lys Gin Ser Lys Ile
20   25   30
Leu Asn Pro Leu Pro Pro Lys Ile Pro Gly Gly Trp Pro Val Ile Gly
35   40   45
His Leu Phe Tyr Phe Asn Asn Gly Asp Asp Arg Arg His Phe Ser
50   55   60
Gln Lys Leu Gly Asp Leu Ala Asp Tyr Gly Pro Val Phe Thr Phe
65   70   75   80
Arg Leu Gly Phe Arg Arg Phe Ala Val Ser Ser Tyr Glu Ala Met
85   90   95
Lys Glu Cys Phe Ser Thr Asn Asp Ile His Phe Ala Asp Arg Pro Ala
100  105  110
Leu Leu Tyr Gly Glu Tyr Leu Cys Tyr Asn Asn Ala Met Leu Ala Val
115  120  125
Ala Lys Tyr Gly Pro Tyr Trp Lys Asn Arg Lys Leu Val Asn Gin
130  135  140
Glu Leu Leu Ser Val Ser Arg Leu Glu Lys Phe His Val Arg Phe
145  150  155  160
Ser Ile Val Gin Lys Asn Ile Lys Gin Leu Tyr Asn Cys Asp Ser Pro
165  170  175
Met Val Lys Ile Asn Leu Ser Asp Trp Ile Asp Lys Leu Thr Phe Asp
180  185  190
Ile Ile Leu Lys Met Val Val Gly Lys Thr Tyr Asn Asn Gly His Gly
195  200  205
Glu Ile Leu Lys Ala Ala Phe Gin Lys Phe Met Val Gin Ala Met Glu
210  215  220
Ile Glu Leu Tyr Asp Val Phe His Ile Pro Phe Phe Lys Trp Leu Asp
225  230  235  240
Leu Thr Gly Asn Ile Lys Ala Met Lys Gin Thr Phe Lys Asp Ile Asp
245  250  255
Asn Ile Ile Gin Gly Trp Leu Asp Glu His Ile Lys Lys Arg Glu Thr
260  265  270
Lys Asp Val Gly Gly Glu Asn Glu Gin Asp Phe Ile Asp Val Leu
275  280  285
Ser Lys Arg Ser Asn Glu His Leu Gly Asp Gly Tyr Ser His Asp Thr
290  295  300
Thr Ile Lys Ala Thr Val Phe Thr Leu Val Leu Asp Ala Thr Asp Thr
305  310  315  320
Leu Ala Leu His Ile Lys Trp Val Met Ala Leu Met Ile Asn Asn Lys
325  330  335
Asn Val Met lys Lys Ala Gln Glu Glu Met Asp Thr Ile Val Gly Arg
340  345  350
Amp Arg Trp Val Glu Glu Asn Asp Ile Lys Asn Leu Val Tyr Leu Gln
355  360  365
Ala Ile Val Lys Glu Val Leu Arg Leu His Pro Pro Ala Pro Leu Ser
370  375  380
Val Gln His Leu Ser Val Lys Asp Cys Val Val Asn Gln Tyr His Ile
385  390  395  400
Pro Lys Gly Thr Ala Leu Leu Thr Ann Ile Met Lys Leu Gln Arg Asp
405  410  415
Pro Gln Ile Trp Val Asp Pro Asp Thr Phe Asp Pro Glu Arg Phe Leu
420  425  430
Thr Thr Asn Ala Ala Ile Asp Tyr Arg Gly Gin His Tyr Glu Leu Ile
435  440  445
Pro Phe Gly Ser Gly Arg Arg Ala Cys Pro Ala Met Asn Tyr Ser Leu
450  455  460
Gln Val Glu His Leu Ser Ile Ala His Leu Ile Gin Gly Phe Asn Phe
465  470  475  480
Ala Thr Thr Thr Asn Glu Pro Leu Asp Met Lys Gin Gln Val Gly Leu
485  490  495
Thr Leu Pro Lys Gin Thr Thr Asp Val Glu Val Leu Ile Thr Pro Arg Leu
500  505  510
Pro Pro Thr Leu Tyr Gln Tyr
515

<210> SEQ ID NO 10
<211> LENGTH: 1578
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<400> SEQUENCE: 10
atgtatcatc ttctttctcc catagaacgc atgttaggac tgtaaccttt tgcatatc	60
ctctaccttc atgagaaaa aaaaactca aaatctttta aaccactgcc tcacaaatc	120
ccaggttagc ggcagcatat cggcctcttc ttttattcca acaacaattg cgatgatga c	180
ggccttttt tctaaaaact cggagacctt atgacaaaat atggttcggtt cttaaccttc	240
ggttaggt ttgccggttt tttgcccgtg agtacttatt agagctatgaa aagaagctttc	300
tctaccaatt ataccattt cggcagatg cccagttcact tttacgggaa atacctttgc	360
ttaaaactg ccagtttctc tgggctaaaaa atggtggcctc actgtagaaaa aaactgaaag	420
tagtctcact aagaactctt ctcggattg cggctgaaa aatccaaaca tggtagatctt	480
tctatagtc agagaaaaat taaacacttg atagagtttg actctcaaat ggttgagata	540
aaccctattg atggactgaa taaggagatc ttcgctacta ttttgagaaa gttgtgtgagg	600
aagacacta ataaagagaa ctttggagata cttcttctct ctttgctgaa gttctagattt	660
cagagactg ctaggtgct ttagatgtt tttcgctactt ctttttcctc cgtgtggatgt	720
cctacgagga atatcaaggg tattgaaaca actttctaaag acattgataa tattatccaa	780
ggtgtagtg atgacccatc tsaagagaga gaaacaagag atgttgagag tggaaatgaa	840
caagatttta ttgatgtgct gcccttcagg aggagaacag aacatcgttgg cgatgtgtac
900
ttccatgaca ccaccatcaca agcacaagta ttcatcttttg tcttggtatgc aacagacaaca
960
ttggcaacct atatcaagtg ggtatggtcct ttaatgataa acaataagaa tggatcgaag
1020
aagagcaaca aagagctgga caccatgttt ggtagagata gatggtgtaga agagatgat
1080
atcagaaagt ttggtagcatt tcaagcgaatt gttgaaagaag tatgacagatt acctaacocct
1140
gccatcgttgt cagttacaca cctatcggta aagattgttg tgtctcaaaggt atacatatatt
1200
cctacgccga ctgcaactct tacaaatatt atgaacaccc aacagagacc acaaatattg
1260
gtagaactcg atacactgga tccagaaaga ttcttgacga ctatgtctgc aattgacat
1320
cgagggcgac actagaggctg atcgcggatt ggataagggga ggacagcttg tcccgctatg
1380
aataccta ctgcaagtga acaacattca attgtcatt tgtcagcaggg tttcatttt
1440
gcaactaga ctaacgagcct ttggtatag aacaagaggcg tgggtctaaac tttactaag
1500
aagacagagtt gtagaattgt aattacacct acgcttcctc ctacgccctta tcaatatatt
1560
tatgtttttc tgctgttgaa
1578

<210> SEQ ID NO 11
<211> LENGTH: 509
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 11
Met Glu Gly Thr Arg Leu Thr Thr Tyr Ala Ala Val Phe Leu Gly Thr
1      5      10      15
Leu Phe Leu Leu Leu Leu Ser Lys Phe Leu Arg Glu Arg Lys Leu Arg
20     25     30
Leu Pro Pro Gly Pro Lys Pro Trp Pro Ile Ile Gly Asn Leu Asn Leu
35     40     45
Ile Gly Ala Leu Pro His Arg Ser Ile His Glu Leu Ser Lys Tyr
50     55     60
Gly Pro Ile Met Glu Leu Glu Phe Gly Thr Phe Pro Val Val Val Gly
65     70     75     80
Ser Ser Val Glu Met Ala Lys Val Phe Leu Lys Ser Met Asp Ile Asn
85     90     95
Phe Val Gly Arg Pro Lys Thr Ala Ala Gly Lys Tyr Thr Thr Tyr Asn
100    105    110
Tyr Ser Asp Ile Thr Trp Ser Pro Tyr Gly Pro Tyr Trp Arg Glu Ala
115    120    125
Arg Arg Met Cys Leu Met Glu Leu Phe Ser Thr Lys Arg Leu Asp Ser
130    135    140
Tyr Glu Tyr Ile Arg Ala Glu Leu His Ser Leu Leu His Asn Leu
145    150    155    160
Asn Lys Ile Ser Gly Lys Pro Ile Val Leu Lys Asp Tyr Leu Thr Thr
165    170    175
Leu Ser Leu Arg Val Ile Ser Arg Met Val Leu Gly Lys Arg Tyr Leu
180    185    190
Asp Glu Ser Glu Arg Ser Ile Val Thr Pro Glu Glu Phe Lys Lys Met
195    200    205
Leu Asp Glu Leu Phe Leu Leu Asn Gly Val Leu Asn Ile Gly Asp Ser
210    215    220
Ile Pro Trp Ile Arg Phe Met Asp Leu Glu Gly Tyr Val Lys Arg Met
225    230    235    240
<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>245</td>
<td>250</td>
<td>Lys Phe Val Ser Lys Lys Phe Asp Lys Phe Leu Glu His Val Ile Asp</td>
</tr>
<tr>
<td>255</td>
<td></td>
<td></td>
</tr>
<tr>
<td>260</td>
<td>265</td>
<td>Glu His Asn Val Arg Arg Arg Gly Val Glu Asn Tyr Ile Ala Lys Asp</td>
</tr>
<tr>
<td>270</td>
<td></td>
<td></td>
</tr>
<tr>
<td>275</td>
<td>280</td>
<td>Met Val Asp Val Leu Leu Gin Leu Ala Asp Asp Pro Thr Leu Glu Val</td>
</tr>
<tr>
<td>285</td>
<td></td>
<td></td>
</tr>
<tr>
<td>290</td>
<td>295</td>
<td>Lys Leu Glu Arg His Gly Val Lys Ala Phe Thr Gin Asp Met Leu Ala</td>
</tr>
<tr>
<td>300</td>
<td>310</td>
<td>Gly Gly Thr Glu Ser Ser Ala Val Thr Val Glu Trp Ala Ile Ser Glu</td>
</tr>
<tr>
<td>320</td>
<td></td>
<td></td>
</tr>
<tr>
<td>325</td>
<td>330</td>
<td>Leu Leu Lys Lys Pro Glu Ile Phe Lys Lys Ala Thr Glu Glu Lys Leu Asp</td>
</tr>
<tr>
<td>335</td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>345</td>
<td>Arg Val Ile Gly Gin Asn Arg Trp Val Gin Glu Lys Asp Ile Pro Ann</td>
</tr>
<tr>
<td>350</td>
<td></td>
<td></td>
</tr>
<tr>
<td>355</td>
<td>360</td>
<td>Leu Pro Tyr Ile Glu Ala Ile Val Lys Glu Thr Met Arg Leu His Pro</td>
</tr>
<tr>
<td>365</td>
<td></td>
<td></td>
</tr>
<tr>
<td>370</td>
<td>375</td>
<td>Val Ala Pro Met Leu Val Pro Arg Glu Cys Arg Glu Asp Cys Lys Val</td>
</tr>
<tr>
<td>380</td>
<td></td>
<td></td>
</tr>
<tr>
<td>385</td>
<td>390</td>
<td>Ala Gly Tyr Asp Val Lys Gly Thr Arg Val Leu Val Ser Val Trp</td>
</tr>
<tr>
<td>395</td>
<td>400</td>
<td>Thr Ile Gly Arg Asp Pro Thr Leu Trp Asp Glu Pro Glu Ala Phe Lys</td>
</tr>
<tr>
<td>405</td>
<td>410</td>
<td>Pro Glu Arg Phe His Glu Lys Ser Ile Asp Val Lys Gly His Asp Phe</td>
</tr>
<tr>
<td>415</td>
<td></td>
<td></td>
</tr>
<tr>
<td>420</td>
<td>425</td>
<td>Glu Leu Leu Pro Phe Gly Ala Gly Arg Arg Met Cys Pro Gly Tyr Asn</td>
</tr>
<tr>
<td>430</td>
<td></td>
<td></td>
</tr>
<tr>
<td>435</td>
<td>440</td>
<td>Leu Gly Leu Lys Val Ile Gin Ala Ser Leu Ala Asn Leu Ile His Gly</td>
</tr>
<tr>
<td>445</td>
<td></td>
<td></td>
</tr>
<tr>
<td>450</td>
<td>455</td>
<td>Leu Gly Leu Lys Val Ile Gin Ala Ser Leu Ala Asn Leu Ile His Gly</td>
</tr>
<tr>
<td>460</td>
<td></td>
<td></td>
</tr>
<tr>
<td>465</td>
<td>470</td>
<td>Phe Asn Trp Ser Leu Pro Asp Ann Met Thr Pro Glu Asp Leu Asp Met</td>
</tr>
<tr>
<td>475</td>
<td>480</td>
<td>Asp Glu Ile Phe Gly Leu Ser Thr Pro Lys Lys Phe Pro Leu Ala Thr</td>
</tr>
<tr>
<td>485</td>
<td>490</td>
<td>Val Ile Glu Pro Arg Leu Ser Pro Lys Leu Tyr Ser Val</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 12
<211> LENGTH: 1530
<212> TYPE: DNA
<213> ORGANISM: Nicotiana tabacum
<400> SEQUENCE: 12

<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>60</td>
<td></td>
<td>atggaagaga caaacatgga tacataagca gcagttatttc tgggtactct gtttcttttg</td>
</tr>
<tr>
<td>120</td>
<td></td>
<td>cttctttcca aatctttcgc cccaaagaaa ctctatctac ctctgagcccc aaaaaacctgg</td>
</tr>
<tr>
<td>180</td>
<td></td>
<td>cgagctcgag gaacatctaa cttcatgcgc acctctctct ctgctcaaat ccacgaactt</td>
</tr>
<tr>
<td>240</td>
<td></td>
<td>gccataaag aagagcacat tattgcaactc caatcgggtta ctcttcccgt tggggtgtggc</td>
</tr>
<tr>
<td>300</td>
<td></td>
<td>tctctctgcg aatggtgccg gtttctttctc aatcactatcct atatgctgtat tcagccagg</td>
</tr>
<tr>
<td>360</td>
<td></td>
<td>cctaaagcc cgcgctggaa gcatacaacta tataattt cagatattct attgctctctct</td>
</tr>
<tr>
<td>420</td>
<td></td>
<td>tatggcact atgctggcc gcacagtaga atgtgtctaatgatgtactct tggagacag</td>
</tr>
<tr>
<td>480</td>
<td></td>
<td>cgcttcgatt catctaggtata ggcgggtgct ggtgggtgtgt gcctcttggct ccataatttg</td>
</tr>
<tr>
<td>540</td>
<td></td>
<td>aataaaaaat caggagaaaa atatgtgtct agagattatt tagacagctt ggttttaaat</td>
</tr>
<tr>
<td>600</td>
<td></td>
<td>gttattagca ggtaggtact ggggaagaggg tattggaagc aatcggagaa ctcgattttg</td>
</tr>
</tbody>
</table>
actcgtgagg aatttaagaa gatgttgac gacgtgtcct tgcataatgg tgtacttaat 660
attgagatt caattcctcg gattgatttc atggacttac aaggttatgt taagaggatg 720
aattttgta gcagaaatgt gacagagttt ttgagacagt ttatctgatg gcataaaggt 780
agggagaatt gcggggagga ttcaatgtct catggagatct gattgtcgtt tttgcagcct 840
gtctgtgatc caagcttgag aagtaagctg gagaacagat gagaagaaag atctacattc 900
gatactcctcg ctgtgacatgc cagagtttca gcaattgacag tggagttgagc aatattcgag 960
cgtctaaaga agcgcagagat tttcaaaaaag gttacagaaag aatggagatg agaaaattg 1020
cagaatagag gggtaaaaga aaaaagacct ccaaatcttc ctctactata gcagatagc 1080
aaagagacta tgccactgca cccgtggaca ccaatgtgag tcgccagggga gttcgcagaa 1140
gactggataag agacgcttta cagcgatttg aagagacaaac cgggtcttgggt gacgtatgg 1200
acatattggaag gacacccacac attgtcggagc gacgtctggag cgttacagccg ggcagatgc 1260
caggtggatcg cattgtgatt taaagcagat gattttgaag tgttggcatt tggagctggg 1320
agaagatttg gcgcgggttt taacggttggg cttgagttga ttgaacagtt atggtaa 1380
cctttacagc gttttactgt gtcattgctg gataaatgta cctctggagca ctctgacatg 1440
gttgagatt ttggtccttc cacaactaaa aagtttccac ttgcattcgt gattgagcga 1500
agcctttcct caaactccttc aatctggtttga 1530

<210> SEQ ID NO 13
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene

<400> SEQUENCE: 13

ggcggagaat ttgtctgtga aatgcatttg ggtag 36

<210> SEQ ID NO 14
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene

<400> SEQUENCE: 14

gtacaatagc gaggttgtaca atg 23

<210> SEQ ID NO 15
<211> LENGTH: 18
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene

<400> SEQUENCE: 15

ggtggtgatg aatagcatg 18

<210> SEQ ID NO 16
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene

<400> SEQUENCE: 16

ttatgcagca ataggtgtga agaca 25
<210> SEQ ID NO 17
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene

<400> SEQUENCE: 17

ggggatcca tgtcaatttt cagcttggtt tcc 33

<210> SEQ ID NO 18
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene

<400> SEQUENCE: 18

ggggaattct tactctcag aagttgata agg 33

<210> SEQ ID NO 19
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene

<400> SEQUENCE: 19

cocggatcca tggatatctc tctttotccc 30

<210> SEQ ID NO 20
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene

<400> SEQUENCE: 20

ggggaattct caatattgat aagcgtagg agg 33

<210> SEQ ID NO 21
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene

<400> SEQUENCE: 21

cocggatcca tgtcaatttt cagcttggtt tcc 33

<210> SEQ ID NO 22
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene

<400> SEQUENCE: 22

ggggaagtct cactctgcaag aagattgata agg 33

<210> SEQ ID NO 23
<211> LENGTH: 75
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene
-continued-

<400> SEQUENCE: 23

gcctatatcg ggcaaatct aatctccaa aatcccggtga aaaaatcctaa gctccacct 60
ggctcacaag cagtc 75

<210> SEQ ID NO 24
<211> LENGTH: 75
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene

<400> SEQUENCE: 24

gggggtcataa tcgtcctctc ctctcataaaa ccttcctccttcgctcattt atcgcgcctt 60
atcgcgccta tacta 75

<210> SEQ ID NO 25
<211> LENGTH: 33
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 gene

<400> SEQUENCE: 25

ggggatctct tatcgcgcctagtagttgaacgc 33

<210> SEQ ID NO 26
<211> LENGTH: 7
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum p450 protein
<220> FEATURE:
<223> OTHER INFORMATION: Xaa = Any Amino Acid

<400> SEQUENCE: 26

Lye Glu Thr Leu Arg Leu Xaa
1 5

<210> SEQ ID NO 27
<211> LENGTH: 13
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum p450 protein
<220> FEATURE:
<223> OTHER INFORMATION: Xaa = Any Amino Acid

<400> SEQUENCE: 27

Pro Phe Gly Xaa Gly Arg Arg Xaa Cys Pro Ala
1 5 10

<210> SEQ ID NO 28
<211> LENGTH: 11
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum p450 protein
<220> FEATURE:
<223> OTHER INFORMATION: Xaa = Any Amino Acid

<400> SEQUENCE: 28

Pro Phe Gly Xaa Gly Arg Arg Xaa Cys Pro Gly
1 5 10
<210> SEQ ID NO 29
<211> LENGTH: 6
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum p450 protein
<220> FEATURE:
<221> NAME/KEY: VARIANT
<222> LOCATION: 2
<223> OTHER INFORMATION: Xaa = Any Amino Acid

<400> SEQUENCE: 29

Phe Xaa Pro Glu Arg Phe
1  5

<210> SEQ ID NO 30
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 protein

<400> SEQUENCE: 30

 Ala Ala Arg Gly Ala Arg Ala Cys Ile Tyr Thr Ile Met Gly Ile Tyr
 1  5  10  15

 Thr Ile Cys Ala
 20

<210> SEQ ID NO 31
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 protein

<400> SEQUENCE: 31

 Ala Ala Arg Gly Ala Arg Ala Cys Ile Tyr Thr Ile Met Gly Ile Tyr
 1  5  10  15

 Thr Ile Thr Ala
 20

<210> SEQ ID NO 32
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 protein

<400> SEQUENCE: 32

 Ala Ala Arg Gly Ala Arg Ala Cys Ile Tyr Thr Ile Met Gly Ile Tyr
 1  5  10  15

 Thr Ile Met Gly
 20

<210> SEQ ID NO 33
<211> LENGTH: 19
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 protein

<400> SEQUENCE: 33

 Thr Thr Tyr Ile Ile Cys Cys Ile Gly Ala Arg Met Gly Ile Thr
 1  5  10  15

Thr Tyr

<210> SEQ ID NO 34
<211> LENGTH: 18
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 protein

<400> SEQUENCE: 34

Arg Ala Ala Ile Cys Lys Tyr Thr Cys Ile Gly Ile Ile Ile Arg
1      5     10    15
Ala Ala

<210> SEQ ID NO 35
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 protein

<400> SEQUENCE: 35

Gly Gly Ile Met Gly Ile Met Gly Ile Ile Ile Thr Gly Tyr Cys
1      5     10    15
Cys Ile Gly Ser
20

<210> SEQ ID NO 36
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from Nicotiana tabacum p450 protein

<400> SEQUENCE: 36

Cys Lys Ile Cys Lys Ile Cys Ile Ile Ile Cys Arg Ala
1      5     10    15
Ala Ile Gly Gly
20

<210> SEQ ID NO 37
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from T7 Bacteriophage Promoter

<400> SEQUENCE: 37

tgtagactagtacctagggg
21

<210> SEQ ID NO 38
<211> LENGTH: 21
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: derived from T3 Bacteriophage Promoter

<400> SEQUENCE: 38

tgtaaaccctcactaaaggg
21

<210> SEQ ID NO 39
<211> LENGTH: 560
<212> TYPE: PRT
<213> ORGANISM: Mentha piperita

<400> SEQUENCE: 39

Met Glu Leu Gln Ile Ser Ser Ala Ile Ile Ile Leu Val Val Thr Tyr
1      5     10    15
Thr Ile Ser Leu Leu Ile Ile Lys Gln Trp Arg Lys Pro Lys Pro Gln
20 25 30
Glu Asn Leu Pro Pro Gly Pro Pro Lys Leu Pro Leu Ile Gly His Leu
35 40 45 50
His Leu Leu Trp Gly Lys Leu Pro Pro Gln His Ala Leu Ala Ser Val Ala
55 60 65 70 75 80 85
Lys Gln Tyr Gly Pro Val Ala His Val Gln Leu Gly Glu Val Phe Ser
90 95 100 105 110 115 120 125 130 135 140
Val Val Leu Ser Ser Arg Glu Ala Thr Lys Glu Ala Met Lys Leu Val
145 150 155 160 165 170 175 180 185 190
Arg Glu Ser Ala Cys Ala Asp Arg Phe Glu Ser Ile Gly Thr Lys Ile Met
195 200 205 210 215 220 225 230 235 240
Trp Tyr Asp Asn Asp Ile Ile Phe Ser Pro Tyr Ser Val His Trp
245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350
Gly His Leu Arg Ser Ser Ala Ala Gly Glu Ala Val Asp Leu Thr
355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430
Val Glu Val Pro Leu Ala Glu Leu Tyr His Phe Asp Trp Lys Leu
435 440 445

Ala Glu Gly Met Asn Pro Ser Asp Met Asp Met Ser Glu Ala Glu Gly
450 455 460

Leu Thr Gly Ile Arg Lys Asn Asn Leu Leu Val Pro Thr Pro Tyr
465 470 475 480

Asp Pro Ser Ser
485 490 495

500

<210> SEQ ID NO 40
<211> LENGTH: 496
<212> TYPE: PRT
<213> ORGANISM: Mentha spicata

<400> SEQUENCE: 40
Met Glu Leu Asp Leu Leu Ser Ala Ile Ile Leu Val Ala Thr Tyr
1  5  10  15
Ile Val Ser Leu Leu Ile Glu Trp Arg Lys Ser Lys Ser Glu Gln
20  25  30
Asn Leu Pro Pro Ser Pro Pro Lys Leu Pro Val Ile Gly His Leu His
35  40  45
Phe Leu Trp Gly Gly Leu Pro Gln His Val Phe Arg Ser Ile Ala Gln
50  55  60
Lys Tyr Gly Pro Val Ala His Val Glu Gly Glu Val Tyr Ser Val
65  70  75  80
Val Leu Ser Ser Ala Glu Ala Ala Lys Glu Ala Met Lys Val Leu Asp
85  90  95
Pro Asn Phe Ala Asp Arg Phe Asp Gly Ile Gly Ser Arg Thr Met Trp
100 105 110
Tyr Asp Lys Asp Asp Ile Ile Phe Ser Pro Tyr Asp His Trp Arg
115 120 125
Gln Met Arg Arg Ile Cys Val Thr Glu Leu Leu Ser Pro Lys Asn Val
130 135 140
Arg Ser Phe Gly Tyr Ile Arg Glu Glu Glu Ile Glu Arg Leu Ile Arg
145 150 155 160
Leu Leu Gly Ser Ser Gly Gly Ala Pro Val Asp Val Thr Glu Glu Val
165 170 175
Ser Lys Met Ser Cys Val Val Val Cys Arg Ala Ala Phe Gly Ser Val
180 185 190
Leu Lys Asp Gln Gly Ser Leu Ala Glu Leu Val Lys Glu Ser Leu Ala
195 200 205
Leu Ala Ser Gly Phe Glu Leu Ala Asp Leu Tyr Pro Ser Ser Trp Leu
210 215 220
Leu Asn Leu Leu Ser Leu Asn Lys Tyr Arg Leu Glu Arg Met Arg Arg
225 230 235 240
Arg Leu Asp His Ile Leu Asp Gly Asp Leu Glu Gly His Arg Glu Lys
245 250 255
Lys Ser Gly Gly Phe Gly Gly Glu Asp Ile Val Asp Val Leu Leu Asp
260 265 270
Met Glu Lys Gly Ser Asp Ile Lys Ile Pro Ile Thr Ser Asn Cys Ile
275 280 285
Lys Gly Phe Ile Phe Asp Thr Phe Ser Ala Gly Ala Glu Thr Ser Ser
290 295 300
Thr Thr Ile Ser Trp Ala Leu Ser Glu Leu Met Arg Asn Pro Ala Lys
305 310 315 320
Met Ala Lys Val Gln Ala Glu Val Arg Glu Ala Leu Lys Gly Lys Thr
325 330 335
Val Val Asp Leu Ser Glu Val Gln Glu Leu Lys Tyr Leu Arg Ser Val
340 345 350
Leu Lys Glu Thr Leu Arg Leu His Pro Pro Phe Pro Leu Ile Pro Arg
355 360 365
Gln Ser Arg Glu Glu Cys Glu Val Asn Gly Tyr Thr Ile Pro Ala Lys
370 375 380
Thr Arg Ile Phe Ile Asn Val Trp Ala Ile Gly Arg Asp Pro Gln Tyr
385 390 395 400
Trp Glu Asp Pro Arg Thr Phe Arg Pro Glu Arg Phe Asp Glu Val Ser
405 410 415
Arg Asp Phe Met Gln Asn Asp Phe Glu Ile Pro Phe Gly Ala Gly
420 425 430
Arg Arg Ile Cys Pro Gly Leu His Phe Gly Leu Ala Asn Val Glu Ile
435 440 445
Pro Leu Ala Gln Leu Leu Tyr His Phe Asp Trp Lys Leu Pro Gln Gly
450 455 460
Met Thr Asp Ala Asp Leu Asp Met Thr Glu Thr Pro Gly Leu Ser Gly
465 470 475 480
Pro Lys Lys Lys Asn Val Cys Leu Val Pro Thr Leu Tyr Lys Ser Pro
485 490 495

<210> SEQ ID NO 41
<211> LENGTH: 509
<212> TYPE: PRT
<213> ORGANISM: Nepeta racemosa

<400> SEQUENCE: 41
Met Val Ser Leu Ser Tyr Phe Leu Ile Ala Leu Leu Cys Thr Leu Pro
1  5  10  15
Phe Leu Leu Phe Leu Asn Lys Trp Arg Arg Ser Tyr Ser Gly Lys Thr
20  25  30
Pro Pro Pro Ser Pro Pro Leu Leu Pro Val Ile Gly Asn Leu His Gln
35  40
Leu Gly Leu Tyr Pro His Arg Tyr Leu Gin Ser Leu Ser Arg Arg Tyr
50  55  60
Gly Pro Leu Met Gin Leu His Phe Gly Ser Val Pro Leu Val Ala
65  70  75  80
Ser Ser Pro Glu Ala Ala Arg Glu Ile Met Lys Asn Gln Asp Ile Val
85  90  95
Phe Ser Asn Arg Pro Lys Met Ser Ile Ala Asn Arg Leu Phe Phe Asn
100 105 110
Asn Arg Asp Val Ala Phe Thr Gin Tyr Glu Gly Tyr Trp Arg Gin Ile
115 120 125
Arg Ser Ile Cys Val Leu Gln Leu Leu Ser Asn Lys Arg Val Gln Ser
130 135 140
Phe Arg Arg Val Arg Glu Glu Glu Thr Ser Ile Met Val Glu Lys Ile
145 150 155 160
Met Gin Leu Gly Ser Ser Ser Ser Thr Pro Val Asn Leu Ser Glu Leu
165 170 175
Leu Leu Ser Leu Thr Asn Asp Val Val Cys Arg Val Thr Leu Gly Lys
180 185 190
Lys Tyr Gly Gly Gly Asn Gly Ser Glu Glu Val Asp Lys Leu Lys Glu 195 200 205
Met Leu Thr Glu Ile Gln Asn Leu Met Gly Ile Ser Pro Val Trp Glu 210 215 220
Phe Ile Pro Trp Leu Asn Trp Thr Arg Arg Phe Asp Gly Val Asp Gin 225 230 235 240
Arg Val Asp Arg Ile Val Lys Ala Phe Asp Gly Phe Leu Glu Ser Val 245 250 255
Ile Gln Glu His Lys Glu Gly Arg Asp Gly Asp Lys Asp Gly Asp Gly Asp 260 265 270
Gly Ala Leu Asp Phe Val Asp Ile Leu Leu Gin Phe Gin Arg Glu Asn 275 280 285
Lys Asn Arg Ser Pro Val Glu Asp Thr Val Lys Ala Leu Ile Leu 290 295 300
Asp Met Phe Val Ala Gly Thr Asp Thr Thr Ala Thr Ala Leu Glu Trp 305 310 315 320
Ala Val Ala Glu Leu Ile Lys Asn Pro Arg Ala Met Lys Arg Leu Gin 325 330 335
Asn Glu Val Arg Glu Val Ala Gly Ser Lys Ala Glu Ile Glu Glu Glu 340 345 350
Asp Leu Glu Gly Met Pro Tyr Leu Lys Ala Ser Ile Lys Gly Ser Leu 355 360 365
Arg Leu His Val Pro Val Val Leu Leu Val Pro Arg Glu Ser Thr Arg 370 375 380
Asp Thr Asn Val Leu Gly Tyr Asp Ile Ala Ser Gly Thr Arg Val Leu 385 390 395 400
Ile Asn Ala Trp Ala Ile Ala Arg Pro Ser Val Trp Glu Asn Pro 405 410 415
Glu Glu Phe Leu Pro Glu Arg Phe Leu Asp Ser Ser Ile Asp Tyr Lys 420 425 430
Gly Leu His Phe Glu Leu Leu Pro Phe Gly Ala Gly Arg Arg Gly Cys 435 440 445
Pro Gly Ala Thr Phe Ala Val Ala Ile Asp Glu Leu Ala Leu Ala Lys 450 455 460
Leu Val His Lys Phe Asp Phe Gly Leu Pro Asn Gly Ala Arg Met Glu 465 470 475 480
Glu Leu Asp Met Ser Glu Thr Ser Gly Met Thr Val His Lys Ser 495
Pro Leu Leu Leu Pro Ile Pro His His Ala Ala Pro 500
<210> SEQ ID NO: 42
<211> LENGTH: 471
<212> TYPE: PRT
<213> ORGANISM: Persea americana
<400> SEQUENCE: 42
Met Ala Ile Leu Val Ser Leu Leu Phe Leu Ala Ile Ala Leu Thr Phe 1 5 10 15
Phe Leu Leu Lys Leu Asn Glu Lys Arg Glu Lys Lys Pro Asn Leu Pro 20 25 30
Pro Ser Pro Pro Leu Leu Pro Ile Ile Gly Asn Leu His Gin Leu Gly 35 40 45
Asn Leu Pro His Arg Ser Leu Arg Ser Leu Ala Asn Glu Leu Gly Pro
<table>
<thead>
<tr>
<th>50</th>
<th>55</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu Ile Leu Leu His Leu Gly His Ile Pro Thr Leu Ile Val Ser Thr</td>
<td>70</td>
<td>75</td>
</tr>
<tr>
<td>Ala Glu Ile Ala Glu Ile Leu Lys Thr His Arg Leu Ile Phe Ala</td>
<td>95</td>
<td>90</td>
</tr>
<tr>
<td>Ser Arg Pro Ser Thr Thr Ala Ala Arg Arg Ile Phe Tyr Asp Cys Thr</td>
<td>100</td>
<td>105</td>
</tr>
<tr>
<td>Asp Val Ala Phe Ser Pro Tyr Gly Glu Tyr Trp Arg Gin Val Arg Lys</td>
<td>115</td>
<td>120</td>
</tr>
<tr>
<td>Ile Cys Val Leu Glu Leu Leu Ser Ile Lys Arg Val Asn Ser Tyr Arg</td>
<td>130</td>
<td>135</td>
</tr>
<tr>
<td>Ser Ile Arg Glu Glu Val Gly Leu Met Met Gin Arg Ile Ser Gin</td>
<td>145</td>
<td>150</td>
</tr>
<tr>
<td>Ser Cys Ser Thr Gly Glu Ala Val Asn Leu Ser Glu Leu Leu Leu Leu</td>
<td>165</td>
<td>170</td>
</tr>
<tr>
<td>Leu Ser Ser Gly Thr Ile Thr Arg Val Ala Phe Gly Lys Tyr Gly</td>
<td>180</td>
<td>185</td>
</tr>
<tr>
<td>Gly Glu Glu Arg Lys Asn Lys Phe Ala Asp Leu Ala Thr Glu Leu</td>
<td>195</td>
<td>200</td>
</tr>
<tr>
<td>Thr Thr Leu Met Gly Ala Phe Phe Val Gly Asp Tyr Phe Pro Ser Phe</td>
<td>210</td>
<td>215</td>
</tr>
<tr>
<td>Ala Trp Val Asp Val Leu Thr Gly Met Asp Ala Arg Leu Lys Arg Asn</td>
<td>225</td>
<td>230</td>
</tr>
<tr>
<td>His Gly Glu Leu Asp Ala Phe Val Asp His Val Ile Asp Asp His Leu</td>
<td>245</td>
<td>250</td>
</tr>
<tr>
<td>Leu Ser Arg Lys Ala Asn Gly Ser Asp Gly Val Glu Gin Lys Asp Leu</td>
<td>260</td>
<td>265</td>
</tr>
<tr>
<td>Val Asp Val Leu Leu His Leu Gin Lys Asp Ser Ser Leu Gly Val His</td>
<td>275</td>
<td>280</td>
</tr>
<tr>
<td>Leu Asn Arg Asn Asn Leu Lys Ala Val Ile Leu Asp Met Phe Ser Gly</td>
<td>290</td>
<td>295</td>
</tr>
<tr>
<td>Gly Thr Asp Thr Thr Ala Val Thr Leu Glu Trp Ala Met Ala Glu Leu</td>
<td>305</td>
<td>310</td>
</tr>
<tr>
<td>Ile Lys His Pro Asp Val Met Glu Lys Ala Gin Gin Glu Val Arg Arg</td>
<td>325</td>
<td>330</td>
</tr>
<tr>
<td>Val Val Gly Lys Lys Ala Lys Val Glu Glu Asp Leu His Gin Leu</td>
<td>340</td>
<td>345</td>
</tr>
<tr>
<td>His Tyr Leu Lys Leu Ile Lys Gly Thr Leu Arg Leu Arg His Pro Val</td>
<td>355</td>
<td>360</td>
</tr>
<tr>
<td>Ala Pro Leu Leu Val Arg Glu Ser Thr Arg Asp Val Val Ile Arg</td>
<td>370</td>
<td>375</td>
</tr>
<tr>
<td>Gly Tyr His Ile Pro Ala Lys Thr Arg Val Phe Asn Ala Trp Ala</td>
<td>385</td>
<td>390</td>
</tr>
<tr>
<td>Ile Gly Arg Asp Pro Lys Ser Trp Glu Asn Ala Glu Glu Phe Leu Pro</td>
<td>405</td>
<td>410</td>
</tr>
<tr>
<td>Glu Arg Phe Val Asn Asn Ser Val Asp Phe Lys Gin Gin Gin Asp Phe Gin</td>
<td>420</td>
<td>425</td>
</tr>
<tr>
<td>Leu Ile Pro Phe Gly Ala Gly Arg Gly Cys Pro Gly Ile Ala Phe</td>
<td>435</td>
<td>440</td>
</tr>
<tr>
<td>Gly Ile Ser Ser Val Gin Ile Ser Leu Ala Gin Leu Leu Tyr Thr Phe</td>
<td>450</td>
<td>455</td>
</tr>
<tr>
<td>Asn Trp Glu Leu Pro Gly Ile</td>
<td>465</td>
<td>470</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Met</td>
<td>Ala</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Phe</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Asn</td>
<td>Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gly</td>
<td>Phe</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>His</td>
<td>Lys</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Lys</td>
<td>Met</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Lys</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Ser</td>
<td>Asn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Ser</td>
<td>Asp</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Gly</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Ala</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>His</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Phe</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Tyr</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Asp</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Arg</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Ile</td>
<td>Gln</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cys</td>
<td>Tyr</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gln</td>
<td>Ile</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr</td>
<td>Leu</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Val</td>
<td>Gln</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Lys</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Phe</td>
<td>His</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We claim:

1. A method for producing an isoprenoid compound, said method comprising the steps of:
culturing a host cell under conditions suitable for expressing a first recombinant protein comprising an isoprenoid synthase and a second recombinant protein comprising a CYP71 family cytochrome P450 polypeptide, wherein said first and second recombinant protein catalyze the formation of an isoprenoid compound not normally produced by said host cell; and
recovering said isoprenoid compound, wherein said cytochrome P450 polypeptide is selected from the group consisting of a polypeptide comprising at least 95% identity to the amino acid sequence of SEQ ID NO:1, and a polypeptide comprising at least 95% identity to the amino acid sequence of SEQ ID NO:3.

2. A method for producing an altered isoprenoid compound, said method comprising the step of contacting an isoprenoid compound with an isolated CYP71 family cytochrome P450 polypeptide, wherein said polypeptide catalyzes the formation of an isoprenoid compound, under conditions allowing for the hydroxylation, oxidation, demethylation, methylation, or any combination thereof of said compound and recovering said altered isoprenoid compound, wherein the isolated CYP71 family cytochrome P450 polypeptide comprises an amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO: 1 or of SEQ ID NO: 3.

3. A method for producing an altered isoprenoid compound, said method comprising the step of contacting an isoprenoid compound with an isolated CYP71 family cytochrome P450 polypeptide, wherein said polypeptide catalyzes the formation of an isoprenoid compound, under conditions allowing for the hydroxylation, oxidation, demethylation, methylation, or any combination thereof of said compound and recovering said altered isoprenoid compound, wherein the isolated CYP71 family cytochrome P450 polypeptide comprises an amino acid sequence that is encoded by a nucleic acid sequence at least 95% identical to the nucleic acid sequence of SEQ ID NO: 1 or of SEQ ID NO: 2 or of SEQ ID NO: 3.

4. A host cell expressing a recombinant isoprenoid synthase and a recombinant CYP71 family cytochrome P450 polypeptide, wherein said second recombinant protein has hydroxylase activity, and further wherein said first and second recombinant protein catalyze the formation of an isoprenoid compound not normally produced by said host cell, wherein the recombinant CYP71 family cytochrome P450 polypeptide comprises an amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO: 1 or of SEQ ID NO: 3.

5. A host cell expressing a recombinant isoprenoid synthase and a recombinant CYP71 family cytochrome P450 polypeptide, wherein said second recombinant protein has hydroxylase activity, and further wherein said first and second recombinant protein catalyze the formation of an isoprenoid compound not normally produced by said host cell, wherein the recombinant CYP71 family cytochrome P450 polypeptide comprises an amino acid sequence that is encoded by a nucleic acid sequence at least 95% identical to the nucleic acid sequence of SEQ ID NO: 2 or of SEQ ID NO: 3.

6. A method for producing an isoprenoid compound, said method comprising the steps of:
culturing a host cell under conditions suitable for expressing a first recombinant protein comprising an isoprenoid synthase and a second recombinant protein comprising a CYP71 family cytochrome P450 polypeptide, wherein said second recombinant protein has hydroxylase activity, and further wherein said first and second recombinant protein catalyze the formation of an isoprenoid compound not normally produced by said host cell; and
recovering said isoprenoid compound, wherein the cytochrome P450 comprises an amino acid sequence that is encoded by a nucleic acid sequence at least 95% identical to the nucleic acid sequence of SEQ ID NO: 2 or of SEQ ID NO: 3.