



12-17-2002

Synthases

Joseph Chappell
University of Kentucky, chappell@uky.edu

Kathleen R. Manna

Joseph P. Noel

Courtney M. Starks

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

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

 Part of the [Pharmacy and Pharmaceutical Sciences Commons](#)

Recommended Citation

Chappell, Joseph; Manna, Kathleen R.; Noel, Joseph P.; and Starks, Courtney M., "Synthases" (2002). *Pharmaceutical Sciences Faculty Patents*. 67.
https://uknowledge.uky.edu/ps_patents/67

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.



US006495354B2

(12) **United States Patent**
Chappell et al.

(10) **Patent No.:** **US 6,495,354 B2**
(45) **Date of Patent:** **Dec. 17, 2002**

(54) **SYNTHESES**

(75) Inventors: **Joseph Chappell**, Lexington, KY (US); **Kathleen R. Manna**, Georgetown, IN (US); **Joseph P. Noel**, San Diego, CA (US); **Courtney M. Starks**, La Jolla, CA (US)

(73) Assignees: **University of Kentucky Research Foundation**, Lexington, KY (US); **The Salk Institute for Biological Studies**, San Diego, CA (US)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 0 days.

(21) Appl. No.: **09/887,586**

(22) Filed: **Jun. 22, 2001**

(65) **Prior Publication Data**

US 2002/0094556 A1 Jul. 18, 2002

Related U.S. Application Data

(62) Division of application No. 09/398,395, filed on Sep. 17, 1999.

(60) Provisional application No. 60/150,262, filed on Aug. 23, 1999, provisional application No. 60/130,628, filed on Apr. 22, 1999, and provisional application No. 60/100,993, filed on Sep. 18, 1998.

(51) **Int. Cl.**⁷ **C12N 9/00**; C12N 9/88; C07H 21/04

(52) **U.S. Cl.** **435/183**; 435/4; 435/232; 435/468; 800/283; 800/284; 800/285; 800/289; 536/23.2

(58) **Field of Search** 435/183, 4, 232, 435/468; 800/283, 284, 285, 289; 536/23.2

(56) **References Cited**

U.S. PATENT DOCUMENTS

5,589,619	A	12/1996	Chappell et al.
5,824,774	A	10/1998	Chappell et al.
5,849,526	A	12/1998	Pichersky
5,871,988	A	2/1999	Croteau et al.
5,876,964	A	3/1999	Croteau et al.
5,891,697	A	4/1999	Croteau et al.
5,981,843	A	11/1999	Chappell et al.
5,994,114	A	11/1999	Croteau et al.
6,008,043	A	12/1999	Croteau et al.

FOREIGN PATENT DOCUMENTS

EP	0 768 381	4/1997
WO	WO 95/11913	5/1995
WO	WO 96/36697	11/1996
WO	WO 97/15584	5/1997
WO	WO 97/38571	10/1997
WO	WO 97/38703	10/1997
WO	WO 99/02030	1/1999
WO	WO 99/15624	4/1999
WO	WO 99/18118	4/1999
WO	WO 99/19460	4/1999
WO	WO 99/37139	7/1999
WO	WO 99/38957	8/1999

OTHER PUBLICATIONS

Starks et al., "Structural Basis for Cyclic Terpene Biosynthesis by Tobacco 5-Epi-Aristolochene Synthase," *Science*, 1997, 227:1815-1820.

Chappell, "The Biochemistry and Molecular Biology of Isoprenoid Metabolism," *Plant Physiol.*, 1995, 107:1-6.

Facchini et al., "Gene family for an elicitor-induced sesquiterpene cyclase in tobacco," *Proc. Natl. Acad. Sci. USA*, 1992, 89:11088-11092.

Corey et al., "Isolation of an *Arabidopsis thaliana* gene encoding cycloartenol synthase by functional expression in a yeast mutant lacking lanosterol synthase by the use of a chromatographic screen," *Proc. Natl. Acad. Sci. USA*, 1993, 90:11628-11632.

Bohlmann et al., "Terpenoid-based defenses in conifers: cDNA cloning, characterization, and functional expression of wound-inducible (E)- α -bisabolene synthase from grand fir (*Abies grandis*)," *Proc. Natl. Acad. Sci. USA*, 1998, 95:6756-6761.

Colby et al., "Germacrene C synthase from *Lycopersicon esculentum* cv. VFNT Cherry tomato: cDNA isolation, characterization, and bacterial expression of the multiple product sesquiterpene cyclase," *Proc. Natl. Acad. Sci. USA*, 1998, 95:2216-2221.

Devarenne et al., "Molecular Characterization of Tobacco Squalene Synthase and Regulation in Response to Fungal Elicitor," *Arch. Biochem. Biophys.*, 1998, 349(2):205-215.

Back et al., "Cloning and Bacterial Expression of a Sesquiterpene Cyclase from *Hyoscyamus muticus* and Its Molecular Comparison to Related Terpene Cyclases," *J. Biol. Chem.*, 1995, 270(13):7375-7381.

Yin et al., "Regulation of Sesquiterpene Cyclase Gene Expression—Characterization of an Elicitor—and Pathogen-Inducible Promoter," *Plant Physiol.*, 1997, 115:437-451.

Mathis et al., "Pre-Steady-State Study of Recombinant Sesquiterpene Cyclases," *Biochemistry*, 36(27):8340-8348.

Back et al., "Identifying functional domains within terpene cyclases using a domain-swapping strategy," *Proc. Natl. Acad. Sci. USA*, 1996, 93:6841-6845.

(List continued on next page.)

Primary Examiner—Mary K. Zeman

Assistant Examiner—Lori A. Clow

(74) *Attorney, Agent, or Firm*—Fish & Richardson P.C., P.A.

(57)

ABSTRACT

Novel synthases and the corresponding nucleic acids encoding such synthases are disclosed herein. Such synthases possess an active site pocket that includes key amino acid residues that are modified to generate desired terpenoid reaction intermediates and products. Synthase modifications are designed based on, e.g., the three-dimensional coordinates of tobacco 5-epi-aristolochene synthase with or without a substrate bound in the active site.

86 Claims, 4 Drawing Sheets

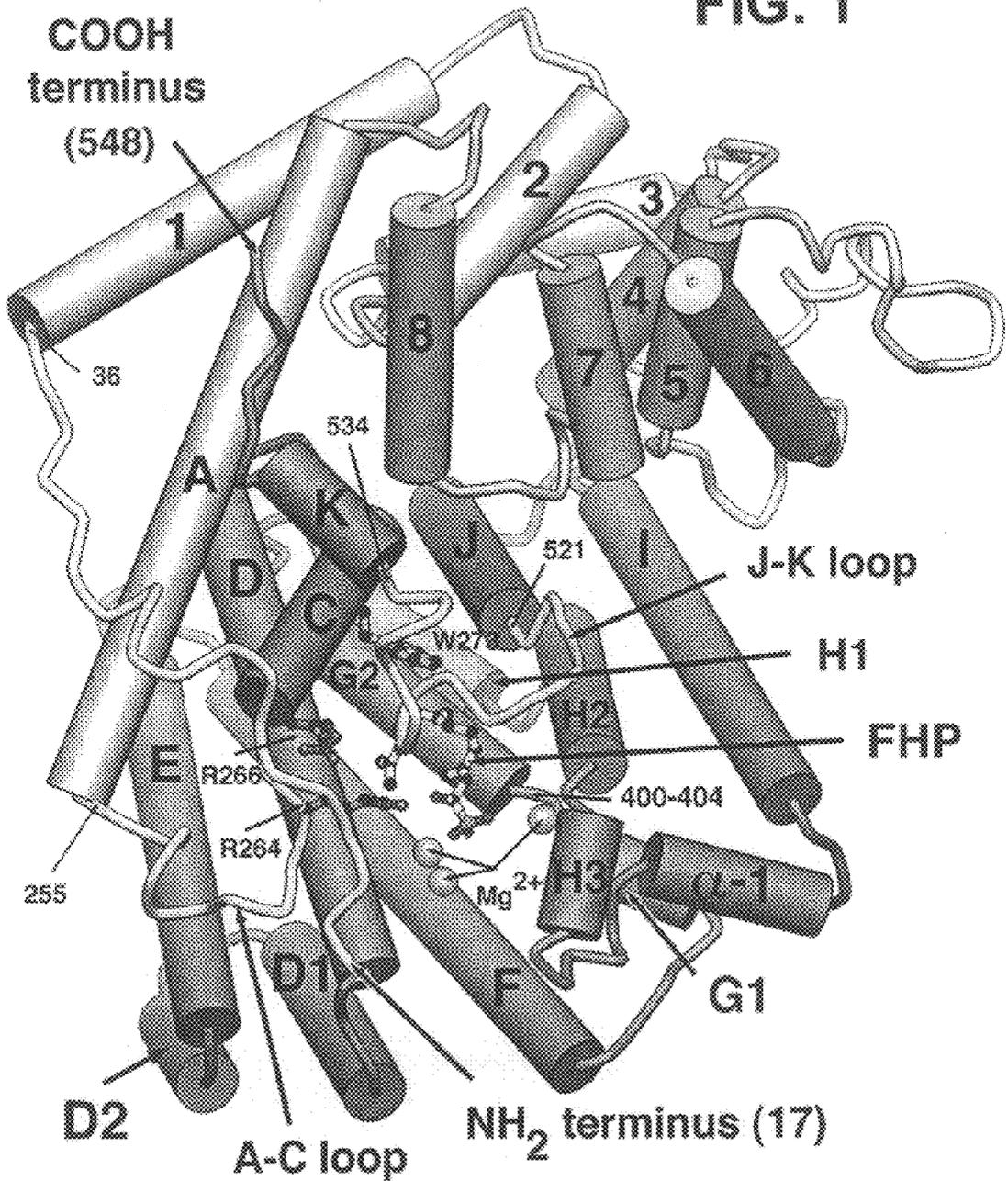
OTHER PUBLICATIONS

- Newman et al., "Characterization of the TAC box, a cis-element within an elicitor-inducible sesquiterpene cyclase promoter," *Plant J.*, 1998, 16(1):1-12.
- Crock et al., "Isolation of bacterial expression of a sesquiterpene synthase cDNA clone from peppermint (*Mentha χ piperita*, L.) that produces the aphid alarm pheromone (E)- β -farnesene," *Proc. Natl. Acad. Sci. USA*, 1997, 94:12833-12838.
- Wildung et al., "A cDNA Clone for Taxadiene Synthase, the Diterpene Cyclase That Catalyzes the Committed Step of Taxol Biosynthesis," *J. Biol. Chem.*, 1996, 271(16):9201-9204.
- Chen et al., "Cloning, Expression, and Characterization of (+)- δ -Cadinene Synthase: A Catalyst for Cotton Phytoalexin Biosynthesis," *Arch. Biochem. Biophys.*, 1995, 324(2):255-266.
- Bohlmann et al., "Plant terpenoid synthases: Molecular biology and phylogenetic analysis," *Proc. Natl. Acad. Sci. USA*, 1998, 95:4126-4133.
- Cane et al., "Trichodiene Synthase. Substrate Specificity and Inhibition," *Biochemistry*, 1995, 34:2471-2479.
- Pyun et al., "Regiospecificity and Isotope Effects Associated with the Methyl-Methylene Eliminations in the Enzyme-Catalyzed Biosynthesis of (R)- and (S)-Limonene," *J. Org. Chem.*, 1993, 58(15):3998-4009.
- Anderson et al., "Farnesyl Diphosphate Synthetase—Molecular Cloning, Sequence, and Expression of an Essential Gene From *Saccharomyces cerevisiae*," *J. Biol. Chem.*, 1989, 264(32):19176-19184.
- Song et al., "Yeast farnesyl-diphosphate synthase: Site-directed mutagenesis of residues in highly conserved prenyltransferase domains I and II," *Proc. Natl. Acad. Sci. USA*, 1994, 91:3044-3048.
- Ohnuma et al., "A Role of the Amino Acid Residue Located on the Fifth Position before the First Aspartate-rich Motif of Farnesyl Diphosphate Synthase on Determination of the Final Product," *J. Biol. Chem.*, 1996, 271(48):30748-30754.
- Tarshis et al., "Regulation of product chain length by isoprenyl diphosphate synthases," *Proc. Natl. Acad. Sci. USA*, 1996, 93:15018-15023.
- Tarshis et al., "Crystal Structure of Recombinant Farnesyl Diphosphate Synthase at 2.6-Å Resolution," *Biochemistry*, 1994, 33:10871-10877.
- Mau et al., "Cloning of casbene synthase cDNA: Evidence for conserved structural features among terpenoid cyclases in plants," *Proc. Natl. Acad. Sci. USA*, 1994, 91:8497-8501.
- Wendt et al., "Structure and Function of a Squalene Cyclase," *Science*, 1997, 277:1811-1815.
- Cane et al., "Trichodiene Synthase. Identification of Active Site Residues by Site-Directed Mutagenesis," *Biochemistry*, 1995, 34:2480-2488.
- Cane et al., "Trichodiene Biosynthesis and the Stereochemistry of the Enzymatic Cyclization of Farnesyl Pyrophosphate," *Bioorg. Chem.*, 1985, 13(3):246-265.
- Wheeler et al., "Direct demonstration of the isomerization component of the monoterpene cyclase reaction using a cyclopropylcarbinyl pyrophosphate substrate analog," *Proc. Natl. Acad. Sci. USA*, 1987, 84(14):4856-4859.
- Pyun et al., "Stereochemistry of the Proton Elimination in the Formation of (+)- and (-)- α -Pinene by Monoterpene Cyclases from Sage (*Salvia Officinalis*)," *Arch. Biochem. Biophys.*, 1994, 308(2):488-496.
- Croteau, "Evidence for the Ionization Steps in Monoterpene Cyclization Reactions Using 2-Fluorogeranyl and 2-Fluorolinalyl Pyrophosphates as Substrates," *Arch. Biochem. Biophys.*, 1986, 251(2):777-782.
- Croteau et al., "Irreversible Inactivation of Monoterpene Cyclases by a Mechanism-Based Inhibitor," *Arch. Biochem. Biophys.*, 1993, 307(2):397-404.
- Rajaonarivony et al., "Characterization and Mechanism of (4S)-Limonene Synthase, A Monoterpene Cyclase from the Glandular Trichomes of Peppermint (*Mentha X piperita*)," *Arch. Biochem. Biophys.*, 1992, 296(1):49-57.
- Rajaonarivony et al., "Evidence for an Essential Histidine Residue in 4S-Limonene Synthase and Other Terpene Cyclases," *Arch. Biochem. Biophys.*, 1992, 299(1):77-82.
- Aleshin et al., "Refined Crystal Structures of Glucoamylase from *Aspergillus awamori* var. X100," *J. Mol. Biol.*, 1994, 238:575-591.
- Juy et al., "Three-dimensional structure of a thermostable bacterial cellulase," *Nature*, 1992, 357(6373):89-91.
- Altschul et al., "Gapped Blast and PSI-Blast: a new generation of protein database search programs," *Nucleic Acids Res.*, 1997, 25(17):3389-3402.
- Back, et al., "Expression of a Plant Sesquiterpene Cyclase Gene in *Escherichia coli*," *Arch. Biochem. Biophys.*, 1994, 315(2):527-532.
- Cane, "Enzymatic Formation of Sesquiterpenes," *Chem. Rev.*, 1990, 90:1089-1103.
- Cane et al., "Aristolochene Biosynthesis and Enzymatic Cyclization of Farnesyl Pyrophosphate," *J. Am. Chem. Soc.*, 1989, 111:8914-8916.
- Cane et al., "Overexpression in *Escherichia coli* of Soluble Aristolochene Synthase from *Penicillium roqueforti*," *Arch. Biochem. Biophys.*, 1993, 304(2):415-419.
- Hohn et al., "Purification and Characterization of the Sesquiterpene Cyclase Aristolochene Synthase from *Penicillium roqueforti*," *Arch. Biochem. Biophys.*, 1989, 272(1):137-143.
- Laskovics et al., "Prenyltransferase: Determination of the Binding Mechanism and Individual Kinetic Constants for Farnesylpyrophosphate Synthetase by Rapid Quench and Isotope Partitioning Experiments," *Biochemistry*, 1981, 20(7):1893-1901.
- Lesburg et al., "Crystal Structure of Pentalenene Synthase: Mechanistic Insights on Terpenoid Cyclization Reactions in Biology," *Science*, 1997, 277(5333):1820-1824.
- Munck et al., "Purification and Characterization of the Sesquiterpene Cyclase Pathchoulol Synthase from *Pogostemon cablin*," *Arch. Biochem. Biophys.*, 1990, 282(1):58-64.
- Proctor et al., "Aristolochene Synthase. Isolation, characterization, and bacterial expression of a sesquiterpenoid biosynthetic gene (Ari1) from *Penicillium roqueforti*," *J. Biol. Chem.*, 1993, 268(6):4543-4548.
- Vogel et al., "Abietadiene Synthase from Grand Fir (*Abies grandis*)," *J. Biol. Chem.*, 1996, 271(38):23262-23268.
- Bohlmann et al., "Monoterpene Synthases from Grand Fir (*Abies grandis*)," *J. Biol. Chem.*, 1997, 272(35):21784-21792.
- Starks et al., "Structural Basis for Cyclic Terpene Biosynthesis by Tobacco 5-Epi-Aristolochene Synthase," *Science*, 1997, 277:1815-1819.
- Genbank Accession No: Q40577.
- Genbank Accession No: AB022598.
- Genbank Accession No: Y18484.
- Genbank Accession No: U48796.

Genbank Accession No: AF035631.
Genbank Accession No: L13459.
Genbank Accession No: AF051901.
Genbank Accession No: AF051900.
Genbank Accession No: AF051899.
Genbank Accession No: AF006194.
Genbank Accession No: U92267.
Genbank Accession No: U92266.
Genbank Accession No: AF024615.
Genbank Accession No: U87909.
Genbank Accession No: U87908.
Genbank Accession No: AF006193.
Genbank Accession No: U50768.
Genbank Accession No: L32134.
Genbank Accession No: AF006195.
Genbank Accession No: AJ005588.

Genbank Accession No: Q43714.
Genbank Accession No: AF061285.
Genbank Accession No: AF043299.
Genbank Accession No: AB022719.
Genbank Accession No: AB023816.
Genbank Accession No: AF043298.
Genbank Accession No: AF043300.
Genbank Accession No: AF042382.
Genbank Accession No: B56118.
Genbank Accession No: C56118.
Genbank Accession No: U20187.
Genbank Accession No: U20189.
Genbank Accession No: U20190.
Lesburg et al., *Current Opinion in Structural Biology*, 1998,
8:695-703.

FIG. 1



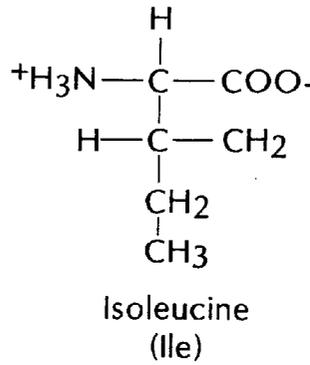
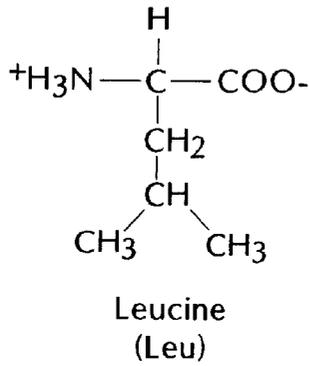
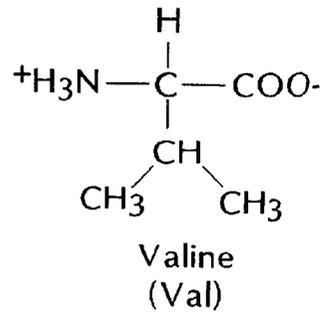
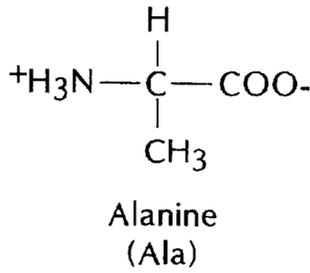
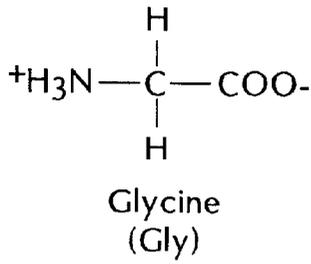
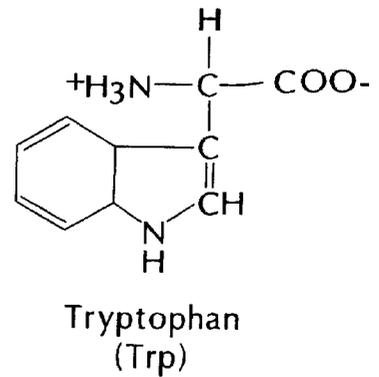
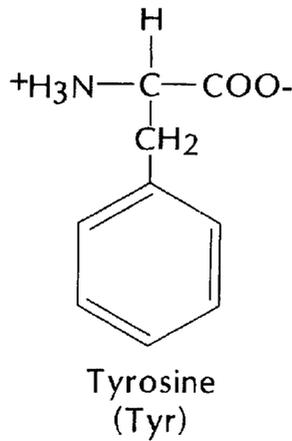
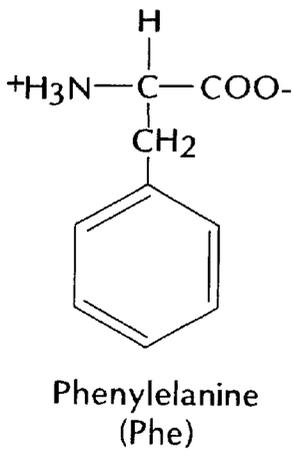
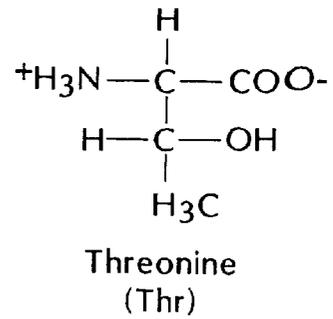
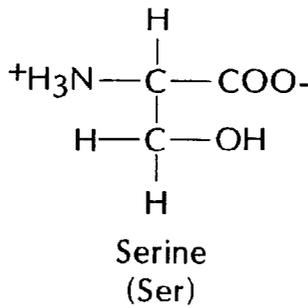
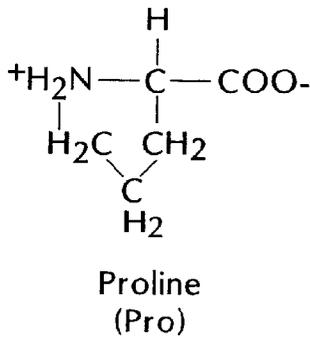


FIG. 2-1



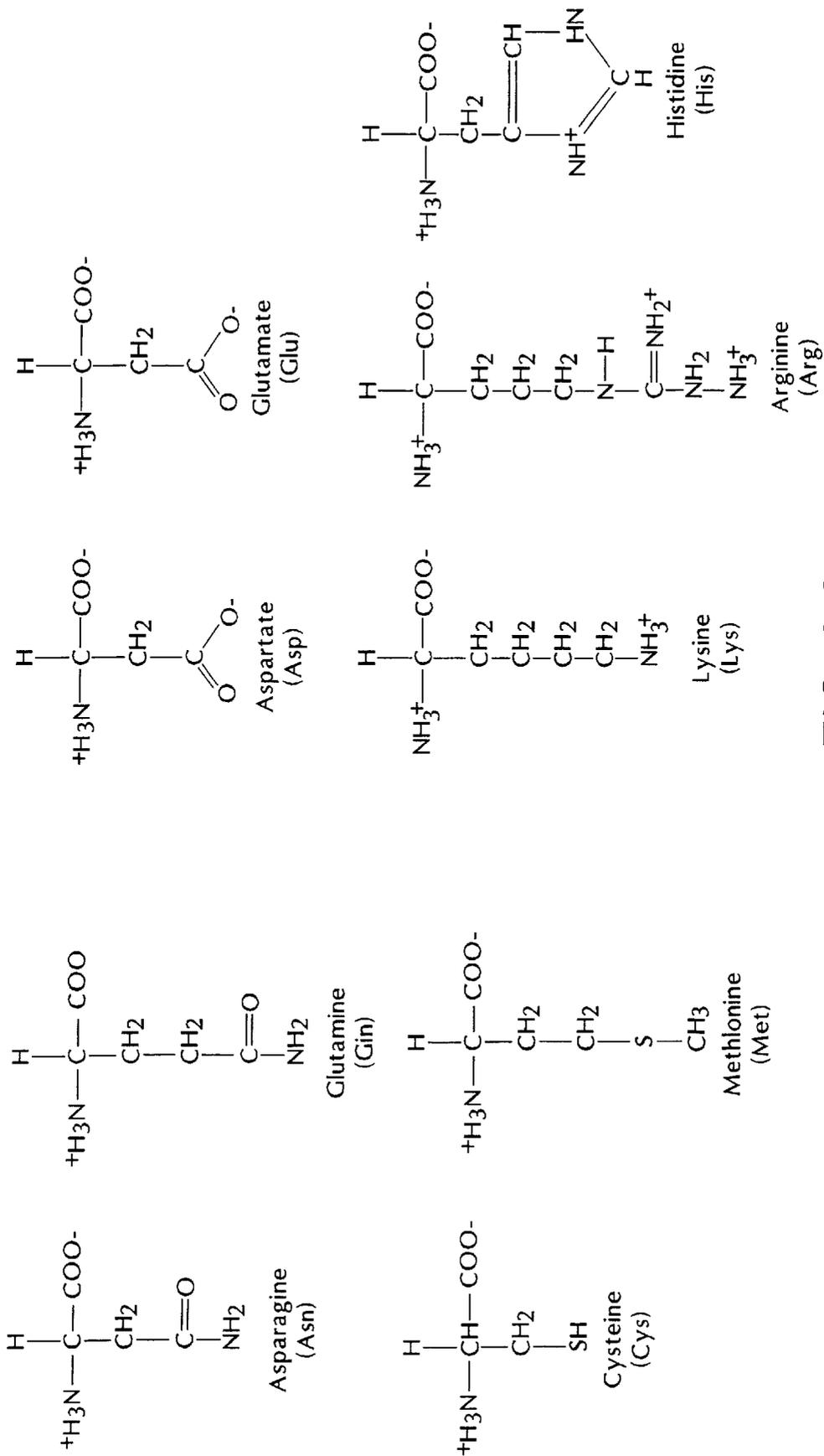
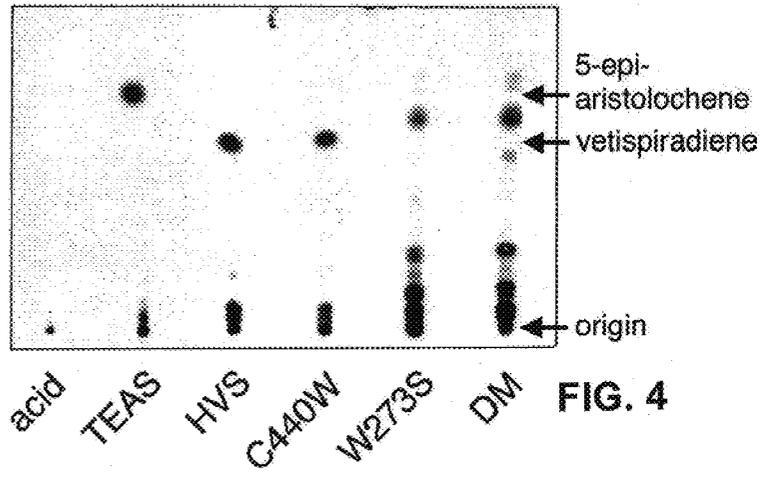
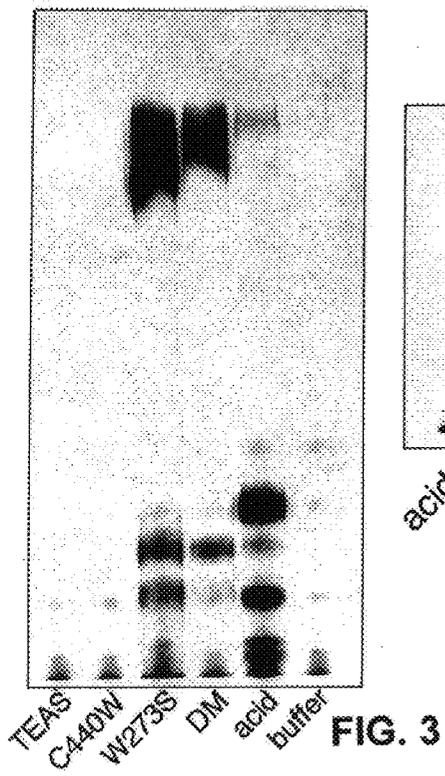


FIG. 2-2



SYNTHESES

CROSS REFERENCE TO RELATED APPLICATION

This application is a divisional of and claims priority under 35 U.S.C. §120 to U.S. application No. 09/398,395, filed Sep. 17, 1999, which claims the benefit of U.S. Provisional Application No. 60/150,262, filed Aug. 23, 1999, U.S. Provisional Application No. 60/130,628, filed Apr. 22, 1999, and U.S. Provisional Application No. 60/100,993 filed Sep. 18, 1998.

This work was supported, in part, with funding from NIH (GM54029 and GM07240) and NSF (IBW-94808152). Therefore, the United States Federal Government may have certain rights in the invention.

BACKGROUND OF THE INVENTION

Isoprenoid compounds are organic molecules produced by a wide range of organisms (e.g., plants, bacteria, fungi, etc). To date, over 23,000 individual isoprenoid molecules have been characterized with tens to hundreds of new structures identified each year. These molecules can fulfill a variety of roles. For example, monoterpenes can be used as fragrances and flavors. Sesquiterpenes and diterpenes can serve as pheromones, defensive agents, visual pigments, antitumor drugs, and components of signal transduction pathways. Triterpenes can serve important functions as membrane constituents and precursors of steroid hormones and bile acids. Polyphenols function as photoreceptive agents and cofactor side chains, and can also exist as natural polymers.

The diverse molecular compounds produced by the isoprenoid pathway are created from diphosphate esters of monounsaturated isoprene units. Isoprenes are added together in multiples of 2, 3, or 4 by prenyl transferases to make C₁₀, C₁₅, and C₂₀ units, respectively. The C₁₀, C₁₅, and C₂₀ molecules, named geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP), respectively, serve as substrates for terpene synthases.

Terpene synthases catalyze the production of isoprenoid compounds via one of the most complex reactions known in chemistry or biology. In general, terpene synthases are moderately sized enzymes having molecular weights of about 40 to 100 kD. As an enzyme, terpene synthases can be classified as having low to moderate turnover rates coupled with exquisite reaction specificity and preservation of chirality. Turnover comprises binding of substrate to the enzyme, establishment of substrate conformation, conversion of substrate to product and product release. Reactions can be performed in vitro in aqueous solvents, typically require magnesium ions as cofactors, and the resulting products, which are often highly hydrophobic, can be recovered by partitioning into an organic solvent.

Terpene synthase genes are found in a variety of organisms including bacteria, fungi and plants. Swapping regions approximating exons between different terpene synthases has identified functional domains responsible for terminal enzymatic steps. For example, work performed on 4-epi-aristolochene synthase (TEAS) from *Nicotiana tabacum* (tobacco) and *Hyoscyamus muticus vetispiradiene* synthase (HVS) from henbane revealed that exon 4 and exon 6, respectively, were responsible for reaction product specificity. Combining functional domains resulted in novel enzymes capable of synthesizing new reaction products (U.S. Pat. No. 5,824,774).

Studies have led to proposed reaction mechanisms for isoprenoid production; see, e.g., Cane et al., 1985, *Bioorg. Chem.*, 13:246-265; Wheeler and Croteau, 1987, *Proc. Natl. Acad. Sci. USA*, 84:4856-4859; and Pyun et al., 1994, *Arch. Biochem. Biophys.*, 308:488-496. The studies used substrate analogs and suicide inhibitors (Croteau, 1994, *Arch. Biochem. Biophys.*, 251:777-782; Cane et al., 1995, *Biochemistry*, 34:2471-2479; and Croteau et al., 1993, *Arch. Biochem. Biophys.*, 307:397-404), as well as chemical-modifying reagents and site-directed mutagenesis in efforts to identify amino acids essential for catalysis (Cane et al., 1995, *Biochemistry*, 34:2480-2488; Rajaonarivony et al., 1992, *Arch. Biochem. Biophys.* 296:49-57; and Rajaonarivony et al., 1992, *Arch. Biochem. Biophys.*, 299:77-82). However, these studies have resulted in limited success in defining the active site due to inherent limitations with these techniques.

SUMMARY OF THE INVENTION

The invention describes a method of identifying alpha-carbon atoms found in the active site of a terpene synthase and describes these atoms in three-dimensional space as well as the spatial relationships among them. The present invention also describes R-groups associated with such alpha-carbons and methods of altering these R-groups in order to create novel terpene synthases capable of generating novel reaction products.

Until the invention taught in this present application, the active site of synthase proteins, the amino acid residues located therein, the amino acid residues involved in catalysis, and the configuration of alpha-carbons and R-groups within the active site have not been known. The current invention now teaches the structure of synthases, as well as provides the means of making and using the information obtained therefrom to develop and produce new and novel synthases having new and novel synthetic capabilities. The data generated using the methods described herein are useful for creation and production of synthase mutants that can use a variety of isoprenoid substrates and produce a variety of isoprenoid products.

In one embodiment, the invention features an isolated terpene synthase having about 20% or greater sequence identity to residues 265 to 535 of SEQ ID NO: 2. Such a synthase comprises nine alpha-carbons having interatomic distances in Angstroms between the alpha-carbons that are ± 2.3 Angstroms of the interatomic distances shown in Table 6. The center point of each alpha-carbon is positioned within a sphere having a radius of 2.3 Angstroms. The center point of each such sphere has the structural coordinates given in Table 5. Each alpha-carbon has an associated R-group, and the synthase has an ordered arrangement of R-groups associated with each alpha-carbon other than the ordered arrangements of R-groups shown in Table 9. The synthase can have about 25% or greater sequence identity to residues 265 to 535 of SEQ ID 2, or about 35% or greater sequence identity to residues 265 to 535 of SEQ ID 2. Such a synthase can catalyze the formation of a terpenoid product from a monoterpene substrate, a sesquiterpene substrate, or a diterpene substrate. The product can be a cyclic terpenoid hydrocarbon or an acyclic terpenoid hydrocarbon. Either type of product can be hydroxylated or non-hydroxylated. The R-group associated with alpha-carbon 1 can be selected from one of the following groups: the group consisting of Cys, Ser, and Thr, the group consisting of Phe, Tyr and Trp, the group consisting of Pro, Gly, and Ala, the group consisting of Glu and Asp, the group consisting of Met, Ile, Val and Leu, the group consisting of Arg and Lys, and the group consisting of Gln,

Asn and His. R-groups associated with α -carbons 2 to 9 can be any amino acid except those having the ordered arrangements of Table 9. Similarly, the R-group associated with each of α -carbons 2-9 can be selected independently from the group consisting of Cys, Ser and Thr, the group consisting of Phe, Tyr and Trp, the group consisting of Pro, Gly, and Ala, the group consisting of Glu and Asp, the group consisting of Met, Ile, Val and Leu, the group consisting of Arg and Lys, and the group consisting of Gln, Asn and His. In these embodiments, R-groups associated with the remaining eight α -carbons except those having the ordered arrangements of Table 9.

In some embodiments, the ordered arrangement of R-groups associated with α -carbons 1 to 9 is Trp, Ile, Thr, Thr, Tyr, Leu, Cys, Thr and Phe, respectively, Ser, Ile, Thr, Thr, Tyr, Leu, Cys, Thr and Tyr, respectively, Trp, Ile, Thr, Thr, Tyr, Leu, Trp, Thr and Tyr, respectively, Ser, Ile, Thr, Thr, Tyr, Leu, Trp, Thr and Tyr, respectively, or Glu, Ile, Thr, Thr, Tyr, Leu, Cys, Thr and Tyr, respectively.

The invention also features a terpene synthase made by aligning the primary amino acid sequence of a preselected terpene synthase polypeptide to the amino acid sequence of residues 265 to 535 of SEQ ID NO: 2, mutating a nucleic acid encoding the preselected polypeptide at one or more codons for nine amino acid residues in a region of the polypeptide primary amino acid sequence having about 20% or greater sequence identity to residues 265 to 535 of SEQ ID NO: 2, the nine residues in the polypeptide aligning with residues 273, 294, 402, 403, 404, 407, 440, 519 and 520 of SEQ ID NO: 2; and expressing the mutated nucleic acid so that a mutated terpene synthase is made.

The invention also features an isolated terpene synthase having about 20% or greater sequence identity to residues 265 to 535 of SEQ ID NO: 2, the synthase comprising sixteen α -carbons having interatomic distances in Angstroms between the α -carbons that are ± 2.3 Angstroms of the interatomic distances given in Table 4. The center point of each α -carbon is positioned within a sphere having a radius of 2.3 Angstroms. The center point of each of the spheres has the structural coordinates given in Table 3. Each α -carbon has an associated R-group, and the synthase has an ordered arrangement of R-groups other than the ordered arrangements of R-groups given in Table 8. The synthase can have about 25% or greater sequence identity to residues 265 to 535 of SEQ ID NO: 2, or about 35% or greater sequence identity to residues 265 to 535 of SEQ ID NO: 2. The synthase can catalyze the formation of a terpenoid product from a monoterpene substrate, a sesquiterpene substrate, or a diterpene substrate. The product can be, for example, a cyclic terpenoid hydrocarbon. The ordered arrangement of R-groups in the synthase associated with α -carbons 1 to 16 can be Cys, Trp, Ile, Ile, Ser, Thr, Thr, Tyr, Leu, Cys, Val, Thr, Tyr, Asp, Phe and Thr, respectively.

The invention also features an isolated terpene synthase having about 20% or greater sequence identity to residues 265 to 535 of SEQ ID NO: 2, the synthase comprising nineteen α -carbons having interatomic distances in Angstroms between the α -carbons that are ± 2.3 Angstroms of the interatomic distances given in Table 2. The center point of each α -carbon is positioned within a sphere having a radius of 2.3 Angstroms. The center points of each sphere have the structural coordinates given in Table 1. Each α -carbon has an associated R-group, and the synthase has an ordered arrangement of the R-groups other than the ordered arrangements of R-groups given in Table 7. The synthase can have about 25% or greater sequence identity to residues 265 to 535 of SEQ ID NO: 2, or about 35% or greater sequence

identity to residues 265 to 535 of SEQ ID NO: 2. The synthase can catalyze the formation of a terpenoid product from a monoterpene substrate, a sesquiterpene substrate, or a diterpene substrate. The product can be, for example, a cyclic terpenoid hydrocarbon.

The invention also features an isolated protein comprising a first domain having an amino terminal end and a carboxyl terminal end. The first domain comprises amino acids that align structurally in three-dimensional space with a glycosyl hydrolase catalytic core, the glycosyl hydrolase catalytic core selected from the group consisting of amino acids 36 to 230 of glucoamylase protein databank (PDB) code 3GLY of *Aspergillus awamori* and amino acids 36 to 230 of endo-glucanase CeID PDB code 1CLC. The isolated protein also comprises a second domain having an amino terminal end and carboxyl terminal end. The second domain comprises amino acids that align structurally in three-dimensional space with avian FPP synthase. The carboxyl terminal end of the first domain is linked to the amino terminal end of the second domain. The second domain has about 20% or greater sequence identity to residues 265 to 535 of SEQ ID NO: 2, and comprises nine α -carbons having interatomic distances in Angstroms between the α -carbons that are ± 2.3 Angstroms of the interatomic distances given in Table 6. The center point of each α -carbon is positioned within a sphere having a radius of 2.3 Angstroms, the center point of each sphere having the structural coordinates given in Table 5. Each α -carbon has an associated R-group, and the synthase has an ordered arrangement of R-groups other than the ordered arrangements of R-groups given in Table 9. The protein can have about 25% or greater sequence identity to SEQ ID NO: 2, or about 35% or greater sequence identity to SEQ ID NO: 2. The synthase can catalyze the formation of a terpenoid product from a monoterpene substrate, a sesquiterpene substrate, or a diterpene substrate. The product can be, for example, a cyclic terpenoid hydrocarbon.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 343 to 606 of SEQ ID NO: 20, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 348, 351, 372, 375, 376, 454, 479, 480, 481, 482, 485, 519, 523, 597, 600, 601, 605, 607 and 608 of SEQ ID NO: 20 are residues other than amino acids Y, L, C, I, T, Y, S, C, G, H, S, L, G, F, G, Y, D, Y and S, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 316 to 586 of SEQ ID NO: 22, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 321, 324, 345, 348, 349, 427, 452, 453, 454, 455, 458, 496, 569, 572, 573, 577, 579 and 580 of SEQ ID NO: 22 are residues other than amino acids C, W, N, I, T, Y, S, I, S, G, M, L, D, A, M, L, D, H and G, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 352 to 622 of SEQ ID NO: 58, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 357, 360, 381, 384, 385, 463, 487,

488, 489, 490, 493, 528, 532, 606, 609, 610, 614, 616 and 617 of SEQ ID NO: 58 are residues other than amino acids Y, M, C, V, T, F, V, S, S, G, I, L, G, F, V, Y, D, Y and T, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to amino acid residues 272 to 540 encoded by SEQ ID NO: 33, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 277, 280, 301, 304, 305, 383, 408, 409, 410, 411, 414, 448, 452, 524, 527, 528, 532, 534 and 535 encoded by SEQ ID NOS: 33 are residues other than amino acids G, W, I, A, S, Y, T, S, G, Y, L, C, D, M, L, Y, D, Y and T, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 319 to 571 of SEQ ID NO: 42, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 324, 327, 348, 351, 352, 430, 455, 456, 457, 458, 461, 495, 499, 571, 574, 575, 579, 581 and 582 of SEQ ID NO: 42 are residues other than amino acids I, W, V, I, S, Y, T, G, L, V, I, N, T, S, Y, D, Y, and T, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 579 to 847 of SEQ ID NO: 44, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 584, 587, 606, 609, 610, 688, 713, 714, 715, 716, 719, 753, 757, 831, 834, 835, 839, 841 and 842 of SEQ ID NO: 44 are residues other than amino acids V, S, G, Q, V, Y, S, V, G, L, C, W, N, V, F, Y, D, Y and G, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 495 to 767 of SEQ ID NO: 46, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 500, 503, 524, 527, 528, 606, 631, 632, 633, 634, 637, 674, 678, 751, 754, 755, 759, 761 and 762 of SEQ ID NO: 46 are residues other than amino acids F, L, A, Q, T, Y, S, I, G, Q, L, S, D, T, I, F, D, F and G, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 295 to 564 of SEQ ID NO: 48, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 300, 303, 324, 327, 328, 406, 431, 432, 433, 434, 437, 471, 475, 548, 551, 552, 556, 558 and

559 of SEQ ID NO: 48 are residues other than amino acids Y, W, A, C, T, Y, S, S, G, M, L, G, D, L, I, Y, D, L and Y, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 307 to 578 of SEQ ID NO: 50, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 312, 315, 336, 339, 340, 419, 444, 445, 446, 447, 450, 484, 488, 562, 565, 566, 570, 572 and 573 of SEQ ID NO: 50 are residues other than amino acids F, W, A, M, T, Y, N, T, G, M, L, S, D, I, M, Y, D, F and S, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 264 to 533 of SEQ ID NO: 52, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 269, 272, 293, 296, 297, 375, 401, 402, 403, 404, 407, 441, 445, 517, 520, 521, 525, 527 and 528 of SEQ ID NO: 52 are residues other than amino acids C, W, L, T, S, Y, S, A, G, Y, I, A, N, A, L, Y, D, Y and S, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 585 to 853 of SEQ ID NO: 56, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 590, 593, 614, 617, 618, 696, 721, 722, 723, 724, 727, 761, 765, 837, 840, 841, 845, 847 and 848 of SEQ ID NO: 56 are residues other than amino acids I, S, S, T, V, Y, S, I, A, L, V, G, N, M, F, Y, D, L and T, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 307 to 574 of SEQ ID NO: 54, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 312, 315, 336, 339, 340, 418, 443, 444, 445, 446, 449, 483, 487, 560, 563, 564, 566, 568 and 569 of SEQ ID NO: 54 are residues other than amino acids C, W, I, I, T, Y, S, I, S, A, I, L, D, A, I, Y, D, D and G, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 309 to 577 of SEQ ID NO: 24, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 314, 317, 338, 341, 342, 420, 446, 447, 448, 449, 452, 485, 489, 560, 563, 564, 569, 571 and 572 of SEQ ID NO: 24 are residues other than amino acids

C, W, N, V, T, Y, I, G, G, I, L, L, D, A, I, Y, D, F and G, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 315 to 584 of SEQ ID NO: 26, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 320, 323, 344, 347, 348, 426, 451, 452, 453, 454, 457, 492, 496, 568, 571, 572, 576, 578 and 579 of SEQ ID NO: 26 are residues other than amino acids S, W, I, A, T, Y, S, V, A, S, I, L, D, A, I, Y, D, F, and G, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 265 to 536 of SEQ ID NO: 28, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 270, 273, 294, 297, 298, 376, 401, 402, 403, 404, 407, 440, 444, 518, 521, 522, 528, 530 and 531 of SEQ ID NO: 28 are residues other than amino acids A, W, V, C, G, F, T, S, C, I, M, G, N, C, S, Y, D, Y and S, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 342 to 612 of SEQ ID NO: 30, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 347, 350, 371, 374, 375, 453, 478, 479, 480, 481, 483, 518, 522, 596, 599, 600, 604, 606 and 607 of SEQ ID NO: 30 are residues other than amino acids F, L, C, V, T, Y, S, S, A, Y, V, L, G, L, L, Y, D, F and S, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more of the ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features an isolated synthase having a region with about 40% or greater sequence identity to residues 273 to 541 of SEQ ID NO: 32, wherein one or more amino acid residues of the synthase that align with amino acid residues at positions 278, 281, 302, 305, 306, 384, 409, 410, 411, 412, 415, 448, 452, 524, 527, 528, 533, 535 and 536 of SEQ ID NO: 32 are residues other than amino acids C, W, I, I, S, Y, T, S, T, Y, L, C, D, I, T, Y, D, Y and T, respectively. In some embodiments, the sequence identity can be about 20% or greater, 25% or greater, or 35% or greater. In some embodiments, one or more ordered arrangements of residues as given in Table 7 are not found in such a synthase.

The invention also features a method for making a terpene synthase, comprising identifying, in a preselected polypeptide having a region with 20% or greater sequence identity to residues 265 to 535 of SEQ ID NO: 2, nine amino acid residues whose α -carbons have interatomic distances in Angstroms between the α -carbons that are ± 2.3 Angstroms of the interatomic distances given in Table 6. The center point of each α -carbon is positioned within a sphere having

a radius of 2.3 Angstroms. The center point of each sphere has the structural coordinates given in Table 5. The method then comprises synthesizing a polypeptide that is modified from the preselected polypeptide. The modified polypeptide has one or more R-groups associated with the nine α -carbons other than the R-groups associated with the α -carbons in the preselected polypeptide. The synthesizing step can comprise the formation of a nucleic acid encoding the preselected polypeptide in which the coding sequence for one or more amino acids corresponding to the nine α -carbons is replaced by a coding sequence that codes for an amino acid different from the amino acid present in the preselected polypeptide. The preselected polypeptide can be, for example, any one of the polypeptides given in SEQ ID NOS: 2, 4, 6, 8, 10, 12, 20, 22, 24, 26, 28, 30, 32, 34-40, 42, 44, 46, 48, 50, 52, 54, 56, or 58.

The invention also features a method of using a terpene synthase, comprising identifying, in a preselected polypeptide having a region with 20% or greater sequence identity to residues 265 to 535 of SEQ ID NO: 2, amino acid residues at nine positions that align with amino acid residues 273, 294, 402, 403, 404, 407, 440, 519 and 520 of SEQ ID NO: 2; and synthesizing a polypeptide that is modified from the preselected polypeptide. The novel polypeptide is modified by having amino acid residues at one or more of the nine positions other than the amino acid residues present in the preselected polypeptide. In some embodiments, the identifying step can comprise identifying sixteen amino acid residues in the preselected polypeptide that align with amino acid residues 270, 273, 294, 297, 298, 402, 403, 404, 407, 440, 516, 519, 520, 525, 527 and 528 of SEQ ID NO: 2, and the synthesizing step can comprise synthesizing a polypeptide that is modified from the preselected polypeptide, the modified polypeptide having amino acid residues at one or more of the sixteen positions other than the amino acid residues present in the preselected polypeptide. In some embodiments, the identifying step can comprise identifying nineteen amino acid residues in the preselected polypeptide that align with amino acid residues 270, 273, 294, 297, 298, 376, 401, 402, 403, 404, 407, 440, 444, 516, 519, 520, 525, 527 and 528 of SEQ ID NO: 2, and the synthesizing step can comprise synthesizing a polypeptide that is modified from the preselected polypeptide, the modified polypeptide having amino acid residues at one or more of the nineteen positions other than the amino acid residues present in the preselected polypeptide. The synthesizing step can comprise the formation of a nucleic acid encoding the preselected polypeptide in which the coding sequence in the nucleic acid coding for one or more of the identified amino acid residues is replaced by a coding sequence that encodes an amino acid different from the amino acid present in the preselected polypeptide. The preselected polypeptide can be, for example, any one of the polypeptides given in SEQ ID NOS: 2, 4, 6, 8, 10, 12, 20, 22, 24, 26, 28, 30, 32, 34-40, 42, 44, 46, 48, 50, 52, 54, 56, or 58. The method can further comprise: contacting the modified polypeptide with an isoprenoid substrate under conditions effective for the compound to bind the polypeptide; and measuring the ability of the modified polypeptide to catalyze the formation of a reaction product from the isoprenoid substrate. The isoprenoid substrate can be a monoterpene, a sesquiterpene, or a diterpene.

The invention also features a method of making a terpene synthase, comprising creating a population of nucleic acid molecules that encode polypeptides, the population having members that differ from one another at one or more of nine codons specifying amino acids of a preselected terpene

synthase having a region with about 20% or greater sequence identity to residues 265 to 535 of SEQ ID NO: 2, α -carbons of the nine amino acids having interatomic distances in Angstroms between the α -carbons that are ± 2.3 Angstroms of the interatomic distances given in Table 6. The center point of each α -carbon is positioned within a sphere having a radius of 2.3 Angstroms, and the center point of each sphere has the structural coordinates given in Table 5. In some embodiments, the codons specify amino acids as described in Tables 1–2 or 3–4 of a preselected terpene synthase. A portion, or all, of the nucleic acid population is expressed so that a population of polypeptides is made. At least one member of the population of polypeptides is a mutant terpene synthase. The expressing step can comprise in vitro transcription and in vitro translation of the nucleic acid population. In some embodiments, the expressing step comprises cloning members of the nucleic acid population into an expression vector, introducing the expression vector into host cells and expressing the cloned nucleic acid population members in the host cells so that the population of polypeptides is made. The preselected terpene synthase polypeptide can be a monoterpene synthase, a sesquiterpene synthase, or a diterpene synthase. The host cells can be prokaryotic cells or eukaryotic cells, including, without limitation, bacterial cells, fungal cells, and animal cells, e.g., mammalian cells or insect cells. The host cells can also be plant cells, e.g., a cell from a Gramineae plant, a cell from a Legumineae plant, a cell from a Solanaceae plant, a cell from a Brassicaceae plant or a cell from a Conifereae plant.

The invention also features a nucleic acid encoding a synthase as described herein, and a host cell containing such a nucleic acid. The invention also features a transgenic plant containing such a nucleic acid, or a transgenic animal cell culture containing such a nucleic acid.

In some embodiments, a synthase polypeptide of the invention comprises a domain that contains an active site comprised of nine α -carbon atoms having the coordinates of Table 5, and interatomic distances between the α -carbons ± 2.3 angstroms of the distances given in Table 5. The α -carbon atoms align structurally in three dimensional space in the presence or absence of bound substrate or substrate analogue, with avian FPP synthase. In another embodiment, a synthase of this invention comprises the following: (i) a first domain containing amino acid residues that align in three-dimensional space (in solution or crystal form, and either having a bound or unbound substrate) with a glycosyl hydrolase catalytic core selected from the group consisting of (a) amino acids 36–230 of glycosyl hydrolase (PDB code 3GLY) of *Aspergillus awamori*, and (b) amino acids 36–230 of endoglucanase CellB (PDB code 1CLC), and (ii) a second domain that aligns structurally in three dimensional space with or without substrate or substrate analogues bound in the active site with avian FPP synthase. The second domain contains an active site comprised of nine, sixteen or nineteen α -carbon atoms having the structural coordinates and interatomic distances of Tables 1–2, 3–4 or 5–6. These α -carbon atoms have R-groups attached thereto that can interact, either directly or indirectly, with an isoprenoid substrate.

The invention also features a method for generating mutant terpene synthases possessing catalytic activity. The method comprises the steps of (a) providing a crystallographic model of a preselected catalytically active terpene synthase having an active site, and (b) using the model to design a terpene synthase having at least one altered R-group in the active site relative to the preselected synthase. The invention also features terpene synthases having

altered substrate specificity, methods of making the same, and procedures for generating three-dimensional structures thereof.

Although methods and materials similar or equivalent to those described herein can be used to practice the invention, suitable methods and materials are described below. All publications, patent applications, patents and other references mentioned herein are incorporated by reference in their entirety.

Other aspects, embodiments, advantages, and features of the present invention will become apparent from the specification.

BRIEF DESCRIPTION OF DRAWINGS

FIG. 1. Schematic representation of tobacco 5-epi-aristolochene synthase (TEAS) with bound farnesyl hydroxyphosphonate (FHP), prepared using the RIBBONS software program of Carson, M. and Bugg, C., J. Mol. Graphics 4:121 (1986). Cylinders 1–8 and A represent α -helices in the NH₂-terminal domain; cylinders C, D, D1, D2, E, F, G1, G2, H1, H2, H3, I and α -1 represent α -helices in the COOH-terminal domain.

FIG. 2. Structure of twenty natural amino acids showing α -carbons and associated R-groups.

FIG. 3. Autoradiogram of an argentation thin-layer chromatogram of terpenoid hydrocarbon products made by TEAS and mutant TEAS enzymes using GGPP as a substrate. DM: W273S/C440W mutant TEAS enzyme.

FIG. 4. Autoradiogram of an argentation thin-layer chromatogram of terpenoid hydrocarbon products made by TEAS and mutant TEAS enzymes using FPP as a substrate.

BRIEF DESCRIPTION OF TABLES

Table 1. X-ray crystallographic structural coordinates for 19 α -carbons found in the active site of a terpene synthase.

Table 2. Interatomic distances in Angstroms between each α -carbon of Table 1. Each α -carbon occupies a sphere having a radius of 2.3 Angstroms. Interatomic distances are calculated from the center point of each sphere.

Table 3. X-ray crystallographic structural coordinates for 16 α -carbons found in the active site of a terpene synthase.

Table 4. Interatomic distances in Angstroms between each α -carbon of Table 3. Each α -carbon occupies a sphere having a radius of 2.3 Angstroms. Interatomic distances are calculated from the center point of each sphere.

Table 5. X-ray crystallographic structural coordinates for nine α -carbons found in the active site of a terpene synthase.

Table 6. Interatomic distances in Angstroms between each α -carbon of Table 5. Each α -carbon occupies a sphere having a radius of 2.3 Angstroms. Interatomic distances are calculated from the center point of each sphere.

Table 7. Ordered arrangement of R-groups not found associated with the α -carbons of Table 1.

Table 8. Ordered arrangement of R-groups not found associated with the α -carbons of Table 3.

Table 9. Ordered arrangement of R-groups not found associated with the α -carbons of Table 5.

Table 10. X-ray structural coordinates for TEAS having the substrate analog FHP bound in the active site.

Table 11. X-ray structural coordinates for TEAS in the absence of substrate.

Table 12. Alignment of residues 265–535 of TEAS with a limonene synthase, SEQ ID NO: 22, using the BLASTp alignment program.

Table 13. Alignment of residues 579 to 847 of SEQ ID NO:44 with SEQ ID NO:26, using the BLASTp program.

Table 14. Alignment of residues 265 to 535 of TEAS with SEQ ID NO:48, using the BLASTp program.

Table 15. Alignment of residues 307 to 593 of SEQ ID NO:50 with SEQ ID NO:56 using the BLASTp program.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

SEQ ID NO:1 is the DNA coding sequence for a tobacco 5-epi-aristolochene synthase (TEAS) protein. Genbank No: Q40577.

SEQ ID NO:2 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:1.

SEQ ID NO:3 is the DNA coding sequence for a TEAS protein in which the codon for Trp273 has been changed to a codon for Glu.

SEQ ID NO:4 is the amino acid sequence for the W273E protein encoded by the TEAS DNA of SEQ ID NO:3.

SEQ ID NO:5 is the DNA coding sequence for a TEAS protein in which the codon for Tyr520 has been changed to a codon for Phe.

SEQ ID NO:6 is the amino acid sequence for the Y520F protein encoded by the TEAS DNA of SEQ ID NO:5.

SEQ ID NO:7 is the DNA coding sequence for a TEAS protein in which the codon for Tyr527 has been changed to a codon for Phe.

SEQ ID NO:8 is the amino acid sequence for the Y527F protein encoded by the TEAS DNA of SEQ ID NO:7.

SEQ ID NO:9 is the DNA coding sequence for a TEAS protein in which the codon for Trp273 has been changed to a codon for Ser and the codon for Cys440 has been changed to a codon for Trp.

SEQ ID NO:10 is the amino acid sequence for the W273S/C440W protein encoded by the TEAS DNA of SEQ ID NO:9.

SEQ ID NO:11 is the DNA coding sequence for TEAS proteins in which the codons for Tyr406 and Leu407 have each been changed to the nucleotides NNS.

SEQ ID NO:12 is the amino acid sequence for the population of Y406X/L407X proteins encoded by the TEAS DNA of SEQ ID NO:11, where X is any naturally occurring amino acid.

SEQ ID NO:13 is a DNA primer sequence.

SEQ ID NO:14 is a DNA primer sequence.

SEQ ID NO:15 is a DNA primer sequence.

SEQ ID NO:16 is a DNA primer sequence.

SEQ ID NO:17 is a DNA primer sequence.

SEQ ID NO:18 is a DNA primer sequence.

SEQ ID NO:19 is the DNA coding sequence for a grand fir pinene synthase. Genbank Accession No: U87909.

SEQ ID NO:20 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:19.

SEQ ID NO:21 is the DNA coding sequence for a spearmint limonene synthase. Genbank Accession No: L13459.

SEQ ID NO:22 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:21.

SEQ ID NO:23 is the DNA coding sequence for a sage 1, 8 cineole synthase. Genbank Accession No: AF051899.

SEQ ID NO:24 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:23.

SEQ ID NO:25 is the DNA coding sequence for a sage bornyl diphosphate synthase. Genbank Accession No: AF051900.

5 SEQ ID NO:26 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:25.

SEQ ID NO:27 is the DNA coding sequence for a mint E-b-farnesene synthase. Genbank Accession No: AF024615.

10 SEQ ID NO:28 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:27.

SEQ ID NO:29 is the DNA coding sequence for a grand fir myrcene synthase. Genbank Accession No: U87908.

SEQ ID NO:30 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:29.

15 SEQ ID NO:31 is the DNA coding sequence for a potato vetaspiradiene synthase. Genbank Accession No: AB022598.

SEQ ID NO:32 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:31.

20 SEQ ID NO:33 is the genomic DNA coding sequence for a cotton delta-cadinene synthase. Genbank Accession No: Y18484.

SEQ ID NOS:34-40 are the amino acid sequences for the exons encoded by the DNA of SEQ ID NO:33.

SEQ ID NO:41 is the DNA coding sequence for a castor bean casbene synthase. Genbank Accession No: L32134.

SEQ ID NO:42 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:41.

30 SEQ ID NO:43 is the DNA coding sequence for a yew taxadiene synthase. Genbank Accession No: U48796.

SEQ ID NO:44 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:43.

35 SEQ ID NO:45 is the DNA coding sequence for a grand fir E-alpha-bisabolene synthase. Genbank Accession No: AF006194.

SEQ ID NO:46 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:45.

40 SEQ ID NO:47 is the DNA coding sequence for a grand fir delta-selinene synthase. Genbank Accession No: U92266.

SEQ ID NO:48 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:47.

45 SEQ ID NO:49 is the DNA coding sequence for a grand fir gamma-humulene synthase. Genbank Accession No: U92267.

SEQ ID NO:50 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:49.

50 SEQ ID NO:51 is the DNA coding sequence for a tomato germacrene C synthase. Genbank Accession No: AF035631.

SEQ ID NO:52 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:51.

55 SEQ ID NO:53 is the DNA coding sequence for a sage+sabinene synthase. Genbank Accession No: AF051901.

SEQ ID NO:54 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:53.

60 SEQ ID NO:55 is the DNA coding sequence for a grand fir abietadiene synthase. Genbank Accession No: U50768.

SEQ ID NO:56 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:55.

65 SEQ ID NO:57 is the DNA coding sequence for a grand fir limonene synthase. Genbank Accession No: AF006193.

SEQ ID NO:58 is the amino acid sequence for the protein encoded by the DNA of SEQ ID NO:57.

DETAILED DESCRIPTION

The following terms are used herein:

" α -carbon" refers to the chiral carbon atom found in an amino acid residue. Four substituents are covalently bound to the α -carbon, including an amino group, a carboxyl group, a hydrogen atom, and an R-group.

"R-group" refers to a substituent attached to the α -carbon of an amino acid residue that is not involved in peptide bond formation in a protein. An R-group is an important determinant of the overall chemical character of an amino acid. The twenty naturally occurring amino acids found in proteins and the R-groups associated with the α -carbon of each amino acid are listed in FIG. 2. The three-letter and one-letter abbreviations for naturally occurring amino acids are sometimes used herein to refer to the R-group associated with a particular amino acid.

"Naturally occurring amino acid" includes L-isomers of the twenty amino acids naturally occurring in proteins. Naturally occurring amino acids are glycine, alanine, valine, leucine, isoleucine, serine, methionine, threonine, phenylalanine, tyrosine, tryptophan, cysteine, proline, histidine, aspartic acid, asparagine, glutamic acid, glutamine, arginine, and lysine. Unless specially indicated, all amino acids referred to in this application are in the L-form. Three-letter and one-letter abbreviations are sometimes used herein to refer to naturally occurring amino acids. These abbreviations are known in the art.

"Unnatural amino acid" includes amino acids that are not naturally found in proteins. Examples of unnatural amino acids included herein are racemic mixtures of selenocysteine and selenomethionine. In addition, unnatural amino acids include the D or L forms of norleucine, para-nitrophenylalanine, homophenylalanine, para-fluorophenylalanine, 3-amino-2-benzylpropionic acid, homoarginine, D-phenylalanine, and the like.

"Positively charged amino acid" includes any naturally occurring or unnatural amino acid having an R-group that carries a positive charge under normal physiological conditions. Examples of positively charged, naturally occurring amino acids include arginine and lysine.

"Negatively charged amino acid" includes any naturally occurring or unnatural amino acid having an R-group that carries a negative charge under normal physiological conditions. Examples of negatively charged, naturally occurring amino acids include aspartic acid and glutamic acid.

"Hydrophobic amino acid" includes any naturally occurring or unnatural amino acid having an uncharged, nonpolar side chain under normal physiological conditions. Examples of naturally occurring hydrophobic amino acids are leucine, isoleucine, valine and methionine.

"Hydrophilic amino acid" includes any naturally occurring or unnatural amino acid having a charged polar side chain. Examples of naturally occurring hydrophilic amino acids include serine, threonine and cysteine.

"Mutant terpene synthase" or "mutated terpene synthase" refers to a synthase polypeptide having a primary amino acid sequence. The center point of the α -carbon of nine residues of the polypeptide is positioned within a sphere having a radius of 2.3 Angstroms; the center points of the nine spheres have the structural coordinates of Table 5 or coordinates which can be rotated and/or translated to coincide with the coordinates of Table 5. The relative interatomic distances between the nine α -carbons is ± 2.3 angstroms of the interatomic distances given in Table 6. Each α -carbon has an associated R-group. A mutant synthase differs from a

non-mutant synthase in the ordered arrangement of R-groups associated with the nine α -carbons. A mutant synthase has an ordered arrangement of R-groups on the nine α -carbons other than the ordered arrangements of R-groups listed in Table 9. R-groups associated with other α -carbons of the synthase primary amino acid sequence may or may not be the same as in a non-mutated synthase.

In some embodiments, a mutant synthase refers to a synthase in which the center point of the α -carbon of sixteen residues of the polypeptide is positioned within a sphere having a radius of 2.3 Angstroms; the center points of the sixteen spheres have the structural coordinates of Table 3 or coordinates which can be rotated and/or translated to coincide with the coordinates of Table 3. The relative interatomic distances between the nine α -carbons is ± 2.3 angstroms of the interatomic distances given in Table 4. Each α -carbon has an associated R-group. A mutant synthase differs from a non-mutant synthase in the ordered arrangement of R-groups associated with the sixteen α -carbons. A mutant synthase has an ordered arrangement of R-groups on the sixteen α -carbons other than the ordered arrangements of R-groups listed in Table 8. R-groups associated with other α -carbons of the synthase primary amino acid sequence may or may not be the same as in a non-mutated synthase.

In some embodiments, a mutant synthase refers to a synthase in which the center point of the α -carbon of nineteen residues of the polypeptide is positioned within a sphere having a radius of 2.3 Angstroms; the center points of the nineteen spheres have the three dimensional coordinates of Table 1 or coordinates which can be rotated and/or translated to coincide with the coordinates of Table 1. The relative interatomic distances between the nineteen α -carbons is ± 2.3 angstroms of the interatomic distances given in Table 2. Each α -carbon has an associated R-group. A mutant synthase differs from a non-mutant synthase in the ordered arrangement of R-groups associated with the nineteen α -carbons. A mutant synthase has an ordered arrangement of R-groups on the nineteen α -carbons other than the ordered arrangements of R-groups listed in Table 7. R-groups associated with other α -carbons of the synthase primary amino acid sequence may or may not be the same as in a non-mutated synthase.

"Nonmutated synthase" or "non-mutant synthase" includes a synthase having a primary amino acid sequence comprising nine, sixteen, or nineteen amino acid residues. The center point of each α -carbon of these residues is positioned within a sphere having a radius of 2.3 Angstroms; the center points of the spheres have the three dimensional coordinates of Tables 5, 3, or 1, respectively, or coordinates which can be rotated and/or translated to coincide with the coordinates of Tables 5, 3, or 1. The relative interatomic distances between the nine, sixteen, or nineteen α -carbons is ± 2.3 angstroms of the interatomic distances given in Tables 6, 4, or 2, respectively. Each α -carbon has an associated R-group. A non-mutant synthase has an ordered arrangement of R-groups on the nine, sixteen, or nineteen α -carbons as listed in Tables 9, 8, or 7, respectively.

"Degenerate variations thereof" refers to variants of a gene coding sequence by which the same polypeptide is encoded by different nucleotide sequences, due to the degeneracy of the genetic code. For example, synthases of the present invention have a primary amino acid sequence. Degenerate synthase variations are different nucleic acid coding sequences that nevertheless encode the same primary amino acid sequence due to the degeneracy of the genetic code.

"Expression" refers to transcription of a gene or nucleic acid molecule and the translation of that nucleic acid into a

polypeptide. Expression of genes also involves processing of RNA into mRNA in eukaryotic systems. It is not necessary for the genes to integrate into the genome of a cell in order to achieve expression. This definition is not limited to expression in a particular system or a particular cell type and includes, without limitation, stable, transient, in vitro, and in vivo expression.

"Promoter" and "promoter regulatory element", refers to a nucleic acid that is involved in controlling expression of a gene. Promoter regulatory elements, and the like, from a variety of sources can be used efficiently to promote gene expression. Promoter regulatory elements include constitutive, tissue-specific, developmental-specific, inducible, subgenomic promoters, and the like. Promoter regulatory elements may also include certain enhancer elements or silencing elements that improve or regulate transcriptional efficiency.

"Active Site" refers to a site in a terpene synthase that binds the hydrophobic portion of a terpene substrate, GPP, FPP, and/or GGPP. The active site can, under certain conditions, catalyze a biosynthetic reaction that allows one or more reaction products to be produced.

"Altered enzymatic specificity" includes an alteration in the ability of a mutant synthase to use a particular terpene substrate or a change in the profile of reaction product(s) from a mutant synthase, compared to the substrate specificity of and the reaction products made by a corresponding non-mutated synthase. Altered specificity may include the ability of a synthase to exhibit different enzymatic parameters relative to a non-mutated synthase (K_m , V_{max} , etc), and/or to produce products that are different from those that are produced by a corresponding non-mutant synthase.

"Structure coordinates" or "structural coordinates" refers to Cartesian coordinates (x, y, and z positions) derived from mathematical equations involving Fourier synthesis as determined from patterns obtained via diffraction of a monochromatic beam of X-rays by the atoms (scattering centers) of a synthase molecule in crystal form. Diffraction data are used to calculate electron density maps of repeating protein units in the crystal (unit cell). Electron density maps are used to establish the positions of individual atoms within a crystal's unit cell. The absolute values for structural coordinates listed herein convey relative spatial relationships between atoms because the absolute values ascribed to structural coordinates can be changed by rotational and/or translational movement along the x, y and/or z axes, together or separately, while maintaining the same relative spatial relationships among atoms. Thus, a terpene synthase whose absolute values for a set of structural coordinates can be rotationally or translationally adjusted to coincide with the particular values listed in Tables 1, 3, or 5 is considered to have the same structural coordinates as those of Tables 1, 3 or 5. An example of structural coordinates that coincide with the absolute values listed herein after rotation and/or translation are the coordinates of Table 11.

"Heavy atom derivatization" refers to a method of producing a chemically modified form of a synthase crystal. In practice, a crystal is soaked in a solution containing heavy atom salts or organometallic compounds, e.g., lead chloride, gold thiomalate, thimerosal, uranyl acetate and the like, which can diffuse through the crystal and bind to the protein's surface. Locations of the bound heavy atoms can be determined by X-ray diffraction analysis of the soaked crystal. The information is then used to construct phase information which can then be used to construct three-dimensional structures of the enzyme as described in

Blundel, T. L., and Johnson, N. L., Protein Crystallography, Academic Press (1976).

"Unit cell" refers to a basic parallelepiped shaped block. Regular assembly of such blocks may construct the entire volume of a crystal. Each unit cell comprises a complete representation of the unit pattern, the repetition of which builds up the crystal.

"Mutagenesis" refers to the substitution of a different amino acid residue at a particular position in the primary amino acid sequence of a protein, thereby changing the R-group present at that position. Mutagenesis can be most easily performed by changing the coding sequence of a nucleic acid encoding the protein so that the coding sequence in the nucleic acid specifies an amino acid residue different from the residue initially present at that position.

"Space Group" refers to the arrangement of symmetry elements within a crystal.

"Molecular replacement" refers to the generation of a preliminary model of a synthase whose structural coordinates are unknown, by orienting and positioning a molecule whose structural coordinates are known within the unit cell of the unknown crystal so as best to account for the observed diffraction pattern of the unknown crystal. Phases can then be calculated from this model and combined with the observed amplitudes to give an approximate Fourier synthesis of the structure whose coordinates are unknown. This in turn can be subject to any of the several forms of refinement to provide a final, accurate structure of the unknown crystal (Lattman, E., 1985, in *Methods in Enzymology*, 115:55-77; Rossmann, MG., ed., "The Molecular Replacement Method" 1972, *Int. Sci. Rev. Ser.*, No. 13, Gordon & Breach, New York). Using structure coordinates and interatomic distance matrices, molecular replacement may be used to determine the structural coordinates of a crystalline mutant, homologue, or a different crystal form of terpene synthase.

"Recombinant protein" includes a protein that is chemically synthesized or derived biosynthetically from an isolated gene.

"Gene" includes naturally derived or genetically manipulated nucleic acids that contain the information needed to produce a polypeptide.

"Nucleic acid" includes any genetic material comprised of the nucleotides guanine, adenine, thymine, cytosine, uracil, inosine and the like. Nucleic acids may be single-, double-, or triple stranded. Nucleic acids may be deoxyribonucleic acid or ribonucleic acid.

"Genetically manipulated" includes genes that have been modified to contain a different nucleotide sequence from that present in a preselected nucleic acid. Genes can be manipulated by synthetically or via traditional cloning, PCR, chemical gene synthesis, direct or random mutagenesis, and gene shuffling. Genetically manipulated also includes the process of making genes that are degenerate variations of nucleic acids encoding preselected proteins.

"First domain" includes polypeptides having a first and second end wherein the first end can have an amino terminal amino acid with a free amino group and can be linked by a peptide bond to a second amino acid. The first end may also be modified through acetylation and the like. The second end of the first domain may or may not have a free carboxyl terminal group.

"Second domain" includes polypeptides having a first and second end wherein the first end can have an amino terminal amino acid and can be linked by a peptide bond to a second

amino acid. The second end of the second domain may or may not have a carboxyl terminal group. Typically, the first end of the second domain is linked to the second end of the first domain via a peptide bond.

“Isoprenoid substrate” refers to the C₁₀, C₁₅, and C₂₀ molecules, named geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP), respectively.

“Sequence identity” or “percent sequence identity” refers to the percentage of amino acids or nucleotides that occupy the same relative position when two protein sequences or nucleic acid sequences, a query sequence and a subject sequence, are aligned. The number of amino acid or nucleotide residues that are identical between both the subject and query sequences are counted, divided by the number of residues in the query sequence, and multiplied by 100. The process is repeated until the alignment resulting in the highest percent sequence identity is found. Percent sequence identity can be determined by visual inspection and/or by using various computer programs, e.g., MegAlign (DNASTAR, Inc., Madison, Wis.) or BLAST programs available on the world wide web from the National Center for Biotechnology Information (NCBI). Gaps of one or more residues may sometimes be inserted to maximize sequence alignments to structurally conserved domains of the query sequence, i.e., α -helices, β -sheets and loops.

“Monoterpene product” refers to linear, cyclized, and/or hydroxylated reaction products made from the substrate GPP. “Sesquiterpene produce refers to linear, cyclized, and/or hydroxylated reaction products made from the substrate FPP.

“Diterpene product” refers to linear, cyclized, and/or hydroxylated reaction products made from the substrate GGPP.

The present invention relates to terpene synthases and mutants thereof from which the position of specific α -carbon atoms and R-groups associated therewith comprising the active site can be determined in three-dimensional space. The invention also relates to structural coordinates of the synthases, use of the structural coordinates to develop structural information related to synthase homologues, mutants, and the like, and to crystal forms of such synthases. Furthermore, the invention provides a method whereby α -carbon structural coordinates for atoms comprising the active site of a preselected terpene synthase can be used to develop synthases in which R-groups associated with active site α -carbon atoms are different from the R-groups found in the preselected terpene synthase. In addition, the present invention provides for the production of novel terpene synthases based on the structural information provided herein and for the use of such synthases to make a variety of isoprenoid compounds.

The present invention further provides, for the first time, crystals of a synthase, as exemplified by tobacco 5epi-aristolochene synthase (TEAS), which are grown in the presence or absence of substrate and substrate analogues, thus allowing definition of the structural coordinates associated therewith. The structural coordinates allow determination of the carbon atoms comprising the active site and R-groups associated therewith. The crystals of the present invention belong to the tetragonal space group P4₁2₁2; the unit cell dimensions vary by a few angstroms between crystals but on average a=126 angstroms, c=122 angstroms, a=b, $\alpha=90^\circ$, $\beta=90^\circ$, and $\gamma=90^\circ$.

Structural coordinates are preferably obtained at a resolution of about 2.2 to about 2.8 angstroms for a synthase in

the presence and in the absence of bound substrate or substrate analog. Coordinates for a synthase with a substrate analog bound in the active site are given in Table 10. Coordinates for a synthase in the absence of a substrate analog bound in the active site are given in Table 11. Those skilled in the art understand that a set of structure coordinates determined by X-ray crystallography is not without standard error. Therefore, for the purpose of this invention, any set of structure coordinates wherein the active site α -carbons of a synthase, synthase homologue, or mutants thereof, have a root mean square deviation less than ± 2.3 angstroms when superimposed using the structural coordinates listed in Table 1, 3, or 5, are considered identical.

A schematic representation of the three-dimensional shape of a synthase is shown in FIG. 1 which was prepared by RIBBONS (Carson and Bugg, 1986, J. Mol. Graphics, 4:121). The synthase shown in FIG. 1 consists entirely of α -helices and short connecting loops and turns, organized into first and second structural domains.

In one embodiment, an isolated synthase of the invention comprises sixteen active site α -carbons having the structural coordinates of Table 3 and the relative distances ± 2.3 angstroms of the distances given in Table 4. The active site α -carbons of Table 3 generally are not all contiguous, i.e., are not adjacent to one another in the primary amino acid sequence of a synthase, due to intervening amino acid residues between various active site α -carbons. On the other hand, it should be appreciated that some of the active site α -carbons can be adjacent to one another in some instances. In the embodiment depicted in the TEAS Y527F protein (SEQ ID NO:8), for example, active site α -carbons are adjacent to one another in the primary amino acid sequence at positions 402, 403 and 404, respectively, whereas active site α -carbons at residues 273 and 294 are separated and thus are not adjacent. Thus, the numbering of active site α -carbons given in Tables 1, 2, 3, 4, 5, or 6 is merely for convenience and such α -carbons may reside at any position in the primary amino acid sequence that achieves the structural coordinates given in Tables 1, 3, or 5 and the relative interatomic distances ± 2.3 angstroms given in Tables 2, 4, or 6.

An appropriate combination of R-groups, linked to active site α -carbons, can facilitate the formation of one or more desired reaction products. The combination of R-groups selected for use in a terpene synthase of the invention can be any combination other than the ordered arrangements of R-groups and corresponding active site α -carbons shown in Tables 7, 8, or 9. An illustrative example of a suitable arrangement of R-groups and α -carbons is Cys, Trp, Ile, Ile, Ser, Thr, Thr, Tyr, Leu, Cys, Val, Thr, Phe, Asp, Tyr and Thr, associated with active site α -carbons 1 to 16, respectively, of Table 3. Another example of a suitable arrangement of R-groups and α -carbons is Cys, Trp, Ile, Ile, Ser, Thr, Thr, Tyr, Leu, Cys, Val, Thr, Tyr, Asp, Phe, and Thr at active site α -carbons 1 to 16, respectively, of Table 3. In some embodiments, a synthase of the invention may have primary amino acid sequences as listed in SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, and SEQ ID NO:10, DNA molecules encoding the same, which are listed in SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, and SEQ ID NO:9, respectively, and degenerate variations thereof. Typically, R-groups found on active site α -carbons are those found in naturally occurring amino acids. See, e.g., FIG. 2. In some embodiments, however, R-groups other than naturally occurring amino acids can be used.

Some arrangements of R-groups and active site α -carbons result in mutant terpene synthases that form reaction prod-

ucts. Such enzymatically active synthases and their corresponding genes are useful to make known terpenoid hydrocarbons, e.g., monoterpenes such as pinene, sesquiterpenes such as delta-cadinene and diterpenes such as abietadiene. Other enzymatically active synthases can be used to make novel terpenoid products.

Some arrangements of R-groups and active site α -carbons may result in mutant terpene synthases that do not form reaction product(s) at a desired rate. Such synthases and their genes are useful as controls in analyses of product formation by enzymatically active mutant synthases. Such synthases and their genes can also be useful in analyses of translation of enzymatically active mutant synthase genes, or as nutritional supplements. Such synthases can be attached to Sepharose beads and used for affinity purification of isoprenoid compounds from crude preparations. In addition, such synthases and their genes can also be useful to develop reagents for various purposes, e.g., immunological reagents to monitor expression of a terpene synthase protein or nucleic acid probes or primers to monitor inheritance of a terpene synthase gene in a plant breeding program.

In some embodiments, the α -carbon backbone of a synthase first domain aligns structurally with the catalytic core of glycosyl hydrolases, as exemplified by glucoamylase (Brookhaven Protein Database (PDB) code 3GLY) from *Aspergillus awamori* (Aleshin et al., 1994, J. Mol. Biol., 238:575) and endoglucanase CelD (PDB code ICLC) from *Clostridium thermocelum* (Juy et al., 1992, Nature, 357:89), and the α -carbon backbone of a synthase second domain, which contains the active site, aligns structurally with avian farnesyl diphosphate synthase (FPS), wherein the active site is comprised of 9, 16, or 19 amino acid residues with α -carbon structural coordinates as listed in Tables 1, 3, or 5 and interatomic distances as described in Tables 2, 4, or 6. Such α -carbons have an ordered arrangement of R-groups different from that observed in a non-mutated synthase.

In the present invention, the first domain forms a twisted α -barrel made up of eight short (10 to 15 amino acid residues) helices surrounding a surface cavity filled by ordered water molecules when hydrated. The second domain comprises a two-layered barrel of α -helices surrounding a hydrophobic and aromatic-rich active site pocket. Typically, the second domain contains a substrate binding site. As exemplified in FIG. 1, helix H is disrupted between segments H1 and H2 by an amino acid such as proline, but its interhelical packing with helix G is accommodated by a corresponding kink in helix G between G1 and G2. Within this kink, hydrogen bonds between a hydroxyl group, such as that found on a threonine, and the carbonyl oxygen of other amino acids disrupt the main chain intrahelical hydrogen bonding of helix G thus assisting in producing the structure as determined.

As exemplified by TEAS, terpene synthases of the present invention can have a first domain segment comprising helices A and C (an A-C loop), and a second domain comprising helices J and K (a J-K loop) (FIG. 1). The ordering of these loops upon substrate binding results in a closed, solvent-inaccessible active site pocket. As the J-K loop becomes ordered, a lid-type structure is formed that clamps down over the active site entrance in the presence of substrate and an extended aromatic patch deep within the active site pocket is formed. As the A-C loop becomes ordered, it translates inward toward the active site, positioning certain R groups in this loop at or near the active site. Thus, substrate binding to the active site results in a change in protein conformation.

To identify or create mutant terpene synthases, sequence alignments can be performed to locate specific residues and α -carbons in a preselected polypeptide that have the structural coordinates and interatomic distances of Tables 1–2, 3–4 or 5–6. The preselected polypeptide is used as the subject sequence in the alignment, e.g., the full-length primary amino acid sequence, a region 190 residues in length, a region 220 residues in length, or a region 300 residues in length. The alignment can use residues 265 to 535 of TEAS (SEQ ID NO: 2), which includes the α -carbons of Tables 1, 3 or 5, as the query sequence to align with the preselected polypeptide. The preselected polypeptide and the query sequence can be aligned using the BLASTp 2.0.9 computer program with a BLOSUM 62 scoring matrix, an expect value of 10, a gap open value of 11, an $x_dropoff$ value of 50, a gap extension value of 1, a wordsize of 3 and no filtering of low complexity sequences. As an alternative, the BLASTp 2.0.9 program can be used with a BLOSUM 50 scoring matrix, an expect value of 10, a gap open value 13, an $x_dropoff$ value of 50, a gap extension value of 2, a wordsize of 3 and no filtering of low complexity sequences. Other parameter values can also be used, e.g., a gap extension value from 0 to 4. See Altschul, et al., Nucl. Acids Res. 25:3389–3402.

Regions of the preselected polypeptide with significant sequence identity to residues 265–535 of TEAS, e.g., 20% or greater sequence identity, 25% or greater sequence identity, 35% or greater sequence identity, 40% or greater sequence identity, 50% or greater sequence identity, 60% or greater sequence identity, 70% or greater sequence identity, or 80% or greater sequence identity are examined for specific residues that align with the TEAS residues corresponding to those listed in Tables 1, 3, or 5. In some cases, the output of the computer program alignment identifies a specific residue in the preselected polypeptide for each of the nine, sixteen, or nineteen residues in the query sequence having the structural coordinates and interatomic distances of Tables 1–2, 3–4 or 5–6, with or without gaps introduced by the alignment program. In other cases, a gap is introduced by the alignment program in either the query sequence or the subject sequence such that no direct alignment or a misalignment occurs between one or more of the nine, sixteen, or nineteen residues in the query sequence that are of interest. In either case, the output can be visually inspected, and specific residues can be chosen in the subject sequence after adjusting the alignment so that alpha-helices and beta-sheet regions in the query sequence are maintained and that gaps or insertions in the subject sequence align with loop regions of the query sentence.

Sequence alignments suggest that other terpene synthases have regions with 20% or greater sequence identity to residues 265–535 of TEAS. Therefore, a region of a terpene synthase other than TEAS can be used as the query sequence, e.g., regions of terpene synthases given in SEQ ID NOS: 4, 6, 8, 10, 12, 20, 22, 24, 26, 28, 30, 32, 34–40, 42, 44, 46, 48, 50, 52, 54, 56, or 58, that have significant sequence identity to residues 265–535 of SEQ ID NO: 2. For example, large sequence insertions are present at the amino terminus in taxadiene synthase (SEQ ID NO: 44) with respect to TEAS, or are within solvent-exposed loops in the amino-terminal domain. Thus, regions of taxadiene synthase with greater than 20% sequence identity to SEQ ID NO: 2 are closer to the carboxy-terminal end, e.g., from residue 579 to residue 847 of SEQ ID NO: 44.

Useful regions of other terpene synthases that can be used as the query sequence include, without limitation, residues 343 to 606 of SEQ ID NO: 20, 316 to 586 of SEQ ID NO:

22, residues 352 to 622 of SEQ ID NO: 58, residues 272 to 540 encoded by SEQ ID NO: 33, residues 319 to 571 of SEQ ID NO: 42, residues 579 to 847 of SEQ ID NO: 44, residues 495 to 767 of SEQ ID NO: 46, residues 295 to 564 of SEQ ID NO: 48, residues 307 to 578 of SEQ ID NO: 50, residues 264 to 533 of SEQ ID NO: 52, residues 585 to 853 of SEQ ID NO: 56, residues 307 to 574 of SEQ ID NO: 54, residues 309 to 577 of SEQ ID NO: 24, residues 315 to 554 of SEQ ID NO: 26, residues 265 to 536 of SEQ ID NO: 28, residues 342 to 612 of SEQ ID NO: 30 and residues 273 to 541 of SEQ ID NO: 32.

One or more of the specific residues in the subject sequence that align with residues in the query sequence are mutated in the preselected polypeptide, e.g., by making mutations in a nucleic acid encoding the polypeptide. The mutant terpene synthase thus created can then be expressed in a host cell and the protein evaluated for enzymatic activity, if desired.

Mutant proteins of the present invention may be prepared in a number of ways including but not limited to oligonucleotide-directed mutagenesis, deletion, chemical mutagenesis, and the like. One or more R-groups associated with the active site α -carbon atoms in a terpene synthase are changed by altering the nucleotide sequence of the corresponding gene. For example, a mutation can be introduced into SEQ ID NO:1, the nucleotide sequence for TEAS, at codons encoding one or more of the following sixteen α -carbons: α -carbon 1=Cys 270; α -carbon 2=Trp 273; α -carbon 3=Ile 294; α -carbon 4=Ile 297; α -carbon 5=Ser298; α -carbon 6=Thr 402; α -carbon 7=Thr 403; α -carbon 8=Tyr 404; α -carbon 9=Leu 407; α -carbon 10=Cys 440; α -carbon 11=Val 516; α -carbon 12=Thr 519; α -carbon 13=Tyr 520; α -carbon 14=Asp 525; α -carbon 15=Tyr 527; or α -carbon 16=Thr 528. The protein encoded by the mutant gene is then produced by expressing the gene in, for example, a bacterial or plant expression system. Alternatively, synthase mutants may be generated by site specific replacement of a particular amino acid with an unnaturally occurring amino acid. As such, synthase mutants may be generated through replacement of an amino acid residue or a particular cysteine or methionine residue with selenocysteine or selenomethionine. This may be achieved by growing a host organism capable of expressing either the wild-type or mutant polypeptide on a growth medium depleted of natural cysteine or methionine or both and growing on medium enriched with either selenocysteine, selenomethionine, or both. These and similar techniques are described in Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd Ed. (1989) Cold Spring Harbor Laboratory Press).

Another suitable method of creating mutant synthases of the present invention is based on a procedure described in Noel and Tsal (1989) J. Cell. Biochem., 40:309–320. In so doing, the nucleic acid encoding the synthase can be synthetically produced using oligonucleotides having overlapping regions, the oligonucleotides being degenerate at specific bases so that mutations are induced.

According to the present invention, nucleic acid sequences encoding a mutated synthase can be produced by the methods described herein, or any alternative methods available to the skilled artisan. In designing the nucleic acid sequence (gene) of interest, it may be desirable to reengineer the gene for improved expression in a particular expression system. For example, it has been shown that many bacterially derived genes do not express well in plant systems. In some cases, plant-derived genes do not express well in bacteria. This phenomenon may be due to the non-optimal

G+C content or A+T content of the gene relative to the expression system being used. For example, the very low G+C content of many bacterial genes results in the generation of sequences mimicking or duplicating plant gene control sequences that are highly A+T rich. The presence of A+T rich sequences within the genes introduced into plants (e.g., TATA box regions normally found in gene promoters) may result in aberrant transcription of the gene(s). In addition, the presence of other regulatory sequences residing in the transcribed mRNA (e.g. polyadenylation signal sequences (AAUAAA) or sequences complementary to small nuclear RNAs involved in pre-mRNA splicing) may lead to RNA instability. Therefore, one goal in the design of genes is to generate nucleic acid sequences that have a G+C content that affords mRNA stability and translation accuracy for a particular expression system.

Due to the plasticity afforded by the redundancy of the genetic code (i.e., some amino acids are specified by more than one codon), evolution of the genomes of different organisms or classes of organisms has resulted in differential usage of redundant codons. This “codon bias” is reflected in the mean base composition of protein coding regions. For example, organisms with relatively low G+C contents utilize codons having A or T in the third position of redundant codons, whereas those having higher G+C contents utilize codons having G or C in the third position. Therefore, in reengineering genes for expression, one may wish to determine the codon bias of the organism in which the gene is to be expressed. Looking at the usage of the codons as determined for genes of a particular organism deposited in GenBank can provide this information. After determining the bias thereof, the new gene sequence can be analyzed for restriction enzyme sites as well as other sites that could affect transcription such as exon:intron junctions, polyA addition signals, or RNA polymerase termination signals.

Genes encoding synthases can be placed in an appropriate vector, depending on the artisan’s interest, and can be expressed using a suitable expression system. An expression vector, as is well known in the art, typically includes elements that permit replication of said vector within the host cell and may contain one or more phenotypic markers for selection of cells containing said gene. The expression vector will typically contain sequences that control expression such as promoter sequences, ribosome binding sites, and translational initiation and termination sequences. Expression vectors may also contain elements such as subgenomic promoters, a repressor gene or various activator genes. The artisan may also choose to include nucleic acid sequences that result in secretion of the gene product, movement of said product to a particular organelle such as a plant plastid (see U.S. Pat. Nos. 4,762,785; 5,451,513 and 5,545,817), or other sequences that increase the ease of peptide purification, such as an affinity tag.

A wide variety of expression control sequences are useful in expressing mutated synthases when operably linked thereto. Such expression control sequences include, for example, the early and late promoters of SV40 for animal cells, the lac system, the trp system, major operator and promoter systems of phage λ , and the control regions of coat proteins, particularly those from RNA viruses in plants. In *E. coli*, a useful transcriptional control sequence is the T7 RNA polymerase binding promoter, which can be incorporated into a pET vector as described by Studier et al., (1990) Methods Enzymology, 185:60–89.

For expression, a desired gene should be operably linked to the expression control sequence and maintain the appropriate reading frame to permit production of the desired

synthase. Any of a wide variety of well-known expression vectors are of use in the present invention. These include, for example, vectors consisting of segments of chromosomal, non-chromosomal and synthetic DNA sequences such as those derived from SV40, bacterial plasmids (including those from *E. coli* such as col E1, pCR1, pBR322 and derivatives thereof, pMB9), wider host range plasmids such as RP4, phage DNA such as phage λ , NM989, M13, and other such systems as described by Sambrook et al., (Molecular Cloning, A Laboratory Manual, 2nd Ed. (1989) Cold Spring Harbor Laboratory Press).

A wide variety of host cells are available for expressing synthase mutants of the present invention. Such host cells include, without limitation, bacteria such as *E. coli*, *Bacillus* and *Streptomyces*, fungi, yeast, animal cells, plant cells, insect cells, and the like. Preferred embodiments of the present invention include terpene synthase mutants that are expressed in *E. coli* or in plant cells. Said plant cells can either be in suspension culture or a culture on a solid support such as an agar-based medium.

Genes encoding synthases of the present invention can also be expressed in transgenic plant cells. In order to produce transgenic plants, vectors containing a nucleic acid construct encoding a mutant terpene synthase are inserted into the plant genome. Preferably, these recombinant vectors are capable of stable integration into the plant genome. One variable in making a transgenic plant is the choice of a selectable marker gene. A selectable marker gene is used to identify transformed cells against a high background of untransformed cells. Such selectable marker genes include but are not limited to aminoglycoside phosphotransferase gene of transposon Tn5 (Aph II) which encodes resistance to the antibiotics kanamycin, neomycin, and G418, as well as those genes which encode for resistance or tolerance to glyphosate, hygromycin, methotrexate, phosphinothricin, imidazolinones, sulfonylureas, and triazolopyrimidine herbicides, such as chlorosulfuron, bromoxynil, dalapon and the like. In addition to a selectable marker gene, it may be desirable to use a reporter gene. In some instances a reporter gene may be used with a selectable marker. Reporter genes allow the detection of transformed cell and may be used at the discretion of the artisan. A list of these reporter genes is provided in K. Weising et al., 1988, Ann. Rev. Genetics, 22:421.

The genes are expressed either by promoters expressing in all tissues at all times (constitutive promoters), by promoters expressing in specific tissues (tissue-specific promoters), promoters expressing at specific stages of development (developmental promoters), and/or promoter expression in response to a stimulus or stimuli (inducible promoters). The choice of these is at the discretion of the artisan.

Several techniques exist for introducing foreign genes into plant cells, and for obtaining plants that stably maintain and express the introduced gene. Such techniques include acceleration of genetic material coated directly into cells (U.S. Pat. No. 4,945,050). Plant may also be transformed using *Agrobacterium* technology (U.S. Pat. Nos. 5,177,010, 5,104,310, 5,149,645, 5,469,976, 5,464,763, 4,940,838, 4,693,976, 5,591,616, 5,231,019, 5,463,174, 4,762,785, 5,004,863, and 5,159,135; European Patent Applications 116718, 290799, 320500, 604662, 627752, 0267159, and 0292435. Other transformation technologies include whiskers technology, see U.S. Pat. Nos. 5,302,523 and 5,464,765. Electroporation technology has also been used to transform plants, see WO 87/06614, WO 92109696 and WO 93/21335 and U.S. Pat. Nos. 5,472,869 and 5,384,253. Viral vector expression systems can also be used such as those

described in U.S. Pat. Nos. 5,316,931, 5,589,367, 5,811,653, and 5,866,785.

In addition to numerous technologies for transforming plants, the type of tissue that is contacted with the genes of interest may vary as well. Suitable tissue includes, but is not limited to, embryogenic tissue, callus tissue, hypocotyl, meristem and the like. Almost all plant tissues may be transformed during dedifferentiation using the appropriate techniques described herein.

Regardless of the transformation system used, a gene encoding a mutant synthase is preferably incorporated into a gene transfer vector adapted to express said gene in a plant cell by including in the vector an expression control sequence (plant promoter regulatory element). In addition to plant promoter regulatory elements, promoter regulatory elements from a variety of sources can be used efficiently in plant cells to express foreign genes. For example, promoter regulatory elements of bacterial origin, such as the octopine synthase promoter, the nopaline synthase promoter, the mannopine synthase promoter may be used. Promoters of viral origin, such as the cauliflower mosaic virus (35S and 19S) are also desirable. Plant promoter regulatory elements also include, but are not limited to, ribulose-1,6-bisphosphate carboxylase small subunit promoter, beta-conglycinin promoter, phaseolin promoter, ADH promoter, heat-shock promoters, and tissue specific promoters and the like. Numerous promoters are available to skilled artisans for use at their discretion.

It should be understood that not all expression vectors and expression systems function in the same way to express the mutated gene sequences of the present invention. Neither do all host cells function equally well with the same expression system. However, one skilled in the art may make a selection among these vectors, expression control sequences, and host without undue experimentation and without departing from the scope of this invention.

Once a synthase of the present invention is expressed, the protein obtained therefrom can be purified so that structural analysis, modeling, and/or biochemical analysis can be performed, as exemplified herein. The nature of the protein obtained can be dependent on the expression system used. For example, genes, when expressed in mammalian or other eukaryotic cells, may contain latent signal sequences that may result in glycosylation, phosphorylation, or other post-translational modifications, which may or may not alter function. Once the proteins are expressed, they can be easily isolated and purified using techniques common to the person having ordinary skill in the art of protein biochemistry and as described in Colligan et al., (1997) Current Protocols in Protein Science, Chanda, V. B., Ed., John Wiley & Sons, Inc. Such techniques often include the use of cation-exchange or anion-exchange chromatography, gel filtration-size exclusion chromatography, and the like. Another technique that may be commonly used is affinity chromatography. Affinity chromatography can include the use of antibodies, substrate analogs, or histidine residues (His-tag technology).

Once purified, mutants of the present invention may be characterized by any of several different properties. For example, such mutants may have altered active site surface charges of one or more charge units. In addition, the mutants may have an altered substrate specificity or spectrum of reaction product relative to a non-mutated synthase.

The present invention allows for the characterization of mutant terpene synthase by crystallization followed by X-ray diffraction. Polypeptide crystallization occurs in solutions where the polypeptide concentration exceeds its solu-

bility maximum (i.e., the polypeptide solution is supersaturated). Such solutions may be restored to equilibrium by reducing the polypeptide concentration, preferably through precipitation of the polypeptide crystals. Often polypeptides may be induced to crystallize from supersaturated solutions by adding agents that alter the polypeptide surface charges or perturb the interaction between the polypeptide and bulk water to promote associations that lead to crystallization.

Compounds known as "precipitants" are often used to decrease the solubility of the polypeptide in a concentrated solution by forming an energetically unfavorable precipitating depleted layer around the polypeptide molecules (Weber, 1991, *Advances in Protein Chemistry*, 41:1-36). In addition to precipitants, other materials are sometimes added to the polypeptide crystallization solution. These include buffers to adjust the pH of the solution and salts to reduce the solubility of the polypeptide. Various precipitants are known in the art and include the following: ethanol, 3-ethyl-20-4 pentanediol, and many of the polyglycols, such as polyethylene glycol.

Commonly used polypeptide crystallization methods include the following techniques: batch, hanging drop, seed initiation, and dialysis. In each of these methods, it is important to promote continued crystallization after nucleation by maintaining a supersaturated solution. In the batch method, polypeptide is mixed with precipitants to achieve supersaturation, the vessel is sealed and set aside until crystals appear. In the dialysis method, polypeptide is retained in a sealed dialysis membrane that is placed into a solution containing precipitant. Equilibration across the membrane increases the polypeptide and precipitant concentrations thereby causing the polypeptide to reach supersaturation levels.

In the preferred hanging drop technique (McPherson, 1976, *J. Biol. Chem.*, 6300-6306), an initial polypeptide mixture is created by adding a precipitant to a concentrated polypeptide solution. The concentrations of the polypeptide and precipitants are such that in this initial form, the polypeptide does not crystallize. A small drop of this mixture is placed on a glass slide that is inverted and suspended over a reservoir of a second solution. The system is then sealed. Typically, the second solution contains a higher concentration of precipitant or other dehydrating agent. The difference in the precipitant concentrations causes the protein solution to have a higher vapor pressure than the solution. Since the system containing the two solutions is sealed, an equilibrium is established, and water from the polypeptide mixture transfers to the second solution. This equilibrium increases the polypeptide and precipitant concentration in the polypeptide solution. At the critical concentration of polypeptide and precipitant, a crystal of the polypeptide may form.

Another method of crystallization introduces a nucleation site into a concentrated polypeptide solution. Generally, a concentrated polypeptide solution is prepared and a seed crystal of the polypeptide is introduced into this solution. If the concentration of the polypeptide and any precipitants are correct, the seed crystal will provide a nucleation site around which larger crystal forms. In preferred embodiments, the crystals of the present invention are formed in hanging drops with 15% PEG 8000; 200 mM magnesium acetate or magnesium chloride, 100 mM 3-(N-morpholino)-2-hydroxypropanesulfonic acid (pH 7.0), 1 mM dithiothreitol as precipitant.

Some proteins may be recalcitrant to crystallization. However, several techniques are available to the skilled

artisan to induce crystallization. The removal of polypeptide segments at the amino or carboxyl terminal end of the protein may facilitate production of crystalline protein samples. Removal of such segments can be done using molecular biology techniques or treatment of the protein with proteases such as trypsin, chymotrypsin, subtilisin. Such procedures can result in the removal of flexible polypeptide segments that may negatively affect crystallization.

The crystals so produced have a wide range of uses. For example, high quality crystals are suitable for X-ray or neutron diffraction analysis to determine the three-dimensional structure of a mutant synthase and to design additional mutants thereof. In addition, crystallization can serve as a further purification method. In some instances, a polypeptide or protein will crystallize from a heterogeneous mixture into crystals. Isolation of such crystals by filtration, centrifugation, etc., followed by redissolving the polypeptide affords a purified solution suitable for use in growing the high-quality crystals needed for diffraction studies. The high-quality crystals may also be dissolved in water and then formulated to provide an aqueous solution having other uses as desired.

Because synthases may crystallize in more than one crystal form, the structural coordinates of (carbons of an active site determined from a synthase or portions thereof, as provided by this invention, are particularly useful to solve the structure of other crystal forms of synthases. The structural coordinates, as provided herein, may also be used to solve the structure of synthases having α -carbons position within the active sites in a manner similar to the wild-type yet having R-groups that may or may not be identical. Furthermore, the structural coordinates disclosed herein may be used to determine the structure of the crystalline form of other proteins with significant amino acid or structural homology to any functional domain of a synthase. One method that may be employed for such purpose is molecular replacement. In this method, the unknown crystal structure, whether it is another crystal form of a synthase, a synthase having a mutated active site, or the crystal of some other protein with significant sequence identity and/or structural homology of a synthase may be determined using the coordinates given in Tables 10 and/or 11. This method provides sufficient structural form for the unknown crystal more efficiently than attempting to determine such information *ab initio*. In addition, this method can be used to determine whether or not a given synthase in question falls within the scope of this invention.

As further disclosed herein, synthases and mutants thereof may be crystallized in the presence or absence of substrates and substrate analogs. The crystal structures of a series of complexes may then be solved by molecular replacement and compared to that of the wild-type to assist in determination of suitable replacements for R-groups within the active site, thus making synthase mutants according to the present invention.

All mutants of the present inventions may be modeled using the information disclosed herein without necessarily having to crystallize and solve the structure for each and every mutant. For example, one skilled in the art may use one of several specialized computer programs to assist in the process of designing synthases having mutated active sites. Examples of such programs can be as follows: GRID (Goodford, 1985, *J. Med. Chem.*, 28:849-857); MCSS (Miranker and Karplus, 1991, *Proteins: Structure, Function and Genetics*, 11:29-34); AUTODOCK (Goodsell and Olsen, 1990, *Proteins: Structure, Function, and Genetics*,

8:195–202); and DOCK (Kuntz et al., 1982, *J. Mol. Biol.*, 161:269–288). In addition, specific computer programs are also available to evaluate specific substrate-active site interactions and the deformation energies and electrostatic interactions resulting therefrom. MODELLER is a computer program often used for homology or comparative modeling of the three-dimensional structure of a protein. A. Sali & T. L. Blundell. *J. Mol. Biol.* 234, 779–815, 1993. A preselected polypeptide sequence to be modeled is aligned with one or more terpene synthases whose crystal structures are known and the MODELLER program is used to calculate a full-atom model, based on optimum satisfaction of spatial restraints. Such restraints can include, inter alia, homologous structures, fluorescence spectroscopy, NMR experiments, or atom-atom potentials of mean force.

The present invention enables synthase mutants to be made and crystal structures thereof to be solved. Moreover, by virtue of the present invention, the location of the active site and the interface of substrate therewith permit the identification of desirable R-groups for mutagenesis. The particular embodiments of this invention are further exemplified in the Examples. However, those skilled in the art will readily appreciate that the specific experiments detailed are only illustrative of the invention as described more fully in the claims, which follow thereafter.

EXAMPLE 1

Generation of Mutant Teas Genes

Construct Generation and Expression.

All mutant enzymes were constructed by the Quick-Change method (Stratagene). Manufacturers instructions were followed, except as noted. Mutations were confirmed by DNA sequencing, and plasmids containing the desired mutation were used to transform BL-21 (DE3) expression cells. Protein was expressed, purified, and stored at -80°C .

TEAS W273S.

The TEAS W273S mutant was generated from a TEAS-pET28b(+) template using the following primers:

GTTGAATGCTACTTTTTCGGCATTAGGAGTTTAT (sense) (SEQ ID NO:13) and ATAAACTCCTMTGCGAAAAGTAGCATTCAAC (antisense) (SEQ ID NO:14). Mutagenesis was carried out according to the manufacturer's instructions, except that sense and anti-

sense strands were generated in separate reactions. For each, 30 plasmid-copying cycles of one minute, annealing at 55°C . and 16 minutes extension at 68°C . were carried out. The two reaction mixtures were then combined, heated to 95°C . for 2.5 minutes, and cooled to room temperature before DpnI treatment.

TEAS C440W

The TEAS C440W mutant was generated from the TEAS-pET28b(+) template using the following primers:

GCTAGTGTAATTATATGGCGAGTTATCGATGAC (sense) (SEQ ID NO:15) and GTCATCGATMCTCGCATATMTTACTAGC (antisense) (SEQ ID NO:16).

TEAS W273SIC440W

The TEAS C440WMW273S mutant was constructed from a TEAS W273S-pET28b(+) template using the primers described for generation of TEAS C440W.

TEAS 406/407 Random Library.

For generation of a library of TEAS mutants with random amino acids at positions 406 and 407, two 50 microliter QuickChange reactions were carried out with the TEAS-pET28b(+) template and the primers

GCACTAGCAACTACCATAT-TACNNSNNSGCGACAACATCGTATTTGGGCATG (sense) (SEQ ID NO:17) and CATGCCAAATACGATGTTGTCGCSNNSNNGTAATATGTGG-TAGTTGCTAGTGC (antisense) (SEQ ID NO:18), in which N denotes A, C, G, or T and S denotes C or G.

By this choice of nucleotides, the reaction included primers which coded for all possible amino acid combinations at positions 406 and 407. No adjustment was made for differing numbers of codons among amino acids. In order to ensure efficient reactions, and to minimize the preference for hybridization of wild-type primers to the template, the primers were designed to be longer than those used to generate the mutations described above. In addition, they were HPLC purified prior to use. After 18 cycles of plasmid copying, the reaction was incubated for two hours with DpnI, ethanol precipitated, and redissolved in 5 microliters water. Each of four 40 microliter aliquots of *E. coli* NovaBlue (Novagen) cells were electroporated with 1.5 microliters of the redissolved DNA. After a recovery period, the cells were plated on kanamycin-LB-agar plates. In order to transfer the newly constructed plasmids to expression cells, the colonies were scraped from all four plates, and used to start an 8 mL culture grown in liquid LB medium at 37°C . for 8 hours. Plasmid purified from this culture was used to transform 20 microliters of competent BL-21 (DE3) cells.

For storage of the constructs, each individual colony was used to inoculate 100 microliters of LB medium containing kanamycin (50 micrograms/mL) in 96-well culture plates. The cells were grown at 37°C . until the A_{600} reached approximately one; 100 microliters of 30% glycerol in LB were then added, and the plates were frozen at -80°C . A set of randomly selected colonies were grown from individual glycerol stocks of some colonies, and plasmids were extracted for sequencing. Approximately 30 percent of the colonies were found to be wild-type. Nucleotide and amino acid sequences for TEAS 406/407 mutant genes and proteins are shown in SEQ ID NOS:11 and 12.

TEAS Y520F

The tyrosine residue at position 520 of SEQ ID NO:2 was changed to a phenylalanine residue by site-directed mutagenesis with primers, in a manner similar to that described above. For Y520F the TAT codon was changed to TTC. The nucleotide sequence of the mutant gene is shown in SEQ ID NO:5.

TEAS Y527F

The tyrosine residue at position 527 of SEQ ID NO:2 was changed to a phenylalanine residue by site-directed mutagenesis with primers, in a manner similar to that described above. For Y527F, the TAC codon at position 527 of the TEAS amino acid sequence was changed to TTC. The nucleotide sequence of the mutant TEAS Y527F gene is shown in SEQ ID No: 7.

TEAS W273E

The tryptophan residue at position 273 of SEQ ID NO:2 was changed to a phenylalanine residue by site-directed mutagenesis with primers, in a manner similar to that described above. For W273E, the TGG codon at position 273 of the TEAS amino acid sequence was changed to GAG. The nucleotide sequence of the mutant gene is shown in SEQ ID NO.:3.

EXAMPLE 2

Expression and Isolation of Synthase Polypeptides

Unless otherwise noted, mutated and non-mutated TEAS proteins were expressed in *Escherichia coli*, purified by metal chelation, anion exchange, and gel filtration chromatography.

Constructs of TEAS and mutant TEAS proteins in the vector pET-28b(+) (Novagen) were expressed in *E. coli* cells. For a typical protein preparation of any of these enzymes, *E. coli* strain BL21 (DE3) cells containing the plasmid construct were grown at 37° C. in 4x1 L terrific broth to an $A_{600}=1.0$. The temperature was dropped to 22° C., and protein expression was induced by adding IPTG to a final concentration of 0.1 mM. After 15–20 h, the cells were harvested by centrifugation, resuspended in 5 mL buffer A (20 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 7.9) per 1 g cells (wet weight), and stirred for 0.5 h at 4° C. The cells were then lysed by sonication, and the resulting lysate was centrifuged for 0.7 h at 82,000x g. The supernatant, containing the protein, was loaded over a 2–3 mL Ni^{2+} chelating histidine affinity column (Qiagen) equilibrated in buffer A, and the column was washed with additional buffer A until the A_{280} of the eluent returned to baseline. The protein was then eluted with a 20–200 mM imidazole gradient in buffer A. Protein-containing fractions were pooled and dialyzed against buffer B (50 mM HEPES, 5 mM $MgCl_2$, 1 mM DTT), then loaded onto an 8 mL MonoQ cation-exchange column (Pharmacia). The column was washed with 20 column volumes buffer B, and the protein was eluted with a 0–500 mM NaCl gradient in buffer B. The resulting protein was further purified by gel filtration on a Superdex-200 column (Pharmacia) in 50 mM Tris, 100 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, pH 8.0. Purified protein was then dialyzed against 5 mM Tris, 5 mM NaCl, 1 mM DTT, pH 8.0, concentrated to 18–22 mg/mL, and stored at –80° C. in 100° L aliquots until needed.

EXAMPLE 3

Crystallization and Structural Analysis of Synthase Polypeptides

Crystal Growth and Microseeding: All crystallization attempts were carried out by the hanging-drop vapor diffusion method. Concentrated protein was mixed with an equal volume (2–5 uL each) of reservoir solution on a plastic cover slip. The cover slip was then inverted over a well of a plastic 24-well tissue culture plate, containing 0.5–1.0 mL of reservoir solution, and sealed by a layer of vacuum grease between the well and cover slip. The plates were incubated at 4° C. while the protein concentration in the hanging drop slowly increased by vapor diffusion. Approximately 300 different reservoir solutions, ranging pH 4.5–9 with a variety of precipitants and added salts, were assayed for crystallization of TEAS (SEQ ID NO:2). TEAS crystallized with a reservoir solution of 15% PEG 8000, 100 mM MOPSO (3-[N-morpholino]-2-hydroxypropanesulfonic acid), 200 mM magnesium acetate, 1 mM DTT, pH 6.9–7.3. For microseeding, an existing crystal was crushed in a few uL of precipitant solution, then diluted to 50 microliters. After initial centrifugation to remove large particles, the suspension was serially diluted with additional precipitant solution, and a small volume of a diluted seed stock was added to each new crystallization drop. For macroseeding, crystals which were no longer rapidly growing (usually 2 weeks after drops were set up), were “rinsed” by serially transferring them through two to three drops of reservoir solution. The crystal was then transferred to a fresh drop containing protein and reservoir solution, and equilibrated against a reservoir solution as in the initial growth. Individual crystals varied in their degree of internal order. In some cases, several crystals were screened to identify a well-diffracting crystal with low mosaicity.

Data collection:

Prior to data collection, crystals were transferred to a drop of reservoir solution, or reservoir solution containing a compound to be soaked into the crystal. A small volume of cryoprotectant solution (15% PEG8000, 100 mM MOPSO, 200 mM Mg acetate, 20% ethylene glycol, 1 mM DTT, pH 7) was then added to the drop. After a short equilibration time (1–5 minutes), the crystal was transferred to a drop of cryoprotectant, or cryoprotectant with soaking compound added. After another short equilibration time, the crystal was picked up on a nylon loop, and quickly mounted for data collection in a stream of cold nitrogen gas (90–110K).

The TEAS crystals belonged to the tetragonal space group $P4_12_12$; the unit cell dimensions varied by a few angstroms between crystals, but on average $a=126$ Å, $c=122$ Å. The uncomplexed TEAS structure was initially refined to 2.8 Å (Table 11) against data collected from a crystal grown in the presence of 2 mM FHP (Table 10). Electron density at the active site allowed unambiguous modeling of FHP, the A-C and J-K loops, and nine additional residues at the NH₂ terminus. The refined TEAS-FHP model consisted of residues 17 to 548, three Mg^{2+} ions, 150 water molecules, and one FHP molecule. The three-dimensional coordinates for TEAS in the presence of bound substrate is shown in Table 10. The three-dimensional coordinates for TEAS in the absence of FHP is shown in Table 11.

Crystals of TEAS complexed with trifluoro-farnesyl diphosphate (F3-FPP) were also prepared. In these crystals, a well-ordered diphosphate binding pocket was also observed. The A-C loop and the NH₂-terminal segment exhibited well-defined electron density, the A-C loop was translated toward the active site, and there was strong electron density for the diphosphate moiety of F3-FPP. The hydrophobic pocket, however, remained flexible; the J-K loop and the farnesyl moiety of F3-FPP were disordered.

Homology models were created and energy-minimized using the Swiss PDB viewer interface of the SwissModel program (Peitsch MC (1996), *Biochem. Soc. Trans.*, 24:274279 and Guex N. and Peitsch MC, 1997, *Electrophoresis.*, 18:2714–2723). Active site volumes were calculated with VOIDOO (Kleywegt, G. J., and Jones, T. A., *CCP4/ESF-EACBM Newsletter on Protein Crystallography.*, 29, 26–28, 1993). To make closed active site cavities, the energy-minimized diphosphate moiety from the modeled TEAS cyclase reaction was appended to the residue equivalent to TEAS D301.

TEAS W273S Crystal Structures.

Two TEAS W273S structures, in the presence of FHP, were determined from different crystals; both crystals appeared to be well ordered, as clear main-chain and side-chain density were apparent for residues throughout the protein, including the frequently mobile helices D1, D2, and E. Initial difference electron density maps from both crystals immediately revealed the W273S mutations. The two crystals were designated W273S-1 and W273S-2.

In each structure, the loops surrounding the active site were ordered, resulting in a closed active site pocket. The A/C loop in each structure was translated toward the active site, forming part of its outer rim, as observed in the wild-type TEAS/FHP complex. However, while the J/K loop of W273S-1 adopted the same conformation observed in the wild-type TEAS/FHP complex, the same loop in W273S-2 adopted a different conformation. In this conformation of the J/K loop, Tyr527 moved away from the side chain of residue 273. In addition, Tyr520 and Asp525 were placed distal to the side chain of Asp444. Hydrogen bonds previously observed between the J/K loop, Arg266, and the N-terminal loop were also missing in the W273S-2 structure.

The W273S-2 conformation does not appear to be an effect of the W273S mutation, as it was also observed in a wild-type TEAS crystal soaked with the epi-aristolochene mimic deoxycapsidiol, despite the fact that no electron density was readily apparent for the deoxycapsidiol molecule in that structure. Further, the TEAS active site loops were distant from crystal contacts, and their conformations were not likely to be artifacts of crystal packing. It is possible that at different stages of the TEAS reaction, the enzyme's J/K loop exists in different, defined conformations, and that each of these crystal structures has captured an image of a different conformation. In both W273S structures, residues other than Arg266 and those on the J/K loop did not undergo significant rearrangement from the conformations observed in wild-type TEAS.

In each W273S crystal structure, electron density in the active site suggested that the substrate mimic FHP binds in multiple conformations. Some regions of this density possibly represented bound water molecules in the mutant active site. The presence of water molecules in the mutant active site is consistent with the observation that TEAS W273S gives rise to multiple hydroxylated terpenoid products.

The FHP electron density in each W273S crystal structure was sufficient to suggest that FHP existed in a more extended conformation in the W273S structure, compared to the more tightly folded conformation of FHP in the wild-type TEAS/FHP complex. The observation that the active site of W273S binds multiple conformations of FHP is consistent with the fact that W273S converts FPP to multiple terpenoid hydrocarbon products.

TEAS C440W/W273S: TEAS C440W/W273S crystallized under conditions identical to wild-type TEAS. A 0.3 mm crystal was soaked for 20 minutes in reservoir solution saturated in farnesyl hydroxy phosphate (FHP). After cryoprotection and flash freezing as described for wild-type TEAS, data were collected on a laboratory source with Cu-K α radiation (MacScience Corp., Japan). A starting model of uncomplexed TEAS (Table 11) (Brookhaven Protein Database Code 5EAT (PDB 5EAT), with waters and magnesiums removed, was positioned against the mutant data with the rigid body module of the software program X-PLOR (A. T. Brunge, X-PLOR Version 3.1—A System for X-Ray Crystallography and NMR Yale University Press, New Haven, 1992, pp. 187–207). Rounds of positional and restrained b-factor refinement with bulk solvent modeling were also carried out in X-PLOR, with manual model building and adjustment carried out in the software program O (Jones, T A, Zou, J Y, Cowan, S W, and Kjeldgaard, M., *Acta Cryst. D.*, 49:148–157, 1993). Additional rounds of refinement and map calculation using the CNS program suite resulted in significantly improved maps; this improvement was likely due to improved bulk solvent modeling.

TEAS C440W:

TEAS C440W crystallized under conditions identical to wild-type TEAS, except that crystals nucleated less readily and were generally smaller. A mutant crystal was soaked for 6 hours in reservoir solution saturated in FHP before flash-freezing and data collection at SSRL beamline 7-1 (Stanford Synchrotron Radiation Laboratory, Menlo Park, Calif.). A starting model of TEAS-FHP (Table 10), with water molecules, ligands, and residues 523–532 of SEQ ID NO:2 removed, was positioned against the data with the rigid body module of X-PLOR. Rounds of positional and restrained b-factor refinement with bulk solvent and overall anisotropic temperature factor modeling were also carried out in

X-PLOR, and manual model building and adjustment were carried out in the software program O. As with the double mutant, electron density maps were noticeably improved after refinement and map calculation in CNS.

EXAMPLE 4

Terpene Synthase Enzyme Assays

Synthase activity assays were carried out based on the assay described in Vogeli and Chappell, *Plant Physiol.* 94:1860 (1990) and Vogeli, et al., *Plant Physiol.* 93:182 (1990). In general, radio-labeled (^3H or ^{14}C) substrate was incubated with enzyme at room temperature in a buffered magnesium salt solution (200 mM Tris, pH 8, 50 mM Mg chloride, 1 mM DTT, unless otherwise noted); hydrocarbon products were then selectively extracted into an organic solvent such as hexane. The hexane extract generally was treated with silica gel to remove prenol alcohols and other oxygenated compounds generated by non-enzymatic hydrolysis of substrate, which partition inefficiently into hexane. Hydrocarbon products present in the hexane phase were quantitated by scintillation counting.

A subsequent extraction with a more polar organic solvent such as ethyl acetate was sometimes carried out. Oxygenated compounds more efficiently partition into ethyl acetate-type solvents. Compounds present in the ethyl acetate phase were also quantitated by scintillation counting.

Substrate concentrations typically ranged from 0.1 nanomolar to 100 micromolar. In some assays, the substrate was not radiolabeled. Reactions generally were carried out in triplicate for each substrate concentration. Protein concentration was determined by the Bradford method. For determination of steady-state kinetic parameters, enzyme concentrations were chosen such that generation of products over time was linear throughout the course of the reaction.

Diterpene synthase assays typically were carried out using ^3H geranylgeranyl diphosphate (GGPP) and enzyme in 250 mM Tris, 10 mM Mg chloride, 1 mM DTT, pH 8.0. Sesquiterpene synthase assays typically were carried out using ^{14}C or ^3H FPP and enzyme in 100 mM Tris, 30 mM Mg chloride, 1 mM DTT, pH 8.0. Monoterpene synthase assays typically were carried out using ^3H GPP and enzyme. As a control for nonspecific binding of GPP by protein, identical reactions were set up which contained BSA, rather than enzyme.

Product analysis of wild type and mutant TEAS enzymes by Ag-TLC.

Terpenoid hydrocarbon products are not readily separated by thin layer chromatography on normal or reverse-phase plates; however, some can be separated by argentation TLC (Ag-TLC), in which silica plates are first treated with silver nitrate. Ag-TLC described here generally followed the procedure described by Back et al., *Arch. Biochem. Biophys.* 315:527 (1994). A silica TLC plate was dipped in 15% silver nitrate (aqueous), then dried for 3–5 hours at 110° C. After spotting of tritiated enzymatic products (solvent extract), the plate was developed in benzene:hexane, ethyl acetate (50:50:1, by volume), sprayed with En 3 Hance (NEN) fluorography spray, placed on film, and exposed for several days to several weeks. Long exposure times were generally necessary, as silver-nitrate treatment of the TLC plate appeared to cause quenching of the fluorography reagent's fluorescence. Alternatively, ^{14}C labelled products were detected after one to two days without the use of fluorography spray.

33

EXAMPLE 5

Activity of TEAS W273S

Diterpene Synthase Activity of TEAS W273S.

The TEAS W273S enzyme and radiolabelled GGPP were incubated as described above and hydrocarbon products were extracted with hexane. Oxygenated products were then extracted with ethyl acetate. Reactions using wild-type TEAS gave counts lower than buffer alone. TEAS W273S, on the other hand, gave counts that were significantly higher for both the hexane and ethyl acetate extracts. Hydrocarbon products formed from GGPP by W273S were distinct from the products made by acid-catalyzed loss of diphosphates from GGPP. See FIG. 3.

Sesquiterpene Synthase Activity of TEAS W273S.

Products of FPP turnover by the purified TEAS W273S mutant were analyzed by argentation thin-layer chromatography (Ag-TLC). One major reaction product had an R_f of 0.7 by Ag-TLC, which was distinct from both 5-epi-aristolochene ($R_f=0.78$) and vetispiradiene ($R_f=0.63$). See FIG. 4. Preliminary GC/MS data showed that hexane extracts from FPP turnover by TEAS W273S contained at least four terpene hydrocarbons, with mass spectra distinct from either 5-epi-aristolochene or vetispiradiene. One of these products had a mass spectrum similar to germacrene A.

EXAMPLE 6

Activity of TEAS C440W/W273S

Diterpene Synthase Activity of TEAS C440W/W273S.

The mutant TEAS C440W/W273S protein contains a tryptophan residue at position 440 and a serine residue at position 273. Assays with GGPP were carried out using 0.5 micromolar ^3H GGPP, various concentrations of unlabelled GGPP (Echelon), and enzyme. Reactions were incubated for 60 minutes at room temperature. The TEAS C440W/W273S mutant protein converted GGPP to hexane-extractable products, whereas the wild-type enzyme did not. The results indicated that the product profile was altered compared to wild-type TEAS. Hexane-extractable products of GGPP turnover by the double mutant were analyzed by Ag-TLC. The products included two species ($R_f=0.11$ and 0.28) that were distinct from the hydrolysis product geranyl geraniol ($R_f=0.0$). To verify that products generated by TEAS C440W/W273S from GGPP were not the hydrolysis product, geranylgeraniol, a sample was analyzed by Ag-TLC. A reaction containing ^3H GGPP ($5\ \mu\text{m}$) and enzyme ($40\ \mu\text{m}$) in 100 microliters buffer was incubated overnight at room temperature. As controls, ^3H GGPP was incubated in reaction buffer alone and in reaction buffer adjusted to pH 1.5. Both the enzymatic and control reactions were extracted with hexane, which was spotted on an argentation TLC plate, and developed and exposed as described above. The results, shown in FIG. 3, demonstrated that the products formed by TEAS C440W/W273S were different from those generated by non-enzymatic degradation of geranylgeranyl diphosphate.

Sesquiterpene Synthase Activity of TEAS C440W/W273S.

Reactions with FPP as substrate were carried out with ^{14}C FPP ($9\ \mu\text{m}$) and enzyme ($160\ \mu\text{m}$) in reaction buffer ($20\ \mu\text{l}$). After incubating for 30 minutes at room temperature, products made by TEAS C440W/W273S were analyzed by Ag-TLC. The product profile of the double mutant was similar to that of 1 EAS W273S, with the addition of a major

34

product having an R_f of 0.57. The new product was distinct from both 5-epi-aristolochene and vetispiradiene. Several other products were also formed, many of which migrated slowly upon argentation TLC. See FIG. 4.

EXAMPLE 7

Activity of TEAS C440W

Diterpene Synthase Activity of TEAS C440W

Enzyme assays with TEAS C440W were carried out as described in Example 6. As shown in FIG. 3, no hexane-extractable products were detectable by Ag-TLC after an overnight incubation at room temperature with $160\ \mu\text{m}$ of enzyme and $9\ \mu\text{m}$ radiolabeled GGPP in $20\ \mu\text{l}$ volume.

Sesquiterpene Synthase Activity of TEAS C440W

Ag-TLC analysis of the products made from radiolabelled by purified TEAS C440W detected the formation of at least one major terpenoid hydrocarbon product ($R_f\ 0.63$) that was distinct from 6-epi-aristolochene ($R_f\ 0.78$) and vetispiradiene. The reactions product profile on Ag-TLC is shown in FIG. 4. Small amounts of slowly-migrating products ($R_f\ 0-0.09$) were also formed.

GC/MS analysis of the hexane extract of TEAS C440W terpenoid hydrocarbon reaction products confirmed that this mutant formed a single major sesquiterpene hydrocarbon product as well as a small number of minor hydroxylated products. The mass spectrum of the major product closely matched the published mass spectrum of the spirocyclic compound hinesene. Hinesene differs from vetispiradiene in the stereochemistry at the C3 methyl group.

EXAMPLE 8

Activity of TEAS W273E

Sesquiterpene Synthase Activity of TEAS W273E. Reactions to determine the products made by TEAS W273E using FPP as substrate were carried out essentially as described above, using radiolabeled FPP. The results indicated that at least one product other than 5-epi-aristolochene was formed. The results also indicated that alkylation of TEAS by FPP had occurred. The alkylation was dependent upon the presence of MgCl_2 in the reaction mixture. In control experiments, boiled W273E-TEAS, as well as wild-type TEAS and BSA, were not alkylated. These results indicate that alkylation had occurred at position 273 and that the amino acid residue at position 273 is part of the active site.

EXAMPLE 9

Activity of TEAS Y520F

Sesquiterpene Synthase Activity of TEAS Y520F.

Reactions with radiolabeled FPP and TEAS Y520F enzyme were carried out essentially as described above. Reaction products were analysed by Ag-TLC and by GC/MS. A major product of the TEAS Y520F reaction had the same GC retention time as authentic germacrene A and the same mass spectrum as authentic gernacrene A. The retention time and mass spectrum of this product were different from 5-epi-aristolochene.

EXAMPLE 10

Activity of TEAS Y527F

Enzymatic Activity of TEAS Y527F.

A crude extract of TEAS Y527F enzyme was made by inducing expression in *E. coli* cells, and sonicating the cells. The sonicate was clarified and the supernatant used for enzyme assays. No products were observed in assays using GPP as a substrate, indicating that TEAS Y527F does not have monoterpene synthase activity. Reaction products were obtained using FPP as a substrate. Analysis of these products by Ag-TLC indicated that products other than 5-epi-aristolochene were generated by the TEAS Y527F enzyme.

EXAMPLE 11

Alignment of Terpene Synthase Sequences

Residues 265 to 535 of the TEAS primary amino acid sequence (SEQ ID NO: 2) were aligned with the full-length amino acid sequence of a limonene synthase (SEQ ID NO: 22), using the BLASTp program (NCBI) with a BLOSUM 62 scoring matrix, a gap open value of 11, a gap extension value of 1, an x_dropoff value of 50, an expect value of 10, a wordsize of 3 and no filtering of low complexity sequences. The output of the alignment program, shown in Table 12, included a gap between residues 527 and 528 of the TEAS sequence (numbered as 263 and 264 in the alignment output). Residues 321, 324, 345, 348, 349, 427, 452, 453, 454, 455, 458, 492, 496, 569, 572, 573, 577, 579 and 580 were selected as having the most suitable alignment with the 19 TEAS residues. Residue 580 of limonene cyclase instead of residue 583 was selected as aligning with residue 528 of TEAS, in order to maintain the spatial orientation of structural aspects found in TEAS, i.e., α -helices, β -, sheets and loops shown in FIG. 1 and Table 10.

A region including residues 579 to 847 of the taxadiene primary amino acid sequence of SEQ ID NO: 44 was aligned with the full-length amino acid sequence of a bornyl diphosphate synthase (SEQ ID NO: 26), using the BLASTp program (NCBI) with a BLOSUM 62 scoring matrix, a gap open value of 11, a gap extension value of 1, an x_dropoff value of 50, an expect value of 10, a wordsize of 3 and no filtering of low complexity sequences. The output of the alignment program, shown in Table 13, included a gap between residues 453 and 454 of the bornyl diphosphate synthase sequence. Residues 321, 324, 344, 347, 348, 426, 451, 452, 453, 454, 457, 492, 496, 568, 571, 572, 576, 578 and 579 of the bornyl diphosphate synthase were selected as having the most suitable alignment with residues 584, 587, 606, 609, 610, 688, 713, 714, 715, 716, 719, 753, 757, 831, 834, 835, 839, 841 and 842 of the query region sequence of SEQ ID NO: 44. Residues 453 and 454 of bornyl diphosphate synthase were selected to align with residues 715 and 716 of taxadiene synthase, in order to maintain the spatial orientation of structural aspects expected to be present in taxadiene synthase, i.e., α -helices, β -sheets and loops shown in FIG. 1 and Table 10.

Residues 265 to 535 of the TEAS primary amino acid sequence (SEQ ID NO: 2) were aligned with the full-length amino acid sequence of a δ -selinene synthase (SEQ ID NO: 48), using the BLASTp program (NCBI) with a BLOSUM 50 scoring matrix, a gap open value of 13, a gap extension value of 2, an x_dropoff value of 50, an expect value of 10, a wordsize of 3 and no filtering of low complexity sequences. The output of the alignment program is shown in Table 14. Residues 300, 303, 324, 327, 328, 406, 431, 432, 433, 434, 437, 471, 475, 548, 551, 552, 556, 558 and 559 of SEQ ID NO:48 were selected as having the most suitable alignment with residues 270, 273, 294, 297, 298, 376, 401, 402, 403, 404, 407, 440, 444, 516, 519, 520, 525, 527 and 528 of SEQ ID NO, 2 Residues 307 to 593 of the primary amino acid sequence of γ -humulene synthase (SEQ ID NO: 50) were aligned with the full-length amino acid sequence of abietadiene synthase (SEQ ID NO: 56), using the BLASTp program (NCBI) with a BLOSUM 62 scoring matrix, a gap open value of 11, a gap extension value of 1, an x_dropoff value of 50, an expect value of 10, a wordsize of 3 and no filtering of low complexity sequences. The output of the alignment program is shown in Table 15. Residues 590, 593, 614, 617, 618, 696, 721, 722, 723, 724, 727, 761, 765, 837,

840, 841, 845, 847 and 848 of the diterpene synthase (SEQ ID NO: 56) were selected as having the most suitable alignment with residues 312, 315, 336, 339, 340, 419, 444, 445, 446, 447, 450, 484, 488, 562, 565, 566, 570, 572 and 573 of the sesquiterpene synthase query sequence (SEQ ID NO: 50).

EXAMPLE 12

Generation of Novel Monoterpene Synthase Genes

A DNA sequence encoding a pinene synthase (SEQ ID NO:20) is used to construct a library of mutant pinene synthase genes. Random mutations are introduced at nucleotides encoding one or more of the following nine amino acid residues: L, C, C, G, H, S, L, G and Y, which correspond to positions 351, 372, 480, 481, 482, 485, 519, 600 and 601 of SEQ ID NO:20.

In some cases, the pinene synthase coding sequence is randomly mutated at nucleotides encoding one or more of amino acid residues 348, 375, 376, 597, 605, 607 and 608, which correspond to positions Y, I, T, F, D, Y and S of SEQ ID NO:20. The pinene synthase coding sequence is sometimes mutated at nucleotides encoding one or more of the following amino acid residues: Y, S and G, which correspond to positions 454, 479 and 523 of SEQ ID NO:20. In some cases, mutations at these ten positions are made in addition to mutations at nucleotides encoding the nine residues mentioned above. In other cases, mutations at these ten positions are made without introducing mutations at the nine residues mentioned above.

The pinene synthase coding sequence DNA is inserted in the pET28b(+) vector and mutagenized using the Quick-Change® method, following a protocol similar to that described in Example 1 for the TEAS 406/407 random library. The primers used to generate mutations are synthesized as indicated in Example 1, using N or S as nucleotides in the desired codons in order to generate random mutants.

Specific mutations at one or more of the above 19 pinene synthase amino acid residues are made by site-directed mutagenesis using a protocol similar to that described in Example 1 for TEAS. Primers are made that have specific A, T, C or G substitutions in the codons to be mutated, in order to generate the desired mutant(s).

Random and/or specific mutations are prepared in a manner similar to that described above to alter amino acid residues of other monoterpene synthases, e.g., limonene synthase, (SEQ ID NOS:22 or 58), myrcene synthase (SEQ ID NO:30), +sabinene synthase (SEQ ID NO:54), 1, 8 cineole synthase (SEQ ID NO:24) and +bornyl diphosphate synthase (SEQ ID NO:26), at residues whose α -carbons have the interatomic distances and structural coordinates described in Tables 1–6.

EXAMPLE 13

Generation of Novel Sesquiterpene Synthase Genes

A DNA sequence encoding a cadinene synthase (SEQ ID NO:33) is used to construct a library of mutant cadinene synthases. Random mutations are introduced at nucleotides encoding one or more of the following nine amino acid residues: W, I, S, G, Y, L, C, L and Y, which correspond to amino acid residues 280, 301, 409, 410, 411, 414, 448, 527 and 528 encoded by SEQ ID NO:33.

In some cases, the cadinene synthase coding sequence is mutated at nucleotides encoding one or more of amino acid residues G, A, S, M, D, Y and T, which correspond to amino

acid residues 277, 304, 305, 524, 532, 534 and 535 encoded by SEQ ID NO:33. In addition, the cadinene synthase coding sequence is sometimes mutated at nucleotides encoding one or more of the following amino acid residues: 383, 408 and 452, which correspond to amino acids Y, T and D encoded by SEQ ID NO:33. In some cases, these mutations are made in addition to mutations at the nine residues mentioned above. In other cases, mutations at these ten residues are made without introducing mutations at the nine residues mentioned above.

The cadinene synthase coding sequence is mutated using the QuickChange® method in the pET28b(+) vector, following a protocol similar to that described in Example 1 for the TEAS 406/407 random library. The primers used to generate mutations are synthesized as indicated in Example 11.

Specific mutations at one or more of the above cadinene synthase amino acid residues are made by site-directed mutagenesis using a protocol similar to that described in Example 1 for TEAS.

Random and/or specific mutations are prepared in a manner similar to that described above to alter amino acid residues of other sesquiterpene synthases, e.g., vetispiradiene synthase (SEQ ID NO:32), germacrene C synthase (SEQ ID NO:52), E-alpha-bisabolene synthase (SEQ ID NO:46), gamma-humulene synthase (SEQ ID NO:50), delta-selinene synthase (SEQ ID NO:48), e-b-farnesene synthase (SEQ ID NO:28), at residues whose alpha-carbons have the interatomic distances and structural coordinates described in Tables 1-6.

EXAMPLE 14

Generation of Novel Diterpene Synthase Genes

A DNA sequence encoding an abietadiene synthase (SEQ ID NO:56) is used to construct a library of mutant abietadiene synthases. Random mutations are introduced at nucleotides encoding one or more of the following nine amino acid residues: S, S, I, A, L, V, G, F and Y, which correspond to positions 593, 614, 722, 723, 724, 727, 761, 840 and 841 of SEQ ID NO:56.

In some cases, the abietadiene synthase coding sequence is mutated at nucleotides encoding one or more of amino acid residues I, S, T, M, D, L and T, which correspond to positions 590, 617, 618, 837, 845, 847 and 848 of SEQ ID NO:56. The abietadiene synthase coding sequence is sometimes mutated at nucleotides encoding one or more of the following amino acid residues: Y, S and N, which correspond to positions 696, 721 and 765 of SEQ ID NO:56. In some cases, these mutations are made in addition to mutations at the nine residues mentioned above. In other cases, mutations are made at these ten residues without introducing mutations at the nine residues mentioned above.

The abietadiene synthase coding sequence is mutated using the QuickChange® method in the pET28b(+) vector, following a protocol similar to that described in Example 1 for the TEAS 406/407 random library. The primers used to generate mutations are synthesized as indicated in Example 11.

Specific mutations at one or more of the above abietadiene synthase amino acid residues are made by site-directed mutagenesis using a protocol similar to that described in Example 1 for TEAS.

Random and/or specific mutations are prepared in a manner similar to that described above to alter amino acid

residues of other diterpene synthases at amino acid residues whose alpha-carbons have the interatomic distances and structural coordinates described in Tables 1-6, e.g., casbene synthase (SEQ ID NO:42) and taxadiene synthase (SEQ ID NO:44).

EXAMPLE 15

Expression of Mutant Synthases in Insect, Mammalian and Bacterial Cells

Constructs containing nucleic acids encoding mutant synthases of Examples 12, 13 and/or 14 are introduced into cultured cells of the insect *Spodoptera frugiperda* using a baculovirus expression vector. After expression of the gene, the mutant enzyme is isolated and purified from each clone.

Constructs containing nucleic acids encoding mutant synthases of Examples 12, 13 and/or 14 are introduced into cultured HeLa cells using an expression vector having an SV40 promoter. After expression of the gene, the mutant enzyme is isolated and purified from each clone.

Constructs containing nucleic acids encoding mutant synthases of Examples 12, 13 and/or 14 are introduced into *E. coli* BL-21 on a plasmid vector as described in Example 1. The mutant synthase gene is expressed and the mutant enzyme is isolated and purified as described in Example 2.

Other Embodiments

To the extent not already indicated, it will be understood by those of ordinary skill in the art that any one of the various specific embodiments herein described and illustrated may be further modified to incorporate features shown in other of the specific embodiments.

It is to be understood that while the invention has been described in conjunction with the Detailed Description thereof, that the foregoing description is intended to illustrate, and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

TABLE 1

α-Carbon	X Position	Y Position	Z Position
1	119.144	43.487	44.133
2	120.203	38.695	43.506
3	114.058	43.884	41.015
4	109.327	46.145	41.743
5	110.682	46.410	45.284
6	99.381	42.920	45.148
7	103.445	38.054	44.605
8	106.807	36.336	45.151
9	107.629	38.010	41.804
10	109.375	34.842	40.617
11	111.944	37.854	37.602
12	110.233	31.098	47.361
13	109.178	33.314	52.875
14	115.915	32.218	48.369
15	118.846	34.443	51.796
16	116.461	32.848	54.290
17	114.100	38.006	55.620
18	116.617	41.285	51.702
19	114.855	43.486	54.238

TABLE 2

α -carbon	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	0.0	5.0	6.0	10.5	9.0	19.8	16.6	14.3	13.0	13.5	11.2	15.6	16.7	12.5	11.8	15.0	13.7	8.3	11.0
2	5.0	0.0	8.4	13.3	12.4	21.3	16.8	13.7	12.7	11.9	10.2	13.1	15.4	9.2	9.4	12.8	13.6	9.3	12.9
3	6.0	8.4	0.0	5.3	6.0	15.3	12.6	11.3	8.7	10.2	7.2	14.8	16.6	13.9	15.1	17.4	15.7	11.3	13.3
4	10.5	13.3	5.3	0.0	3.8	11.0	10.4	10.7	8.3	11.4	9.6	16.1	17.0	16.8	18.1	19.6	16.8	13.3	13.9
5	9.0	12.4	6.0	3.8	0.0	11.8	11.1	10.8	9.6	12.5	11.6	15.5	15.2	15.4	15.9	17.3	13.8	10.1	10.3
6	19.8	21.3	15.3	11.0	11.8	0.0	6.4	9.9	10.2	13.6	15.5	16.2	15.8	20.0	22.2	21.8	18.7	18.5	18.0
7	16.6	16.8	12.6	10.4	11.1	6.4	0.0	3.8	5.0	7.8	11.0	10.1	11.1	14.3	17.4	17.0	15.3	15.3	15.9
8	14.3	13.7	11.3	10.7	10.8	9.9	3.8	0.0	3.8	5.4	9.3	6.6	8.6	10.5	13.9	13.7	12.9	12.8	14.1
9	13.0	12.7	8.7	8.3	9.6	10.2	5.0	3.8	0.0	3.8	6.0	9.2	12.1	12.1	15.4	16.1	15.3	13.8	15.4
10	13.5	11.9	10.2	11.4	12.5	13.6	7.8	5.4	3.8	0.0	5.0	7.8	12.4	10.5	14.6	15.5	16.0	14.7	17.0
11	11.2	10.2	7.2	9.6	11.6	15.5	11.0	9.3	6.0	5.0	0.0	12.0	16.2	12.8	16.1	18.0	18.2	15.3	17.8
12	15.6	13.1	14.8	16.1	15.5	16.2	10.1	6.6	9.2	7.8	12.0	0.0	6.0	5.9	10.2	9.5	11.4	12.8	14.9
13	16.7	15.4	16.6	17.0	15.2	15.8	11.1	8.6	12.1	12.4	16.2	6.0	0.0	8.2	9.8	7.4	7.3	11.0	11.7
14	12.5	9.2	13.9	16.8	15.4	20.0	14.3	10.5	12.1	10.5	12.8	5.9	8.2	0.0	5.0	6.0	9.5	9.7	12.8
15	11.8	9.4	15.1	18.1	15.9	22.2	17.4	13.9	15.4	14.6	16.1	10.2	9.8	5.0	0.0	3.8	7.1	7.2	10.2
16	15.0	12.8	17.4	19.6	17.3	21.8	17.0	13.7	16.1	15.5	18.0	9.5	7.4	6.0	3.8	0.0	5.8	8.8	10.8
17	13.7	13.6	15.7	16.8	13.8	18.7	15.3	12.9	15.3	16.0	18.2	11.4	7.3	9.5	7.1	5.8	0.0	5.7	5.7
18	8.3	9.3	11.3	13.3	10.1	18.5	15.3	12.8	13.8	14.7	15.3	12.8	11.0	9.7	7.2	8.8	5.7	0.0	3.8
19	11.0	12.9	13.3	13.9	10.3	18.0	15.9	14.1	15.4	17.0	17.8	14.9	11.7	12.8	10.2	10.8	5.7	3.8	0.0

TABLE 3

TABLE 3-continued

α -Carbon	X Position	Y Position	Z Position		α -Carbon	X Position	Y Position	Z Position
1	119.144	43.487	44.133	25	10	110.233	31.098	47.361
2	120.203	38.695	43.506		11	115.915	32.218	48.369
3	114.058	43.884	41.015		12	118.846	34.443	51.796
4	109.327	46.145	41.743		13	116.461	32.848	54.290
5	110.682	46.410	45.284	35	14	114.100	38.006	55.620
6	106.807	36.336	45.151		15	116.617	41.285	51.702
7	107.629	38.010	41.804		16	114.855	43.486	54.238
8	109.375	34.842	40.617					
9	111.944	37.854	37.602					

TABLE 4

α -Carbon	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0.0	5.0	6.0	10.5	9.0	14.3	13.0	13.5	11.2	15.6	12.5	11.8	15.0	13.7	8.3	11.0
2	5.0	0.0	8.4	13.3	12.4	13.7	12.7	11.9	10.2	13.1	9.2	9.4	12.8	13.6	9.3	12.9
3	6.0	8.4	0.0	5.3	6.0	11.3	8.7	10.2	7.2	14.8	13.9	15.1	17.4	15.7	11.3	13.3
4	10.5	13.3	5.3	0.0	3.8	10.7	8.3	11.4	9.6	16.1	16.8	18.1	19.6	16.8	13.3	13.9
5	9.0	12.4	6.0	3.8	0.0	10.8	9.6	12.5	11.6	15.5	15.4	15.9	17.3	13.8	10.1	10.3
6	14.3	13.7	11.3	10.7	10.8	0.0	3.8	5.4	9.3	6.6	10.5	13.9	13.7	12.9	12.8	14.1
7	13.0	12.7	8.7	8.3	9.6	3.8	0.0	3.8	6.0	9.2	12.1	15.4	16.1	15.3	13.8	15.4
8	13.5	11.9	10.2	11.4	12.5	5.4	3.8	0.0	5.0	7.8	10.5	14.6	15.5	16.0	14.7	17.0
9	11.2	10.2	7.2	9.6	11.6	9.3	6.0	5.0	0.0	12.0	12.8	16.1	18.0	18.2	15.3	17.8
10	15.6	13.1	14.8	16.1	15.5	6.6	9.2	7.8	12.0	0.0	5.9	10.2	9.5	11.4	12.8	14.9
11	12.5	9.2	13.9	16.8	15.4	10.5	12.1	10.5	12.8	5.9	0.0	5.0	6.0	9.5	9.7	12.8
12	11.8	9.4	15.1	18.1	15.9	13.9	15.4	14.6	16.1	12.2	5.0	0.0	3.8	7.1	7.2	10.2
13	15.0	12.8	17.4	19.6	17.3	13.7	16.1	15.5	18.0	9.5	6.0	3.8	0.0	5.8	8.8	10.8
14	13.7	13.6	15.7	16.8	13.8	12.9	15.3	16.0	18.2	11.4	9.5	7.1	5.8	0.0	5.7	5.7
15	8.3	9.3	11.3	13.3	10.1	12.8	13.8	14.7	15.3	12.8	9.7	7.2	8.8	5.7	0.0	3.8
16	11.0	12.9	13.3	13.9	10.3	14.1	15.4	17.0	17.8	14.9	12.8	10.2	10.8	5.7	3.8	0.0

TABLE 5

α -Carbon	X Position	Y Position	Z Position
1	120.203	38.695	43.506
2	114.058	43.884	41.015
3	106.807	36.336	45.151
4	107.629	38.010	41.804
5	109.375	34.842	40.617
6	111.944	37.854	37.602
7	110.233	31.098	47.361
8	118.846	34.443	51.796
9	116.461	32.848	54.290

TABLE 6

α -Carbon	1	2	3	4	5	6	7	8	9
1	0	8.4	13.7	12.7	11.9	10.2	13.1	9.4	12.8
2	8.4	0	11.3	8.7	10.2	7.2	14.8	15.1	17.4
3	13.7	11.3	0	3.8	5.4	9.3	6.6	13.9	13.7
4	12.7	8.7	3.8	0	3.8	6	9.2	15.4	16.1
5	11.9	10.2	5.4	3.8	0	5	7.8	14.6	15.5
6	10.2	7.2	9.3	6	5	0	12	16.1	18
7	13.1	14.8	6.6	9.2	7.8	12	0	10.2	9.5
8	9.4	15.1	13.9	15.4	14.6	16.1	10.2	0	3.8
9	12.8	17.4	13.7	16.1	15.5	18	9.5	3.8	0

TABLE 7

Ordered Arrangement of R-Groups at α -carbons 1-19																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
A	C	W	I	I	S	Y	T	T	T	Y	L	C	D	V	T	Y	D	Y	T
B	C	W	I	I	S	Y	T	S	T	Y	L	C	D	I	T	Y	D	Y	T
C	G	W	I	A	S	Y	T	C	G	Y	L	C	D	M	L	Y	D	Y	T
D	G	W	I	A	S	Y	T	S	G	Y	L	C	D	M	L	Y	D	Y	T
E	C	W	L	T	S	Y	S	A	G	Y	I	A	N	A	L	Y	D	Y	T
F	G	W	L	L	S	Y	S	T	V	H	L	G	D	A	V	Y	D	Y	T
G	C	W	L	T	S	Y	S	A	G	Y	I	A	N	A	L	Y	D	Y	S
H	L	W	I	T	T	Y	S	V	G	N	L	F	D	V	L	Y	D	F	T
I	P	W	I	V	D	Y	S	T	A	G	L	S	D	A	C	Y	D	Y	T
J	A	W	V	C	G	F	T	S	C	I	M	G	N	C	S	Y	D	Y	S
K	N	F	F	L	G	A	E	I	T	A	T	G	N	I	T	Y	E	F	T
L	C	W	N	I	T	Y	S	I	S	G	M	L	D	A	M	Y	D	H	Q
M	S	W	V	L	T	Y	S	S	S	Y	L	G	G	V	L	Y	D	F	T
N	N	F	F	L	V	N	A	T	L	A	L	G	N	L	S	Y	E	F	T
O	C	W	N	I	T	Y	I	S	G	P	L	L	D	A	M	Y	D	H	G
P	C	W	N	V	T	Y	I	G	G	I	L	L	D	A	I	Y	D	F	G
Q	C	Y	L	L	T	F	A	V	T	M	T	G	N	I	T	Y	D	Y	T
R	C	W	I	I	T	Y	S	I	S	A	I	L	D	A	I	Y	D	D	G
S	S	W	F	I	V	F	S	S	S	V	I	L	N	V	I	Y	D	H	G
T	S	W	I	A	T	Y	S	V	A	S	I	L	D	A	I	Y	D	F	G
U	N	W	N	L	T	Y	S	I	S	S	I	F	N	S	M	Y	D	H	G
V	F	L	A	Q	T	Y	S	I	G	Q	L	S	D	T	I	F	D	F	G
W	I	S	S	T	V	Y	S	I	A	L	V	G	N	M	F	Y	D	L	T
X	Y	L	C	I	T	Y	S	C	G	H	S	L	G	F	G	Y	D	Y	S
Y	G	S	F	I	T	F	S	S	S	V	I	L	N	A	V	Y	D	H	G
Z	Y	W	A	C	T	Y	S	S	G	M	L	G	D	L	I	Y	D	L	Y
AA	A	A	N	L	T	N	A	L	T	S	T	C	M	L	L	Y	D	Y	N
BB	F	L	C	V	T	Y	S	S	A	Y	V	L	G	L	L	Y	D	F	S
CC	F	W	A	M	T	Y	N	T	G	M	L	S	D	I	M	Y	D	F	S
DD	Y	M	C	V	T	F	V	S	S	G	I	L	G	F	V	Y	D	Y	T
EE	V	S	G	Q	V	Y	S	V	G	L	C	W	N	V	F	Y	D	Y	G
FF	C	S	G	T	T	M	F	A	L	G	V	G	N	L	F	Y	D	F	T
GG	C	S	G	T	T	M	S	F	A	L	I	G	N	L	F	Y	D	F	T
HH	C	A	G	T	T	M	S	F	A	L	I	G	N	V	F	Y	D	Y	T
II	I	W	V	I	S	Y	T	T	G	L	V	I	N	T	S	Y	D	Y	T
JJ	Y	W	A	C	T	Y	S	S	G	M	L	G	D	L	I	Y	D	L	Y
KK	C	W	I	I	S	Y	T	S	T	Y	L	C	D	V	T	Y	D	Y	T
LL	C	W	I	I	S	Y	T	T	T	Y	L	C	D	I	T	Y	D	Y	T
MM	C	W	N	I	T	Y	S	I	S	G	M	L	D	A	M	Y	D	H	G
NN	F	A	A	Q	T	Y	S	I	G	Q	L	S	D	T	I	F	D	F	G
OO	F	A	I	A	T	Y	S	V	A	S	I	L	D	A	I	Y	D	F	G

TABLE 8

Ordered Arrangement of R-Groups at α -carbons 1-16																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
A	C	W	I	I	S	T	T	Y	L	C	V	T	Y	D	Y	T
B	C	W	I	I	S	S	T	Y	L	C	I	T	Y	D	Y	T
C	G	W	I	A	S	C	G	Y	L	C	M	L	Y	D	Y	T
D	G	W	I	A	S	S	G	Y	L	C	M	L	Y	D	Y	T
E	C	W	L	T	S	A	G	Y	I	A	A	L	Y	D	Y	T
F	G	W	L	L	S	T	V	H	L	G	A	V	Y	D	Y	T
G	C	W	L	T	S	A	G	Y	I	A	A	L	Y	D	Y	S
H	L	W	I	T	T	V	G	N	L	F	V	L	Y	D	F	T
I	P	W	I	V	D	T	A	G	L	S	A	C	Y	D	Y	T
J	A	W	V	C	G	S	C	I	M	G	C	S	Y	D	Y	S
K	N	F	F	L	G	I	T	A	T	G	I	T	Y	E	F	T
L	C	W	N	I	T	I	S	G	M	L	A	M	Y	D	H	Q
M	S	W	V	L	T	S	S	Y	L	G	V	L	Y	D	F	T
N	N	F	V	L	V	T	L	A	L	G	L	S	Y	E	F	T
O	C	W	N	I	T	S	G	P	L	L	A	M	Y	D	H	G
P	C	W	N	V	T	G	G	I	L	L	A	I	Y	D	F	G
Q	C	Y	L	L	T	V	T	M	T	G	I	T	Y	D	Y	T
R	C	W	I	I	T	I	S	A	I	L	A	I	Y	D	D	G
S	S	W	F	I	V	S	S	V	I	L	V	I	Y	D	H	G
T	S	W	I	A	T	V	A	S	I	L	A	I	Y	D	F	G
U	N	W	N	L	T	I	S	S	I	F	S	M	Y	D	H	G
V	F	L	A	Q	T	I	G	Q	L	S	T	I	F	D	F	G
W	I	S	S	T	V	I	A	L	V	G	M	F	Y	D	L	T
X	Y	L	C	I	T	C	G	H	S	L	F	G	Y	D	Y	S
Y	G	S	F	I	T	S	S	V	I	L	A	V	Y	D	H	G
Z	Y	W	A	C	T	S	G	M	L	G	L	I	Y	D	L	Y
AA	A	A	N	L	T	L	T	S	T	C	L	L	Y	D	Y	N
BB	F	L	C	V	T	S	A	Y	V	L	L	Y	D	F	S	S
CC	F	W	A	M	T	T	G	M	L	S	I	M	Y	D	F	S
DD	Y	M	C	V	T	S	S	G	I	L	F	V	Y	D	Y	T
EE	V	S	G	Q	V	V	G	L	C	W	V	F	Y	D	Y	G
FF	C	S	G	T	T	A	L	G	V	G	L	F	Y	D	F	T
GG	C	S	G	T	T	F	A	L	I	G	L	F	Y	D	F	T
HH	C	A	G	T	T	F	A	L	I	G	V	F	Y	D	Y	T
II	I	W	V	I	S	T	G	L	V	I	T	S	Y	D	Y	T
JJ	Y	W	A	C	T	S	G	M	L	G	L	I	Y	D	L	Y
KK	C	W	I	I	S	S	T	Y	L	C	V	T	Y	D	Y	T
LL	C	W	I	I	S	T	T	Y	L	C	I	T	Y	D	Y	T
MM	C	W	N	I	T	I	S	G	M	L	A	M	Y	D	H	G
NN	F	A	A	Q	T	I	G	Q	L	S	T	I	F	D	F	G
OO	F	A	I	A	T	V	A	S	I	L	A	I	Y	D	F	G

TABLE 9

Ordered Arrangements of α -Carbons 1-9									
	1	2	3	4	5	6	7	8	9
A	W	I	T	T	Y	L	C	T	Y
B	W	I	S	T	Y	L	C	T	Y
C	W	I	C	G	Y	L	C	L	Y
D	W	I	S	G	Y	L	C	L	Y
E	W	L	A	G	Y	I	A	L	Y
F	W	L	T	V	H	L	G	V	Y
G	W	L	A	G	Y	I	A	L	Y
H	W	I	V	G	N	L	F	L	Y
I	W	I	T	A	G	L	S	C	Y
J	W	V	S	C	I	M	G	S	Y
K	F	F	I	T	A	T	G	T	Y
L	W	N	I	S	G	M	L	M	Y
M	W	V	S	S	Y	L	G	L	Y
N	F	F	T	L	A	L	G	S	Y
O	W	N	S	G	P	L	L	M	Y
P	W	N	G	G	I	L	L	I	Y
Q	Y	L	V	T	M	T	G	T	Y
R	W	I	I	S	A	I	L	I	Y
S	W	F	S	S	V	I	L	I	Y
T	W	I	V	A	S	I	L	I	Y
U	W	N	I	S	S	I	F	M	Y
V	L	A	I	G	Q	L	S	I	F
W	S	S	I	A	L	V	G	F	Y

TABLE 9-continued

Ordered Arrangements of α -Carbons 1-9									
	1	2	3	4	5	6	7	8	9
X	L	C	C	G	H	S	L	G	Y
Y	S	F	S	S	V	I	L	V	Y
Z	W	A	S	G	M	L	G	I	Y
AA	A	N	L	T	S	T	C	L	Y
BB	L	C	S	A	Y	V	L	L	Y
CC	W	A	T	G	M	L	S	M	Y
DD	M	C	S	S	G	I	L	V	Y
EE	S	G	V	G	L	C	W	F	Y
FF	S	G	A	L	G	V	G	F	Y
GG	S	G	F	A	L	I	G	F	Y
HH	A	G	F	A	L	I	G	F	Y
II	W	V	T	G	L	V	I	S	Y
JJ	W	A	S	G	M	L	G	I	Y
KK	W	I	S	T	Y	L	C	T	Y
LL	W	I	T	T	Y	L	C	T	Y
MM	W	N	I	S	G	M	L	M	Y
NN	A	A	I	G	Q	L	S	I	F
OO	A	I	V	A	S	I	L	I	Y

TABLE 12-continued

Query:	61	INEIDRLPDYMKISYKAILDLKYDEKELSSAGRSHIVCHAIERMKEVVRNYNVESTWFI	120
		IN ID+LPDYN++ + A+ + D ++ +++ + + ++ Y VE+ WF	
Sbjct:	376	INSIDQLPDYMQLCFLALNNFVDDTSDYVMKEKGVNVIPLRQSWVDLADKYMVEARWFY	435
Query:	121	EGYMPVSEYLSNALATTTYYLATTSYLGM-KSATEQDFEWLSKNPKILEASVVICRVI	179
		G+ P + EYL N+ + + + T + + S T++ + L K ++ S + R+	
Sbjct:	436	GGHKPSLEEYLENSQWISGPCMLTHIFFRVTDSTFKETVDSLYKYHDLVRWSSFVLR	495
Query:	180	DDTATYEVEKSRGQIATGIECCMRDYGISTKEAMAKFQNMETAWKDIN-EGLLRPTVPS	238
		DD T E SRG + ++C M DY S EA + + WK +N E + + +P	
Sbjct:	496	DDLGTSSVEEVRSGDVPKSLQCYMSDYNASEAEARKHVKWLIAEVWKKMNAERVSKDSPFG	555
Query:	239	TEFLTPILNLRARIVEVITYIHNLGDY--THP	266
		+F+ ++L R+ ++ Y HN DG+ HP	
Sbjct:	556	KDFIGCAVDLGRMAQLMY-HNGDGHGTQHP	584

TABLE 13

Score = 116 bits (289), Expect = 1e-25
 Identities = 77/270 (28%), Positives = 126/270 (46%), Gaps = 6/270 (2%)

Query:	3	VAEVYFSSATFEP-EYSATRIAFTKIGCLQVLFDDMADIFATLDELKSFTEGVKRWDTSL	61
		V +++ FEP ++ R I L + DD+ D++ TLDEL+ FT+ KRWD	
Sbjct:	318	VESFFWAVGMFEPHQHYQRKMAATIIIVLATVDDIYDVYGTLELELFTDFKRWDTES	377
Query:	62	LNEIPECMQTCFKVWFVLMEEVNNNDVVKVQGRDMLAHIRKPWELYNFCYQEREWLEAGY	121
		+ +P MQ C+ + + D++K G L ++RK Y E +W +GY	
Sbjct:	378	ITRLLPYMQLCYGVHNYISDAAYDILKEHGFCLQYLRKSVVDLVEAYFHEAKWYHSGY	437
Query:	122	IPTFEYLKTYAISVGLGPCQLPILMGEVLKDD--VVEKVHYPSNMFELVSLWRILT	179
		P+ +EYL ISV P + P D V++ ++ ++ L + RL +	
Sbjct:	438	TPSLDEYLNIAKISVA-SPAIISPTYFTFANASHDTAVIDSLYQYHDLCLAGIILRLPD	496
Query:	180	DTKTYQAEKARGQQASGIACYMKNPGEEDAIKHICRVDRALKEASFEYFKPSNDIP	239
		D T E ARG I CYMK+ A+EE+A++H+ ++ A K+ + P	
Sbjct:	497	DLGTSYFELARGDVPKTIQCYMKET-NASEEEAVEHVKFLIREAWKDMN-TAIAAGYPPF	554
Query:	240	MGCKSFIFNLRLCVQIFYKFDYGIANE 269	
		G + N+ Q Y DG+G+ ++	
Sbjct:	555	DGMVAGAAANIGRVAQFIYHLHGDGFGVQHSK	584

TABLE 14

Score = 120 bits (299), Expect = 6e-27
 Identities = 70/272 (25%), Positives = 137/272 (49%), Gaps = 3/272 (1%)

Query:	2	RVVECYFWALGVYFEPQYSQARVMLVKTISMISIVDDTFDAYGTVELEAYTDAIQRWDI	61
		R VE Y W + FEP++S++R+ KT + +++DD +D + T+ E++ T+ ++RWD+	
Sbjct:	296	RHVEYYSWVVMCIFEPFESRIAFAKTAIILCTVLDLDTHTATLHEIKIMTEGVRWRDL	355
Query:	62	NEIDRLPDYMKISYKAILDLKYDEKELSSAGRSHIVCHAIERMKEVVRNYNVESTWFI	121
		+ D LPDY+KI+++ + + E+ + + K + +Y E+ W	
Sbjct:	356	SLTDDLPDYIKIAFPFFNTVNLIVEIVKRGQRDMTIVKDCWKRYTESYLQEAEWIAT	415
Query:	122	GYMPVSEYLSNALATTTYYLATTSYLGM-KSATEQDFEWLSKNPKILEASVVICRVID	180
		G++P +EY+ N +A++ L L + K + E + KIL+ + R+ D	
Sbjct:	416	GHIPTFNEYIKNGMASSGMCILNPLLLDCLLPDNIHQIHSKILDLLELTGRIAD	475
Query:	181	DTATYEVEKSRGQIATGIECCMRDYGISTKE-AMAKFQNMETAWKDINEGLLRPTVST	239
		D +E EK RG++A+ ++C M++ ST E A+ + + + ++ N ++ V	
Sbjct:	476	DLKDFEDEKERGEMASSLQCYMKENPESTVENALNHKIGILNRSLEEFNWEFMKQDSVPM	535
Query:	240	EFLTPILNLRARIVEVITYIHNLGDYTHPEKVLK	271
		N+ R ++ Y + DG +K +K	
Sbjct:	536	CCKKFTFNIGRGLQFIYKYR-DGLYISDKVEK	566

TABLE 15

Score = 221 bits (557), Expect = 4e-57
 Identities = 120/263 (42%), Positives = 178/233 (62%), Gaps = 6/283 (2%)

Query:	5	EFYFWMAAAISEPEFSGSRVAFTKIALLMTMLDLDYDTHGTLOQLKIFTEGVRWDVSLV	64
		E YF A+ I EPEFS R +TK + +LDDLYD HG+LD LK+PTE V+RWD+SLV	
Sbjct:	589	EIYFSPASFIFEPFESKREVVYTKSNFTVILDDLYDAHGSLLDKLFTESVSRWDLSLV	648
Query:	65	EGLPDFMKIAFEFVWLTSLNIAEAVKQGDMAAYIRKNAWERYLEAYLQDAEWIATGH	124
		+ +P MKI F + T N++ E + QG+D+ YI +N W+ LEAY ++AEW +	
Sbjct:	649	DQMPQMKICVGVYNTFNDAIKEGRERQGRDVLGYI-QNVWVQVLEAYTKEAEWSEAKY	707
Query:	125	VPTDEYLNNGTPTNGMVLNLIPLLLMGEHLPIDILEQIFLPSRFHLLTELASRLVDDA	184
		VP+P+EY+N + + + + LI L GE L ++L +I SRF L+ L RLV+D	
Sbjct:	708	VPSFNEYENASVSIALGTVVLI SALFTGEVLTDEVLSKIDRESRPLQLMGLTGRLVNDT	767
Query:	185	RDFQAEKDHGDL-SCIECYLKDHPESSTVEDALNHVNGLLGNCLLEMNWFLKQDSVPLS	243
		+ +QAE+ G++ S I+CY+KDHP+ + E+AL HV ++ N L E+N +F+ + +P	

TABLE 15-continued

Sbjct: 768 KTYQAERGQGEVASAIQCVMKDHPIKISEEEALQHVYSVMENALEELNREFV--NNKIPDI 825
 Query: 244 CKKYSFHVLRARSIQFMYNQGDGFSISNKV-IKDQVQKVLIVPV 285
 K+ F AR +Q Y QGDG ++S+ + IK+ V+ L PV
 Sbjct: 826 YKRLVFET-ARIMQLFYMQGDGLTSLSHDMEIKEHVKNCLFQPV 867

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 58

<210> SEQ ID NO 1
 <211> LENGTH: 1671
 <212> TYPE: DNA
 <213> ORGANISM: Nicotiana tabacum
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (25)...(1668)

<400> SEQUENCE: 1

atggtccatc atcatcatca tcat atg gcc tca gca gca gtt gca aac tat	51
Met Ala Ser Ala Ala Val Ala Asn Tyr	
1 5	
gaa gaa gag att gtt cgc ccc gtc gcc gac ttc tcc cct agt ttg tgg	99
Glu Glu Glu Ile Val Arg Pro Val Ala Asp Phe Ser Pro Ser Leu Trp	
10 15 20 25	
ggt gat cag ttc ctt tca ttc tcc att aaa aat cag gtt gca gaa aag	147
Gly Asp Gln Phe Leu Ser Phe Ser Ile Lys Asn Gln Val Ala Glu Lys	
30 35 40	
tat gct caa gag att gaa gca ttg aag gaa caa acg agg aat atg ctg	195
Tyr Ala Gln Glu Ile Glu Ala Leu Lys Glu Gln Thr Arg Asn Met Leu	
45 50 55	
tta gca act gga atg aaa ttg gct gat aca ctg aat ttg ata gac act	243
Leu Ala Thr Gly Met Lys Leu Ala Asp Thr Leu Asn Leu Ile Asp Thr	
60 65 70	
att gaa cgc ctt ggc ata tcc tac cac ttt gag aaa gaa att gat gat	291
Ile Glu Arg Leu Gly Ile Ser Tyr His Phe Glu Lys Glu Ile Asp Asp	
75 80 85	
att ttg gat cag att tac aac caa aac tca aac tgc aac gat ttg tgc	339
Ile Leu Asp Gln Ile Tyr Asn Gln Asn Ser Asn Cys Asn Asp Leu Cys	
90 95 100 105	
act tct gca ctt caa ttt cga ttg ctc agg caa cat ggt ttc aac atc	387
Thr Ser Ala Leu Gln Phe Arg Leu Leu Arg Gln His Gly Phe Asn Ile	
110 115 120	
tct cct gaa att ttc agc aaa ttc caa gac gaa aat ggc aaa ttc aag	435
Ser Pro Glu Ile Phe Ser Lys Phe Gln Asp Glu Asn Gly Lys Phe Lys	
125 130 135	
gaa tct ctt gct agt gat gtc tta gga tta ttg aac ttg tat gaa gct	483
Glu Ser Leu Ala Ser Asp Val Leu Gly Leu Leu Asn Leu Tyr Glu Ala	
140 145 150	
tca cat gta agg act cat gct gac gat atc tta gaa gac gca ctt gct	531
Ser His Val Arg Thr His Ala Asp Asp Ile Leu Glu Asp Ala Leu Ala	
155 160 165	
ttc tcc act atc cat ctt gaa tct gca gct cca cat ttg aaa tct cca	579
Phe Ser Thr Ile His Leu Glu Ser Ala Ala Pro His Leu Lys Ser Pro	
170 175 180 185	
ctt agg gag caa gtg aca cat gcc ctt gag caa tgt ttg cac aag ggt	627
Leu Arg Glu Gln Val Thr His Ala Leu Glu Gln Cys Leu His Lys Gly	
190 195 200	
gtt cct aga gtc gag acc cga ttc ttc atc tca tca atc tat gac aag	675
Val Pro Arg Val Glu Thr Arg Phe Phe Ile Ser Ser Ile Tyr Asp Lys	

-continued

	525		530		535				1671
--	-----	--	-----	--	-----	--	--	--	------

att att aac cta ctt gtg gac tcc atc aaa att tga
Ile Ile Asn Leu Leu Val Asp Ser Ile Lys Ile
540 545

<210> SEQ ID NO 2
<211> LENGTH: 548
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 2

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

-continued

Leu Ser Ser Ala Gly Arg Ser His Ile Val Cys His Ala Ile Glu Arg
 355 360 365
 Met Lys Glu Val Val Arg Asn Tyr Asn Val Glu Ser Thr Trp Phe Ile
 370 375 380
 Glu Gly Tyr Thr Pro Pro Val Ser Glu Tyr Leu Ser Asn Ala Leu Ala
 385 390 395 400
 Thr Thr Thr Tyr Tyr Tyr Leu Ala Thr Thr Ser Tyr Leu Gly Met Lys
 405 410 415
 Ser Ala Thr Glu Gln Asp Phe Glu Trp Leu Ser Lys Asn Pro Lys Ile
 420 425 430
 Leu Glu Ala Ser Val Ile Ile Cys Arg Val Ile Asp Asp Thr Ala Thr
 435 440 445
 Tyr Glu Val Glu Lys Ser Arg Gly Gln Ile Ala Thr Gly Ile Glu Cys
 450 455 460
 Cys Met Arg Asp Tyr Gly Ile Ser Thr Lys Glu Ala Met Ala Lys Phe
 465 470 475 480
 Gln Asn Met Ala Glu Thr Ala Trp Lys Asp Ile Asn Glu Gly Leu Leu
 485 490 495
 Arg Pro Thr Pro Val Ser Thr Glu Phe Leu Thr Pro Ile Leu Asn Leu
 500 505 510
 Ala Arg Ile Val Glu Val Thr Tyr Ile His Asn Leu Asp Gly Tyr Thr
 515 520 525
 His Pro Glu Lys Val Leu Lys Pro His Ile Ile Asn Leu Leu Val Asp
 530 535 540
 Ser Ile Lys Ile
 545

<210> SEQ ID NO 3
 <211> LENGTH: 1644
 <212> TYPE: DNA
 <213> ORGANISM: *Nicotiana tabacum*
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (1)...(1644)

<400> SEQUENCE: 3

atg gcc tca gca gca gtt gca aac tat gaa gaa gag att gtt cgc ccc	48
Met Ala Ser Ala Ala Val Ala Asn Tyr Glu Glu Ile Val Arg Pro	
1 5 10 15	
gtc gcc gac ttc tcc cct agt ctc tgg ggt gat cag ttc ctt tca ttc	96
Val Ala Asp Phe Ser Pro Ser Leu Trp Gly Asp Gln Phe Leu Ser Phe	
20 25 30	
tcc att gat aat cag gtt gcg gaa aag tat gct caa gag att gaa gca	144
Ser Ile Asp Asn Gln Val Ala Glu Lys Tyr Ala Gln Glu Ile Glu Ala	
35 40 45	
ttg aag gaa caa acg agg agt atg ctg tta gca acc gga agg aaa ttg	192
Leu Lys Glu Gln Thr Arg Ser Met Leu Leu Ala Thr Gly Arg Lys Leu	
50 55 60	
gcc gat aca ttg aat ttg att gac att att gaa cgc ctt ggt ata tcc	240
Ala Asp Thr Leu Asn Leu Ile Asp Ile Ile Glu Arg Leu Gly Ile Ser	
65 70 75 80	
tac cac ttt gag aaa gaa att gat gag att ttg gat cag att tac aac	288
Tyr His Phe Glu Lys Glu Ile Asp Glu Ile Leu Asp Gln Ile Tyr Asn	
85 90 95	
caa aac tca aac tgc aat gat ttg tgc acc tct gca ctt caa ttt cga	336
Gln Asn Ser Asn Cys Asn Asp Leu Cys Thr Ser Ala Leu Gln Phe Arg	
100 105 110	

-continued

ttg ctc agg caa cac ggt ttc aac atc tct cct gaa att ttc agc aaa Leu Leu Arg Gln His Gly Phe Asn Ile Ser Pro Glu Ile Phe Ser Lys 115 120 125	384
ttc caa gat gaa aat ggc aaa ttc aag gag tct ctt gct agt gat gtc Phe Gln Asp Glu Asn Gly Lys Phe Lys Glu Ser Leu Ala Ser Asp Val 130 135 140	432
tta gga tta tta aac ttg tat gaa gct tca cat gta agg act cat gct Leu Gly Leu Leu Asn Leu Tyr Glu Ala Ser His Val Arg Thr His Ala 145 150 155 160	480
gac gat atc tta gaa gac gca ctt gct ttc tcc act atc cat ctt gaa Asp Asp Ile Leu Glu Asp Ala Leu Ala Phe Ser Thr Ile His Leu Glu 165 170 175	528
tct gca gct cca cat ttg aaa tct cca ctt agg gag caa gtg aca cat Ser Ala Ala Pro His Leu Lys Ser Pro Leu Arg Glu Gln Val Thr His 180 185 190	576
gcc ctt gag caa tgt ttg cac aag ggt gtt cct aga gtc gag acc cga Ala Leu Glu Gln Cys Leu His Lys Gly Val Pro Arg Val Glu Thr Arg 195 200 205	624
ttc ttc atc tca tca atc tat gac aag gaa caa tcg aag aat aat gtg Phe Phe Ile Ser Ser Ile Tyr Asp Lys Glu Gln Ser Lys Asn Asn Val 210 215 220	672
tta ctt cga ttt gcc aaa ttg gat ttc aac ttg ctc cag atg ttg cac Leu Leu Arg Phe Ala Lys Leu Asp Phe Asn Leu Leu Gln Met Leu His 225 230 235 240	720
aaa caa gaa ctt gct caa gta tca agg tgg tgg aaa gat ttg gat ttt Lys Gln Glu Leu Ala Gln Val Ser Arg Trp Trp Lys Asp Leu Asp Phe 245 250 255	768
gta aca aca ctt cca tat gct aga gat cga gta gtt gaa tgc tac ttt Val Thr Thr Leu Pro Tyr Ala Arg Asp Arg Val Val Glu Cys Tyr Phe 260 265 270	816
gag gca tta gga gtt tat ttt gag cct caa tac tct caa gct cgc gtc Glu Ala Leu Gly Val Tyr Phe Glu Pro Gln Tyr Ser Gln Ala Arg Val 275 280 285	864
atg ctc gtt aag acc ata tca atg att tcg att gtc gat gac acc ttt Met Leu Val Lys Thr Ile Ser Met Ile Ser Ile Val Asp Asp Thr Phe 290 295 300	912
gat gct tac ggt aca gtt aaa gaa ctt gag gca tac aca gat gcc ata Asp Ala Tyr Gly Thr Val Lys Glu Leu Glu Ala Tyr Thr Asp Ala Ile 305 310 315 320	960
caa aga tgg gat atc aac gaa att gat cgg ctt cct gat tac atg aaa Gln Arg Trp Asp Ile Asn Glu Ile Asp Arg Leu Pro Asp Tyr Met Lys 325 330 335	1008
atc agt tat aaa gct att cta gat ctc tac aag gat tat gaa aag gaa Ile Ser Tyr Lys Ala Ile Leu Asp Leu Tyr Lys Asp Tyr Glu Lys Glu 340 345 350	1056
ttg tct agt gcc gga aga tct cat att gtc tgc cat gca ata gaa aga Leu Ser Ser Ala Gly Arg Ser His Ile Val Cys His Ala Ile Glu Arg 355 360 365	1104
atg aaa gaa gta gta aga aat tat aat gtc gag tca aca tgg ttt att Met Lys Glu Val Val Arg Asn Tyr Asn Val Glu Ser Thr Trp Phe Ile 370 375 380	1152
gaa gga tat atg cca cct gtt tct gaa tac cta agc aat gca cta gca Glu Gly Tyr Met Pro Pro Val Ser Glu Tyr Leu Ser Asn Ala Leu Ala 385 390 395 400	1200
act acc aca tat tac tac ctc gcg aca aca tcg tat ttg ggc atg aag Thr Thr Thr Tyr Tyr Tyr Leu Ala Thr Thr Ser Tyr Leu Gly Met Lys 405 410 415	1248
tct gcc acg gag caa gat ttt gag tgg ttg tca aag aat cca aaa att Ser Ala Thr Glu Gln Asp Phe Glu Trp Leu Ser Lys Asn Pro Lys Ile 420 425 430	1296

-continued

```

ctt gaa gct agt gta att ata tgt cga gtt atc gat gac aca gcc acg      1344
Leu Glu Ala Ser Val Ile Ile Cys Arg Val Ile Asp Asp Thr Ala Thr
      435                      440                      445

tac gag gtt gag aaa agc agg gga caa att gca act gga att gag tgc      1392
Tyr Glu Val Glu Lys Ser Arg Gly Gln Ile Ala Thr Gly Ile Glu Cys
      450                      455                      460

tgc atg aga gat tat ggt ata tca aca aaa gag gca atg gct aaa ttt      1440
Cys Met Arg Asp Tyr Gly Ile Ser Thr Lys Glu Ala Met Ala Lys Phe
      465                      470                      475                      480

caa aat atg gct gag aca gca tgg aaa gat att aat gaa gga ctt ctt      1488
Gln Asn Met Ala Glu Thr Ala Trp Lys Asp Ile Asn Glu Gly Leu Leu
      485                      490                      495

agg ccc act ccc gtc tct aca gaa ttt tta act cct att ctc aat ctt      1536
Arg Pro Thr Pro Val Ser Thr Glu Phe Leu Thr Pro Ile Leu Asn Leu
      500                      505                      510

gct cgt att gtt gag gtt aca tat ata cac aat cta gat gga tac act      1584
Ala Arg Ile Val Glu Val Thr Tyr Ile His Asn Leu Asp Gly Tyr Thr
      515                      520                      525

cat ccg gag aaa gtc tta aaa cct cac att att aac cta ctt gtg gac      1632
His Pro Glu Lys Val Leu Lys Pro His Ile Ile Asn Leu Leu Val Asp
      530                      535                      540

tcc atc aaa att
Ser Ile Lys Ile
545

```

```

<210> SEQ ID NO 4
<211> LENGTH: 548
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

```

```

<400> SEQUENCE: 4

```

```

Met Ala Ser Ala Ala Val Ala Asn Tyr Glu Glu Glu Ile Val Arg Pro
 1          5          10          15

Val Ala Asp Phe Ser Pro Ser Leu Trp Gly Asp Gln Phe Leu Ser Phe
      20          25          30

Ser Ile Asp Asn Gln Val Ala Glu Lys Tyr Ala Gln Glu Ile Glu Ala
      35          40          45

Leu Lys Glu Gln Thr Arg Ser Met Leu Leu Ala Thr Gly Arg Lys Leu
      50          55          60

Ala Asp Thr Leu Asn Leu Ile Asp Ile Ile Glu Arg Leu Gly Ile Ser
      65          70          75          80

Tyr His Phe Glu Lys Glu Ile Asp Glu Ile Leu Asp Gln Ile Tyr Asn
      85          90          95

Gln Asn Ser Asn Cys Asn Asp Leu Cys Thr Ser Ala Leu Gln Phe Arg
      100         105         110

Leu Leu Arg Gln His Gly Phe Asn Ile Ser Pro Glu Ile Phe Ser Lys
      115         120         125

Phe Gln Asp Glu Asn Gly Lys Phe Lys Glu Ser Leu Ala Ser Asp Val
      130         135         140

Leu Gly Leu Leu Asn Leu Tyr Glu Ala Ser His Val Arg Thr His Ala
      145         150         155         160

Asp Asp Ile Leu Glu Asp Ala Leu Ala Phe Ser Thr Ile His Leu Glu
      165         170         175

Ser Ala Ala Pro His Leu Lys Ser Pro Leu Arg Glu Gln Val Thr His
      180         185         190

Ala Leu Glu Gln Cys Leu His Lys Gly Val Pro Arg Val Glu Thr Arg
      195         200         205

```

-continued

Phe Phe Ile Ser Ser Ile Tyr Asp Lys Glu Gln Ser Lys Asn Asn Val
 210 215 220
 Leu Leu Arg Phe Ala Lys Leu Asp Phe Asn Leu Leu Gln Met Leu His
 225 230 235
 Lys Gln Glu Leu Ala Gln Val Ser Arg Trp Trp Lys Asp Leu Asp Phe
 245 250 255
 Val Thr Thr Leu Pro Tyr Ala Arg Asp Arg Val Val Glu Cys Tyr Phe
 260 265 270
 Glu Ala Leu Gly Val Tyr Phe Glu Pro Gln Tyr Ser Gln Ala Arg Val
 275 280 285
 Met Leu Val Lys Thr Ile Ser Met Ile Ser Ile Val Asp Asp Thr Phe
 290 295 300
 Asp Ala Tyr Gly Thr Val Lys Glu Leu Glu Ala Tyr Thr Asp Ala Ile
 305 310 315
 Gln Arg Trp Asp Ile Asn Glu Ile Asp Arg Leu Pro Asp Tyr Met Lys
 325 330 335
 Ile Ser Tyr Lys Ala Ile Leu Asp Leu Tyr Lys Asp Tyr Glu Lys Glu
 340 345 350
 Leu Ser Ser Ala Gly Arg Ser His Ile Val Cys His Ala Ile Glu Arg
 355 360 365
 Met Lys Glu Val Val Arg Asn Tyr Asn Val Glu Ser Thr Trp Phe Ile
 370 375 380
 Glu Gly Tyr Met Pro Pro Val Ser Glu Tyr Leu Ser Asn Ala Leu Ala
 385 390 395
 Thr Thr Thr Tyr Tyr Tyr Leu Ala Thr Thr Ser Tyr Leu Gly Met Lys
 405 410 415
 Ser Ala Thr Glu Gln Asp Phe Glu Trp Leu Ser Lys Asn Pro Lys Ile
 420 425 430
 Leu Glu Ala Ser Val Ile Ile Cys Arg Val Ile Asp Asp Thr Ala Thr
 435 440 445
 Tyr Glu Val Glu Lys Ser Arg Gly Gln Ile Ala Thr Gly Ile Glu Cys
 450 455 460
 Cys Met Arg Asp Tyr Gly Ile Ser Thr Lys Glu Ala Met Ala Lys Phe
 465 470 475 480
 Gln Asn Met Ala Glu Thr Ala Trp Lys Asp Ile Asn Glu Gly Leu Leu
 485 490 495
 Arg Pro Thr Pro Val Ser Thr Glu Phe Leu Thr Pro Ile Leu Asn Leu
 500 505 510
 Ala Arg Ile Val Glu Val Thr Tyr Ile His Asn Leu Asp Gly Tyr Thr
 515 520 525
 His Pro Glu Lys Val Leu Lys Pro His Ile Ile Asn Leu Leu Val Asp
 530 535 540
 Ser Ile Lys Ile
 545

<210> SEQ ID NO 5
 <211> LENGTH: 1644
 <212> TYPE: DNA
 <213> ORGANISM: Nicotiana tabacum
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (1)...(1644)

<400> SEQUENCE: 5

atg gcc tca gca gca gtt gca aac tat gaa gaa gag att gtt cgc ccc

48

-continued

caa aga tgg gat atc aac gaa att gat cgg ctt cct gat tac atg aaa Gln Arg Trp Asp Ile Asn Glu Ile Asp Arg Leu Pro Asp Tyr Met Lys 325 330 335	1008
atc agt tat aaa gct att cta gat ctc tac aag gat tat gaa aag gaa Ile Ser Tyr Lys Ala Ile Leu Asp Leu Tyr Lys Asp Tyr Glu Lys Glu 340 345 350	1056
ttg tct agt gcc gga aga tct cat att gtc tgc cat gca ata gaa aga Leu Ser Ser Ala Gly Arg Ser His Ile Val Cys His Ala Ile Glu Arg 355 360 365	1104
atg aaa gaa gta gta aga aat tat aat gtc gag tca aca tgg ttt att Met Lys Glu Val Val Arg Asn Tyr Asn Val Glu Ser Thr Trp Phe Ile 370 375 380	1152
gaa gga tat atg cca cct gtt tct gaa tac cta agc aat gca cta gca Glu Gly Tyr Met Pro Pro Val Ser Glu Tyr Leu Ser Asn Ala Leu Ala 385 390 395 400	1200
act acc aca tat tac tac ctc gcg aca aca tcg tat ttg ggc atg aag Thr Thr Thr Tyr Tyr Tyr Leu Ala Thr Thr Ser Tyr Leu Gly Met Lys 405 410 415	1248
tct gcc acg gag caa gat ttt gag tgg ttg tca aag aat cca aaa att Ser Ala Thr Glu Gln Asp Phe Glu Trp Leu Ser Lys Asn Pro Lys Ile 420 425 430	1296
ctt gaa gct agt gta att ata tgt cga gtt atc gat gac aca gcc acg Leu Glu Ala Ser Val Ile Ile Cys Arg Val Ile Asp Asp Thr Ala Thr 435 440 445	1344
tac gag gtt gag aaa agc agg gga caa att gca act gga att gag tgc Tyr Glu Val Glu Lys Ser Arg Gly Gln Ile Ala Thr Gly Ile Glu Cys 450 455 460	1392
tgc atg aga gat tat ggt ata tca aca aaa gag gca atg gct aaa ttt Cys Met Arg Asp Tyr Gly Ile Ser Thr Lys Glu Ala Met Ala Lys Phe 465 470 475 480	1440
caa aat atg gct gag aca gca tgg aaa gat att aat gaa gga ctt ctt Gln Asn Met Ala Glu Thr Ala Trp Lys Asp Ile Asn Glu Gly Leu Leu 485 490 495	1488
agg ccc act ccc gtc tct aca gaa ttt tta act cct att ctc aat ctt Arg Pro Thr Pro Val Ser Thr Glu Phe Leu Thr Pro Ile Leu Asn Leu 500 505 510	1536
gct cgt att gtt gag gtt aca ttc ata cac aat cta gat gga tac act Ala Arg Ile Val Glu Val Thr Phe Ile His Asn Leu Asp Gly Tyr Thr 515 520 525	1584
cat ccg gag aaa gtc tta aaa cct cac att att aac cta ctt gtg gac His Pro Glu Lys Val Leu Lys Pro His Ile Ile Asn Leu Leu Val Asp 530 535 540	1632
tcc atc aaa att Ser Ile Lys Ile 545	1644

<210> SEQ ID NO 6
 <211> LENGTH: 548
 <212> TYPE: PRT
 <213> ORGANISM: Nicotiana tabacum

<400> SEQUENCE: 6

Met Ala Ser Ala Ala Val Ala Asn Tyr Glu Glu Glu Ile Val Arg Pro 1 5 10 15
Val Ala Asp Phe Ser Pro Ser Leu Trp Gly Asp Gln Phe Leu Ser Phe 20 25 30
Ser Ile Asp Asn Gln Val Ala Glu Lys Tyr Ala Gln Glu Ile Glu Ala 35 40 45
Leu Lys Glu Gln Thr Arg Ser Met Leu Leu Ala Thr Gly Arg Lys Leu 50 55 60

-continued

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

-continued

Gln Asn Met Ala Glu Thr Ala Trp Lys Asp Ile Asn Glu Gly Leu Leu
 485 490 495

Arg Pro Thr Pro Val Ser Thr Glu Phe Leu Thr Pro Ile Leu Asn Leu
 500 505 510

Ala Arg Ile Val Glu Val Thr Phe Ile His Asn Leu Asp Gly Tyr Thr
 515 520 525

His Pro Glu Lys Val Leu Lys Pro His Ile Ile Asn Leu Leu Val Asp
 530 535 540

Ser Ile Lys Ile
 545

<210> SEQ ID NO 7
 <211> LENGTH: 1644
 <212> TYPE: DNA
 <213> ORGANISM: Nicotiana tabacum
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (1)...(1644)

<400> SEQUENCE: 7

atg gcc tca gca gca gtt gca aac tat gaa gaa gag att gtt cgc ccc	48
Met Ala Ser Ala Ala Val Ala Asn Tyr Glu Glu Glu Ile Val Arg Pro	
1 5 10 15	
gtc gcc gac ttc tcc cct agt ctc tgg ggt gat cag ttc ctt tca ttc	96
Val Ala Asp Phe Ser Pro Ser Leu Trp Gly Asp Gln Phe Leu Ser Phe	
20 25 30	
tcc att gat aat cag gtt gcg gaa aag tat gct caa gag att gaa gca	144
Ser Ile Asp Asn Gln Val Ala Glu Lys Tyr Ala Gln Glu Ile Glu Ala	
35 40 45	
ttg aag gaa caa acg agg agt atg ctg tta gca acc gga agg aaa ttg	192
Leu Lys Glu Gln Thr Arg Ser Met Leu Leu Ala Thr Gly Arg Lys Leu	
50 55 60	
gcc gat aca ttg aat ttg att gac att att gaa cgc ctt ggt ata tcc	240
Ala Asp Thr Leu Asn Leu Ile Asp Ile Ile Glu Arg Leu Gly Ile Ser	
65 70 75 80	
tac cac ttt gag aaa att gat gag att ttg gat cag att tac aac	288
Tyr His Phe Glu Lys Glu Ile Asp Glu Ile Leu Asp Gln Ile Tyr Asn	
85 90 95	
caa aac tca aac tgc aat gat ttg tgc acc tct gca ctt caa ttt cga	336
Gln Asn Ser Asn Cys Asn Asp Leu Cys Thr Ser Ala Leu Gln Phe Arg	
100 105 110	
ttg ctc agg caa cac ggt ttc aac atc tct cct gaa att ttc agc aaa	384
Leu Leu Arg Gln His Gly Phe Asn Ile Ser Pro Glu Ile Phe Ser Lys	
115 120 125	
ttc caa gat gaa aat ggc aaa ttc aag gag tct ctt gct agt gat gtc	432
Phe Gln Asp Glu Asn Gly Lys Phe Lys Glu Ser Leu Ala Ser Asp Val	
130 135 140	
tta gga tta tta aac ttg tat gaa gct tca cat gta agg act cat gct	480
Leu Gly Leu Leu Asn Leu Tyr Glu Ala Ser His Val Arg Thr His Ala	
145 150 155 160	
gac gat atc tta gaa gac gca ctt gct ttc tcc act atc cat ctt gaa	528
Asp Asp Ile Leu Glu Asp Ala Leu Ala Phe Ser Thr Ile His Leu Glu	
165 170 175	
tct gca gct cca cat ttg aaa tct cca ctt agg gag caa gtg aca cat	576
Ser Ala Ala Pro His Leu Lys Ser Pro Leu Arg Glu Gln Val Thr His	
180 185 190	
gcc ctt gag caa tgt ttg cac aag ggt gtt cct aga gtc gag acc cga	624
Ala Leu Glu Gln Cys Leu His Lys Gly Val Pro Arg Val Glu Thr Arg	
195 200 205	
ttc ttc atc tca tca atc tat gac aag gaa caa tcg aag aat aat gtg	672

-continued

```

cat ccg gag aaa gtc tta aaa cct cac att att aac cta ctt gtg gac 1632
His Pro Glu Lys Val Leu Lys Pro His Ile Ile Asn Leu Leu Val Asp
530 535 540

```

```

tcc atc aaa att 1644
Ser Ile Lys Ile
545

```

```

<210> SEQ ID NO 8
<211> LENGTH: 548
<212> TYPE: PRT
<213> ORGANISM: Nicotiana tabacum

```

```

<400> SEQUENCE: 8

```

```

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

```

-continued

Ile Ser Tyr Lys Ala Ile Leu Asp Leu Tyr Lys Asp Tyr Glu Lys Glu
 340 345 350
 Leu Ser Ser Ala Gly Arg Ser His Ile Val Cys His Ala Ile Glu Arg
 355 360 365
 Met Lys Glu Val Val Arg Asn Tyr Asn Val Glu Ser Thr Trp Phe Ile
 370 375 380
 Glu Gly Tyr Met Pro Pro Val Ser Glu Tyr Leu Ser Asn Ala Leu Ala
 385 390 395 400
 Thr Thr Thr Tyr Tyr Tyr Leu Ala Thr Thr Ser Tyr Leu Gly Met Lys
 405 410 415
 Ser Ala Thr Glu Gln Asp Phe Glu Trp Leu Ser Lys Asn Pro Lys Ile
 420 425 430
 Leu Glu Ala Ser Val Ile Ile Cys Arg Val Ile Asp Asp Thr Ala Thr
 435 440 445
 Tyr Glu Val Glu Lys Ser Arg Gly Gln Ile Ala Thr Gly Ile Glu Cys
 450 455 460
 Cys Met Arg Asp Tyr Gly Ile Ser Thr Lys Glu Ala Met Ala Lys Phe
 465 470 475 480
 Gln Asn Met Ala Glu Thr Ala Trp Lys Asp Ile Asn Glu Gly Leu Leu
 485 490 495
 Arg Pro Thr Pro Val Ser Thr Glu Phe Leu Thr Pro Ile Leu Asn Leu
 500 505 510
 Ala Arg Ile Val Glu Val Thr Tyr Ile His Asn Leu Asp Gly Phe Thr
 515 520 525
 His Pro Glu Lys Val Leu Lys Pro His Ile Ile Asn Leu Leu Val Asp
 530 535 540
 Ser Ile Lys Ile
 545

<210> SEQ ID NO 9
 <211> LENGTH: 1644
 <212> TYPE: DNA
 <213> ORGANISM: Nicotiana tabacum
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (1)...(1644)

<400> SEQUENCE: 9

atg gcc tca gca gca gtt gca aac tat gaa gaa gag att gtt cgc ccc	48
Met Ala Ser Ala Ala Val Ala Asn Tyr Glu Glu Glu Ile Val Arg Pro	
1 5 10 15	
gtc gcc gac ttc tcc cct agt ctc tgg ggt gat cag ttc ctt tca ttc	96
Val Ala Asp Phe Ser Pro Ser Leu Trp Gly Asp Gln Phe Leu Ser Phe	
20 25 30	
tcc att gat aat cag gtt gcg gaa aag tat gct caa gag att gaa gca	144
Ser Ile Asp Asn Gln Val Ala Glu Lys Tyr Ala Gln Glu Ile Glu Ala	
35 40 45	
ttg aag gaa caa acg agg agt atg ctg tta gca acc gga agg aaa ttg	192
Leu Lys Glu Gln Thr Arg Ser Met Leu Leu Ala Thr Gly Arg Lys Leu	
50 55 60	
gcc gat aca ttg aat ttg att gac att att gaa cgc ctt ggt ata tcc	240
Ala Asp Thr Leu Asn Leu Ile Asp Ile Ile Glu Arg Leu Gly Ile Ser	
65 70 75 80	
tac cac ttt gag aaa gaa att gat gag att ttg gat cag att tac aac	288
Tyr His Phe Glu Lys Glu Ile Asp Glu Ile Leu Asp Gln Ile Tyr Asn	
85 90 95	
caa aac tca aac tgc aat gat ttg tgc acc tct gca ctt caa ttt cga	336
Gln Asn Ser Asn Cys Asn Asp Leu Cys Thr Ser Ala Leu Gln Phe Arg	

-continued

Ala Leu Glu Gln Cys Leu His Lys Gly Val Pro Arg Val Glu Thr Arg
 195 200 205

Phe Phe Ile Ser Ser Ile Tyr Asp Lys Glu Gln Ser Lys Asn Asn Val
 210 215 220

Leu Leu Arg Phe Ala Lys Leu Asp Phe Asn Leu Leu Gln Met Leu His
 225 230 235 240

Lys Gln Glu Leu Ala Gln Val Ser Arg Trp Trp Lys Asp Leu Asp Phe
 245 250 255

Val Thr Thr Leu Pro Tyr Ala Arg Asp Arg Val Val Glu Cys Tyr Phe
 260 265 270

Ser Ala Leu Gly Val Tyr Phe Glu Pro Gln Tyr Ser Gln Ala Arg Val
 275 280 285

Met Leu Val Lys Thr Ile Ser Met Ile Ser Ile Val Asp Asp Thr Phe
 290 295 300

Asp Ala Tyr Gly Thr Val Lys Glu Leu Glu Ala Tyr Thr Asp Ala Ile
 305 310 315 320

Gln Arg Trp Asp Ile Asn Glu Ile Asp Arg Leu Pro Asp Tyr Met Lys
 325 330 335

Ile Ser Tyr Lys Ala Ile Leu Asp Leu Tyr Lys Asp Tyr Glu Lys Glu
 340 345 350

Leu Ser Ser Ala Gly Arg Ser His Ile Val Cys His Ala Ile Glu Arg
 355 360 365

Met Lys Glu Val Val Arg Asn Tyr Asn Val Glu Ser Thr Trp Phe Ile
 370 375 380

Glu Gly Tyr Met Pro Pro Val Ser Glu Tyr Leu Ser Asn Ala Leu Ala
 385 390 395 400

Thr Thr Thr Tyr Tyr Tyr Leu Ala Thr Thr Ser Tyr Leu Gly Met Lys
 405 410 415

Ser Ala Thr Glu Gln Asp Phe Glu Trp Leu Ser Lys Asn Pro Lys Ile
 420 425 430

Leu Glu Ala Ser Val Ile Ile Trp Arg Val Ile Asp Asp Thr Ala Thr
 435 440 445

Tyr Glu Val Glu Lys Ser Arg Gly Gln Ile Ala Thr Gly Ile Glu Cys
 450 455 460

Cys Met Arg Asp Tyr Gly Ile Ser Thr Lys Glu Ala Met Ala Lys Phe
 465 470 475 480

Gln Asn Met Ala Glu Thr Ala Trp Lys Asp Ile Asn Glu Gly Leu Leu
 485 490 495

Arg Pro Thr Pro Val Ser Thr Glu Phe Leu Thr Pro Ile Leu Asn Leu
 500 505 510

Ala Arg Ile Val Glu Val Thr Tyr Ile His Asn Leu Asp Gly Tyr Thr
 515 520 525

His Pro Glu Lys Val Leu Lys Pro His Ile Ile Asn Leu Leu Val Asp
 530 535 540

Ser Ile Lys Ile
 545

<210> SEQ ID NO 11

<211> LENGTH: 1644

<212> TYPE: DNA

<213> ORGANISM: Nicotiana tabacum

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (1)...(1644)

<221> NAME/KEY: misc_feature

<222> LOCATION: (1)...(1644)

-continued

<223> OTHER INFORMATION: n = A,T,C or G

<400> SEQUENCE: 11

atg gcc tca gca gca gtt gca aac tat gaa gaa gag att gtt cgc ccc	48
Met Ala Ser Ala Ala Val Ala Asn Tyr Glu Glu Glu Ile Val Arg Pro	
1 5 10 15	
gtc gcc gac ttc tcc cct agt ctc tgg ggt gat cag ttc ctt tca ttc	96
Val Ala Asp Phe Ser Pro Ser Leu Trp Gly Asp Gln Phe Leu Ser Phe	
20 25 30	
tcc att gat aat cag gtt gcg gaa aag tat gct caa gag att gaa gca	144
Ser Ile Asp Asn Gln Val Ala Glu Lys Tyr Ala Gln Glu Ile Glu Ala	
35 40 45	
ttg aag gaa caa acg agg agt atg ctg tta gca acc gga agg aaa ttg	192
Leu Lys Glu Gln Thr Arg Ser Met Leu Leu Ala Thr Gly Arg Lys Leu	
50 55 60	
gcc gat aca ttg aat ttg att gac att att gaa cgc ctt ggt ata tcc	240
Ala Asp Thr Leu Asn Leu Ile Asp Ile Ile Glu Arg Leu Gly Ile Ser	
65 70 75 80	
tac cac ttt gag aaa gaa att gat gag att ttg gat cag att tac aac	288
Tyr His Phe Glu Lys Glu Ile Asp Glu Ile Leu Asp Gln Ile Tyr Asn	
85 90 95	
caa aac tca aac tgc aat gat ttg tgc acc tct gca ctt caa ttt cga	336
Gln Asn Ser Asn Cys Asn Asp Leu Cys Thr Ser Ala Leu Gln Phe Arg	
100 105 110	
ttg ctc agg caa cac ggt ttc aac atc tct cct gaa att ttc agc aaa	384
Leu Leu Arg Gln His Gly Phe Asn Ile Ser Pro Glu Ile Phe Ser Lys	
115 120 125	
ttc caa gat gaa aat ggc aaa ttc aag gag tct ctt gct agt gat gtc	432
Phe Gln Asp Glu Asn Gly Lys Phe Lys Glu Ser Leu Ala Ser Asp Val	
130 135 140	
tta gga tta tta aac ttg tat gaa gct tca cat gta agg act cat gct	480
Leu Gly Leu Leu Asn Leu Tyr Glu Ala Ser His Val Arg Thr His Ala	
145 150 155 160	
gac gat atc tta gaa gac gca ctt gct ttc tcc act atc cat ctt gaa	528
Asp Asp Ile Leu Glu Asp Ala Leu Ala Phe Ser Thr Ile His Leu Glu	
165 170 175	
tct gca gct cca cat ttg aaa tct cca ctt agg gag caa gtg aca cat	576
Ser Ala Ala Pro His Leu Lys Ser Pro Leu Arg Glu Gln Val Thr His	
180 185 190	
gcc ctt gag caa tgt ttg cac aag ggt gtt cct aga gtc gag acc cga	624
Ala Leu Glu Gln Cys Leu His Lys Gly Val Pro Arg Val Glu Thr Arg	
195 200 205	
ttc ttc atc tca tca atc tat gac aag gaa caa tgc aag aat aat gtg	672
Phe Phe Ile Ser Ser Ile Tyr Asp Lys Glu Gln Ser Lys Asn Asn Val	
210 215 220	
tta ctt cga ttt gcc aaa ttg gat ttc aac ttg ctc cag atg ttg cac	720
Leu Leu Arg Phe Ala Lys Leu Asp Phe Asn Leu Leu Gln Met Leu His	
225 230 235 240	
aaa caa gaa ctt gct caa gta tca agg tgg tgg aaa gat ttg gat ttt	768
Lys Gln Glu Leu Ala Gln Val Ser Arg Trp Trp Lys Asp Leu Asp Phe	
245 250 255	
gta aca aca ctt cca tat gct aga gat cga gta gtt gaa tgc tac ttt	816
Val Thr Thr Leu Pro Tyr Ala Arg Asp Arg Val Val Glu Cys Tyr Phe	
260 265 270	
tgg gca tta gga gtt tat ttt gag cct caa tac tct caa gct cgc gtc	864
Trp Ala Leu Gly Val Tyr Phe Glu Pro Gln Tyr Ser Gln Ala Arg Val	
275 280 285	
atg ctc gtt aag acc ata tca atg att tog att gtc gat gac acc ttt	912
Met Leu Val Lys Thr Ile Ser Met Ile Ser Ile Val Asp Asp Thr Phe	
290 295 300	

-continued

gat gct tac ggt aca gtt aaa gaa ctt gag gca tac aca gat gcc ata 960
 Asp Ala Tyr Gly Thr Val Lys Glu Leu Glu Ala Tyr Thr Asp Ala Ile
 305 310 315 320

caa aga tgg gat atc aac gaa att gat cgg ctt cct gat tac atg aaa 1008
 Gln Arg Trp Asp Ile Asn Glu Ile Asp Arg Leu Pro Asp Tyr Met Lys
 325 330 335

atc agt tat aaa gct att cta gat ctc tac aag gat tat gaa aag gaa 1056
 Ile Ser Tyr Lys Ala Ile Leu Asp Leu Tyr Lys Asp Tyr Glu Lys Glu
 340 345 350

ttg tct agt gcc gga aga tct cat att gtc tgc cat gca ata gaa aga 1104
 Leu Ser Ser Ala Gly Arg Ser His Ile Val Cys His Ala Ile Glu Arg
 355 360 365

atg aaa gaa gta gta aga aat tat aat gtc gag tca aca tgg ttt att 1152
 Met Lys Glu Val Val Arg Asn Tyr Asn Val Glu Ser Thr Trp Phe Ile
 370 375 380

gaa gga tat atg cca cct gtt tct gaa tac cta agc aat gca cta gca 1200
 Glu Gly Tyr Met Pro Pro Val Ser Glu Tyr Leu Ser Asn Ala Leu Ala
 385 390 395 400

act acc aca tat tac nns nns gcg aca aca tcg tat ttg ggc atg aag 1248
 Thr Thr Thr Tyr Tyr Xaa Xaa Ala Thr Thr Ser Tyr Leu Gly Met Lys
 405 410 415

tct gcc acg gag caa gat ttt gag tgg ttg tca aag aat cca aaa att 1296
 Ser Ala Thr Glu Gln Asp Phe Glu Trp Leu Ser Lys Asn Pro Lys Ile
 420 425 430

ctt gaa gct agt gta att ata tgt cga gtt atc gat gac aca gcc acg 1344
 Leu Glu Ala Ser Val Ile Ile Cys Arg Val Ile Asp Asp Thr Ala Thr
 435 440 445

tac gag gtt gag aaa agc agg gga caa att gca act gga att gag tgc 1392
 Tyr Glu Val Glu Lys Ser Arg Gly Gln Ile Ala Thr Gly Ile Glu Cys
 450 455 460

tgc atg aga gat tat ggt ata tca aca aaa gag gca atg gct aaa ttt 1440
 Cys Met Arg Asp Tyr Gly Ile Ser Thr Lys Glu Ala Met Ala Lys Phe
 465 470 475 480

caa aat atg gct gag aca gca tgg aaa gat att aat gaa gga ctt ctt 1488
 Gln Asn Met Ala Glu Thr Ala Trp Lys Asp Ile Asn Glu Gly Leu Leu
 485 490 495

agg ccc act ccc gtc tct aca gaa ttt tta act cct att ctc aat ctt 1536
 Arg Pro Thr Pro Val Ser Thr Glu Phe Leu Thr Pro Ile Leu Asn Leu
 500 505 510

gct cgt att gtt gag gtt aca tat ata cac aat cta gat gga tac act 1584
 Ala Arg Ile Val Glu Val Thr Tyr Ile His Asn Leu Asp Gly Tyr Thr
 515 520 525

cat ccg gag aaa gtc tta aaa cct cac att att aac cta ctt gtg gac 1632
 His Pro Glu Lys Val Leu Lys Pro His Ile Ile Asn Leu Leu Val Asp
 530 535 540

tcc atc aaa att 1644
 Ser Ile Lys Ile
 545

<210> SEQ ID NO 12
 <211> LENGTH: 548
 <212> TYPE: PRT
 <213> ORGANISM: Nicotiana tabacum
 <220> FEATURE:
 <221> NAME/KEY: VARIANT
 <222> LOCATION: (1)...(548)
 <223> OTHER INFORMATION: Xaa = Any Amino Acid
 <400> SEQUENCE: 12

Met Ala Ser Ala Ala Val Ala Asn Tyr Glu Glu Glu Ile Val Arg Pro
 1 5 10 15

-continued

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

-continued

Leu Glu Ala Ser Val Ile Ile Cys Arg Val Ile Asp Asp Thr Ala Thr
 435 440 445

Tyr Glu Val Glu Lys Ser Arg Gly Gln Ile Ala Thr Gly Ile Glu Cys
 450 455 460

Cys Met Arg Asp Tyr Gly Ile Ser Thr Lys Glu Ala Met Ala Lys Phe
 465 470 475 480

Gln Asn Met Ala Glu Thr Ala Trp Lys Asp Ile Asn Glu Gly Leu Leu
 485 490 495

Arg Pro Thr Pro Val Ser Thr Glu Phe Leu Thr Pro Ile Leu Asn Leu
 500 505 510

Ala Arg Ile Val Glu Val Thr Tyr Ile His Asn Leu Asp Gly Tyr Thr
 515 520 525

His Pro Glu Lys Val Leu Lys Pro His Ile Ile Asn Leu Leu Val Asp
 530 535 540

Ser Ile Lys Ile
 545

<210> SEQ ID NO 13
 <211> LENGTH: 33
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: primer for mutagenesis

<400> SEQUENCE: 13

gttgaatgct acttttcggc attagagatt tat 33

<210> SEQ ID NO 14
 <211> LENGTH: 33
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: primer for mutagenesis

<400> SEQUENCE: 14

ataaactcct aatgccgaaa agtagcattc aac 33

<210> SEQ ID NO 15
 <211> LENGTH: 33
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: primer for mutagenesis

<400> SEQUENCE: 15

gctagtgtaa ttatatggcg agttatcgat gac 33

<210> SEQ ID NO 16
 <211> LENGTH: 33
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: primer for mutagenesis

<400> SEQUENCE: 16

gtcatcgata actcgccata taattacact agc 33

<210> SEQ ID NO 17
 <211> LENGTH: 54
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: primer for mutagenesis

-continued

<221> NAME/KEY: misc_feature
 <222> LOCATION: (1)...(54)
 <223> OTHER INFORMATION: n = A,T,C or G

 <400> SEQUENCE: 17
 gcactagcaaa ctaccacata ttacnnsnns gcgacaacat cgtatttggg catg 54

<210> SEQ ID NO 18
 <211> LENGTH: 54
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: primer for mutagenesis
 <221> NAME/KEY: misc_feature
 <222> LOCATION: (1)...(54)
 <223> OTHER INFORMATION: n = A,T,C or G

 <400> SEQUENCE: 18
 catgccc aaa tacgatgttg tcgcsnnsnn gtaatatgtg gtagttgcta gtgc 54

<210> SEQ ID NO 19
 <211> LENGTH: 2018
 <212> TYPE: DNA
 <213> ORGANISM: Abies grandis
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (6)...(1889)
 <223> OTHER INFORMATION: pinene synthase

 <400> SEQUENCE: 19
 cagca atg gct cta gtt tct acc gca ccg ttg gct tcc aaa tca tgc ctg 50
 Met Ala Leu Val Ser Thr Ala Pro Leu Ala Ser Lys Ser Cys Leu
 1 5 10 15
 cac aaa tcg ttg atc agt tct acc cat gag ctt aag gct ctc tct aga 98
 His Lys Ser Leu Ile Ser Ser Thr His Glu Leu Lys Ala Leu Ser Arg
 20 25 30
 aca att cca gct cta gga atg agt agg cga ggg aaa tct atc act cct 146
 Thr Ile Pro Ala Leu Gly Met Ser Arg Arg Gly Lys Ser Ile Thr Pro
 35 40 45
 tcc atc agc atg agc tct acc acc gtt gta acc gat gat ggt gta cga 194
 Ser Ile Ser Met Ser Ser Thr Thr Val Val Thr Asp Asp Gly Val Arg
 50 55 60
 aga cgc atg ggc gat ttc cat tcc aac ctc tgg gac gat gat gtc ata 242
 Arg Arg Met Gly Asp Phe His Ser Asn Leu Trp Asp Asp Asp Val Ile
 65 70 75
 cag tct tta cca acg gct tat gag gaa aaa tcg tac ctg gag cgt gct 290
 Gln Ser Leu Pro Thr Ala Tyr Glu Glu Lys Ser Tyr Leu Glu Arg Ala
 80 85 90 95
 gag aaa ctg atc ggg gaa gta aag aac atg ttc aat tcg atg tca tta 338
 Glu Lys Leu Ile Gly Glu Val Lys Asn Met Phe Asn Ser Met Ser Leu
 100 105 110
 gaa gat gga gag tta atg agt ccg ctc aat gat ctc att caa cgc ctt 386
 Glu Asp Gly Glu Leu Met Ser Pro Leu Asn Asp Leu Ile Gln Arg Leu
 115 120 125
 tgg att gtc gac agc ctt gaa cgt ttg ggg atc cat aga cat ttc aaa 434
 Trp Ile Val Asp Ser Leu Glu Arg Leu Gly Ile His Arg His Phe Lys
 130 135 140
 gat gag ata aaa tcg gcg ctt gat tat gtt tac agt tat tgg ggc gaa 482
 Asp Glu Ile Lys Ser Ala Leu Asp Tyr Val Tyr Ser Tyr Trp Gly Glu
 145 150 155
 aat ggc atc gga tgc ggg agg gag agt gtt gtt act gat ctg aac tca 530
 Asn Gly Ile Gly Cys Gly Arg Glu Ser Val Val Thr Asp Leu Asn Ser
 160 165 170 175
 act gcg ttg ggg ctt cga acc cta cga cta cac gga tac ccg gtg tct 578

-continued

ccc ttt cct gat cat atc ctc aag gaa gtt gac ttc cca tca aag ctt	1538
Pro Phe Pro Asp His Ile Leu Lys Glu Val Asp Phe Pro Ser Lys Leu	
500 505 510	
aac gac ttg gca tgt gcc atc ctt cga tta cga ggt gat acg cgg tgc	1586
Asn Asp Leu Ala Cys Ala Ile Leu Arg Leu Arg Gly Asp Thr Arg Cys	
515 520 525	
tac aag gcg gac agg gct cgt gga gaa gaa gct tcc tct ata tca tgt	1634
Tyr Lys Ala Asp Arg Ala Arg Gly Glu Glu Ala Ser Ser Ile Ser Cys	
530 535 540	
tat atg aaa gac aat cct gga gta tca gag gaa gat gct ctc gat cat	1682
Tyr Met Lys Asp Asn Pro Gly Val Ser Glu Glu Asp Ala Leu Asp His	
545 550 555	
atc aac gcc atg atc agt gac gta atc aaa gga tta aat tgg gaa ctt	1730
Ile Asn Ala Met Ile Ser Asp Val Ile Lys Gly Leu Asn Trp Glu Leu	
560 565 570 575	
ctc aaa cca gac atc aat gtt ccc atc tcg gcg aag aaa cat gct ttt	1778
Leu Lys Pro Asp Ile Asn Val Pro Ile Ser Ala Lys Lys His Ala Phe	
580 585 590	
gac atc gcc aga gct ttc cat tac ggc tac aaa tac cga gac ggc tac	1826
Asp Ile Ala Arg Ala Phe His Tyr Gly Tyr Lys Tyr Arg Asp Gly Tyr	
595 600 605	
agc gtt gcc aac gtt gaa acg aag agt ttg gtc acg aga acc ctc ctt	1874
Ser Val Ala Asn Val Glu Thr Lys Ser Leu Val Thr Arg Thr Leu Leu	
610 615 620	
gaa tct gtg cct ttg tag caacagctca aatctatgcc ctatgctatg	1922
Glu Ser Val Pro Leu	
625	
tcgggttaaa atatatgtgg aaggtagccg ttggatgtag aggataagtt tgttataatt	1982
taataaagtt gtaatttaaa aaaaaaaaa aaaaaa	2018

<210> SEQ ID NO 20

<211> LENGTH: 628

<212> TYPE: PRT

<213> ORGANISM: Abies grandis

<400> SEQUENCE: 20

Met Ala Leu Val Ser Thr Ala Pro Leu Ala Ser Lys Ser Cys Leu His	
1 5 10 15	
Lys Ser Leu Ile Ser Ser Thr His Glu Leu Lys Ala Leu Ser Arg Thr	
20 25 30	
Ile Pro Ala Leu Gly Met Ser Arg Arg Gly Lys Ser Ile Thr Pro Ser	
35 40 45	
Ile Ser Met Ser Ser Thr Thr Val Val Thr Asp Asp Gly Val Arg Arg	
50 55 60	
Arg Met Gly Asp Phe His Ser Asn Leu Trp Asp Asp Asp Val Ile Gln	
65 70 75 80	
Ser Leu Pro Thr Ala Tyr Glu Glu Lys Ser Tyr Leu Glu Arg Ala Glu	
85 90 95	
Lys Leu Ile Gly Glu Val Lys Asn Met Phe Asn Ser Met Ser Leu Glu	
100 105 110	
Asp Gly Glu Leu Met Ser Pro Leu Asn Asp Leu Ile Gln Arg Leu Trp	
115 120 125	
Ile Val Asp Ser Leu Glu Arg Leu Gly Ile His Arg His Phe Lys Asp	
130 135 140	
Glu Ile Lys Ser Ala Leu Asp Tyr Val Tyr Ser Tyr Trp Gly Glu Asn	
145 150 155 160	

-continued

Gly	Ile	Gly	Cys	Gly	Arg	Glu	Ser	Val	Val	Thr	Asp	Leu	Asn	Ser	Thr
			165						170					175	
Ala	Leu	Gly	Leu	Arg	Thr	Leu	Arg	Leu	His	Gly	Tyr	Pro	Val	Ser	Ser
		180						185					190		
Asp	Val	Phe	Lys	Ala	Phe	Lys	Gly	Gln	Asn	Gly	Gln	Phe	Ser	Cys	Ser
		195					200					205			
Glu	Asn	Ile	Gln	Thr	Asp	Glu	Glu	Ile	Arg	Gly	Val	Leu	Asn	Leu	Phe
	210					215					220				
Arg	Ala	Ser	Leu	Ile	Ala	Phe	Pro	Gly	Glu	Lys	Ile	Met	Asp	Glu	Ala
225					230					235					240
Glu	Ile	Phe	Ser	Thr	Lys	Tyr	Leu	Lys	Glu	Ala	Leu	Gln	Lys	Ile	Pro
				245					250					255	
Val	Ser	Ser	Leu	Ser	Arg	Glu	Ile	Gly	Asp	Val	Leu	Glu	Tyr	Gly	Trp
			260					265					270		
His	Thr	Tyr	Leu	Pro	Arg	Leu	Glu	Ala	Arg	Asn	Tyr	Ile	Gln	Val	Phe
		275					280					285			
Gly	Gln	Asp	Thr	Glu	Asn	Thr	Lys	Ser	Tyr	Val	Lys	Ser	Lys	Lys	Leu
	290					295					300				
Leu	Glu	Leu	Ala	Lys	Leu	Glu	Phe	Asn	Ile	Phe	Gln	Ser	Leu	Gln	Lys
305					310					315					320
Arg	Glu	Leu	Glu	Ser	Leu	Val	Arg	Trp	Trp	Lys	Glu	Ser	Gly	Phe	Pro
				325					330					335	
Glu	Met	Thr	Phe	Cys	Arg	His	Arg	His	Val	Glu	Tyr	Tyr	Thr	Leu	Ala
			340					345					350		
Ser	Cys	Ile	Ala	Phe	Glu	Pro	Gln	His	Ser	Gly	Phe	Arg	Leu	Gly	Phe
		355					360					365			
Ala	Lys	Thr	Cys	His	Leu	Ile	Thr	Val	Leu	Asp	Asp	Met	Tyr	Asp	Thr
	370					375						380			
Phe	Gly	Thr	Val	Asp	Glu	Leu	Glu	Leu	Phe	Thr	Ala	Thr	Met	Lys	Arg
385					390					395					400
Trp	Asp	Pro	Ser	Ser	Ile	Asp	Cys	Leu	Pro	Glu	Tyr	Met	Lys	Gly	Val
			405						410					415	
Tyr	Ile	Ala	Val	Tyr	Asp	Thr	Val	Asn	Glu	Met	Ala	Arg	Glu	Ala	Glu
		420						425					430		
Glu	Ala	Gln	Gly	Arg	Asp	Thr	Leu	Thr	Tyr	Ala	Arg	Glu	Ala	Trp	Glu
		435					440					445			
Ala	Tyr	Ile	Asp	Ser	Tyr	Met	Gln	Glu	Ala	Arg	Trp	Ile	Ala	Thr	Gly
	450					455					460				
Tyr	Leu	Pro	Ser	Phe	Asp	Glu	Tyr	Tyr	Glu	Asn	Gly	Lys	Val	Ser	Cys
465					470					475					480
Gly	His	Arg	Ile	Ser	Ala	Leu	Gln	Pro	Ile	Leu	Thr	Met	Asp	Ile	Pro
			485						490					495	
Phe	Pro	Asp	His	Ile	Leu	Lys	Glu	Val	Asp	Phe	Pro	Ser	Lys	Leu	Asn
			500					505					510		
Asp	Leu	Ala	Cys	Ala	Ile	Leu	Arg	Leu	Arg	Gly	Asp	Thr	Arg	Cys	Tyr
		515					520					525			
Lys	Ala	Asp	Arg	Ala	Arg	Gly	Glu	Glu	Ala	Ser	Ser	Ile	Ser	Cys	Tyr
	530					535					540				
Met	Lys	Asp	Asn	Pro	Gly	Val	Ser	Glu	Glu	Asp	Ala	Leu	Asp	His	Ile
545					550					555					560
Asn	Ala	Met	Ile	Ser	Asp	Val	Ile	Lys	Gly	Leu	Asn	Trp	Glu	Leu	Leu
			565						570					575	
Lys	Pro	Asp	Ile	Asn	Val	Pro	Ile	Ser	Ala	Lys	Lys	His	Ala	Phe	Asp

-continued

580	585	590	
Ile Ala Arg Ala Phe His Tyr Gly Tyr Lys Tyr Arg Asp Gly Tyr Ser 595 600 605			
Val Ala Asn Val Glu Thr Lys Ser Leu Val Thr Arg Thr Leu Leu Glu 610 615 620			
Ser Val Pro Leu 625			
 <210> SEQ ID NO 21 <211> LENGTH: 2170 <212> TYPE: DNA <213> ORGANISM: Mentha spicata <220> FEATURE: <221> NAME/KEY: CDS <222> LOCATION: (29)...(1825) <223> OTHER INFORMATION: 4S-limonene synthase <400> SEQUENCE: 21			
agagagagag aggaaggaaa gattaatc atg gct ctc aaa gtg tta agt gtt Met Ala Leu Lys Val Leu Ser Val 1 5			52
gca act caa atg gcg att cct agc aac cta acg aca tgt ctt caa ccc Ala Thr Gln Met Ala Ile Pro Ser Asn Leu Thr Thr Cys Leu Gln Pro 10 15 20			100
tca cac ttc aaa tct tct cca aaa ctg tta tct agc act aac agt agt Ser His Phe Lys Ser Ser Pro Lys Leu Leu Ser Ser Thr Asn Ser Ser 25 30 35 40			148
agt cgg tct cgc ctc cgt gtg tat tgc tcc tcc tcg caa ctc act act Ser Arg Ser Arg Leu Arg Val Tyr Cys Ser Ser Ser Gln Leu Thr Thr 45 50 55			196
gaa aga cga tcc gga aac tac aac cct tct cgt tgg gat gtc aac ttc Glu Arg Arg Ser Gly Asn Tyr Asn Pro Ser Arg Trp Asp Val Asn Phe 60 65 70			244
atc caa tcg ctt ctc agt gac tat aag gag gac aaa cac gtg att agg Ile Gln Ser Leu Leu Ser Asp Tyr Lys Glu Asp Lys His Val Ile Arg 75 80 85			292
gct tct gag ctg gtc act ttg gtg aag atg gaa ctg gag aaa gaa acg Ala Ser Glu Leu Val Thr Leu Val Lys Met Glu Leu Glu Lys Glu Thr 90 95 100			340
gat caa att cga caa ctt gag ttg atc gat gac ttg cag agg atg ggg Asp Gln Ile Arg Gln Leu Glu Leu Ile Asp Asp Leu Gln Arg Met Gly 105 110 115 120			388
ctg tcc gat cat ttc caa aat gag ttc aaa gaa atc ttg tcc tct ata Leu Ser Asp His Phe Gln Asn Glu Phe Lys Glu Ile Leu Ser Ser Ile 125 130 135			436
tat ctc gac cat cac tat tac aag aac cct ttt cca aaa gaa gaa agg Tyr Leu Asp His His Tyr Tyr Lys Asn Pro Phe Pro Lys Glu Glu Arg 140 145 150			484
gat ctc tac tcc aca tct ctt gca ttt agg ctc ctc aga gaa cat ggt Asp Leu Tyr Ser Thr Ser Leu Ala Phe Arg Leu Leu Arg Glu His Gly 155 160 165			532
ttt caa gtc gca caa gag gta ttc gat agt ttc aag aac gag gag ggt Phe Gln Val Ala Gln Glu Val Phe Asp Ser Phe Lys Asn Glu Glu Gly 170 175 180			580
gag ttc aaa gaa agc ctt agc gac gac acc aga gga ttg ttg caa ctg Glu Phe Lys Glu Ser Leu Ser Asp Asp Thr Arg Gly Leu Leu Gln Leu 185 190 195 200			628
tat gaa gct tcc ttt ctg ttg acg gaa ggc gaa acc acg ctc gag tca Tyr Glu Ala Ser Phe Leu Leu Thr Glu Gly Glu Thr Thr Leu Glu Ser 205 210 215			676

-continued

gcg agg gaa ttc gcc acc aaa ttt ttg gag gaa aaa gtg aac gag ggt Ala Arg Glu Phe Ala Thr Lys Phe Leu Glu Glu Lys Val Asn Glu Gly 220 225 230	724
ggt gtt gat ggc gac ctt tta aca aga atc gca tat tct ttg gac atc Gly Val Asp Gly Asp Leu Leu Thr Arg Ile Ala Tyr Ser Leu Asp Ile 235 240 245	772
cct ctt cat tgg agg att aaa agg cca aat gca cct gtg tgg atc gaa Pro Leu His Trp Arg Ile Lys Arg Pro Asn Ala Pro Val Trp Ile Glu 250 255 260	820
tgg tat agg aag agg ccc gac atg aat cca gta gtg ttg gag ctt gcc Trp Tyr Arg Lys Arg Pro Asp Met Asn Pro Val Val Leu Glu Leu Ala 265 270 275 280	868
ata ctc gac tta aat att gtt caa gca caa ttt caa gaa gag ctc aaa Ile Leu Asp Leu Asn Ile Val Gln Ala Gln Phe Gln Glu Glu Leu Lys 285 290 295	916
gaa tcc ttc agg tgg tgg aga aat act ggg ttt gtt gag aag ctg ccc Glu Ser Phe Arg Trp Trp Arg Asn Thr Gly Phe Val Glu Lys Leu Pro 300 305 310	964
ttc gca agg gat aga ctg gtg gaa tgc tac ttt tgg aat act ggg atc Phe Ala Arg Asp Arg Leu Val Glu Cys Tyr Phe Trp Asn Thr Gly Ile 315 320 325	1012
atc gag cca cgt cag cat gca agt gca agg ata atg atg ggc aaa gtc Ile Glu Pro Arg Gln His Ala Ser Ala Arg Ile Met Met Gly Lys Val 330 335 340	1060
aac gct ctg att acg gtg atc gat gat att tat gat gtc tat ggc acc Asn Ala Leu Ile Thr Val Ile Asp Asp Ile Tyr Asp Val Tyr Gly Thr 345 350 355 360	1108
tta gaa gaa ctc gaa caa ttc act gac ctc att cga aga tgg gat ata Leu Glu Glu Leu Glu Gln Phe Thr Asp Leu Ile Arg Arg Trp Asp Ile 365 370 375	1156
aac tca atc gac caa ctt ccc gat tac atg caa ctg tgc ttt ctt gca Asn Ser Ile Asp Gln Leu Pro Asp Tyr Met Gln Leu Cys Phe Leu Ala 380 385 390	1204
ctc aac aac ttc gtc gat gat aca tcg tac gat gtt atg aag gag aaa Leu Asn Asn Phe Val Asp Asp Thr Ser Tyr Asp Val Met Lys Glu Lys 395 400 405	1252
ggc gtc aac gtt ata ccc tac ctg cgg caa tcg tgg gtt gat ttg gcg Gly Val Asn Val Ile Pro Tyr Leu Arg Gln Ser Trp Val Asp Leu Ala 410 415 420	1300
gat aag tat atg gta gag gca cgg tgg ttc tac ggc ggg cac aaa cca Asp Lys Tyr Met Val Glu Ala Arg Trp Phe Tyr Gly Gly His Lys Pro 425 430 435 440	1348
agt ttg gaa gag tat ttg gag aac tca tgg cag tcg ata agt ggg ccc Ser Leu Glu Glu Tyr Leu Glu Asn Ser Trp Gln Ser Ile Ser Gly Pro 445 450 455	1396
tgt atg tta acg cac ata ttc ttc cga gta aca gat tcg ttc aca aag Cys Met Leu Thr His Ile Phe Phe Arg Val Thr Asp Ser Phe Thr Lys 460 465 470	1444
gag acc gtc gac agt ttg tac aaa tac cac gat tta gtt cgt tgg tca Glu Thr Val Asp Ser Leu Tyr Lys Tyr His Asp Leu Val Arg Trp Ser 475 480 485	1492
tcc ttc gtt ctg cgg ctt gct gat gat ttg gga acc tcg gtg gaa gag Ser Phe Val Leu Arg Leu Ala Asp Asp Leu Gly Thr Ser Val Glu Glu 490 495 500	1540
gtg agc aga ggg gat gtg ccg aaa tca ctt cag tgc tac atg agt gac Val Ser Arg Gly Asp Val Pro Lys Ser Leu Gln Cys Tyr Met Ser Asp 505 510 515 520	1588
tac aat gca tcg gag gcg gag gcg cgg aag cac gtg aaa tgg ctg ata Tyr Asn Ala Ser Glu Ala Glu Ala Arg Lys His Val Lys Trp Leu Ile 525 530 535	1636

-continued

gcg gag gtg tgg aag atg aat gcg gag agg gtg tcg aag gat tct 1684
 Ala Glu Val Trp Lys Lys Met Asn Ala Glu Arg Val Ser Lys Asp Ser
 540 545 550

cca ttc ggc aaa gat ttt ata gga tgt gca gtt gat tta gga agg atg 1732
 Pro Phe Gly Lys Asp Phe Ile Gly Cys Ala Val Asp Leu Gly Arg Met
 555 560 565

gcg cag ttg atg tac cat aat gga gat ggg cac ggc aca caa cac cct 1780
 Ala Gln Leu Met Tyr His Asn Gly Asp Gly His Gly Thr Gln His Pro
 570 575 580

att ata cat caa caa atg acc aga acc tta ttc gag ccc ttt gca tga 1828
 Ile Ile His Gln Gln Met Thr Arg Thr Leu Phe Glu Pro Phe Ala
 585 590 595

gagatgatga cgagccatcg ttacttact taaattctac caaagttttt cgaaggcata 1888

gttcgtaatt tttcaagcac caataataa ggagaatcgg ctcaaacaaa cgtggcattt 1948

gccaccacgt gagcacaagg gagagtctgt cgctgtttat ggatgaacta ttcaattttt 2008

atgcatgtaa taattaagtt caagttcaag agccttctgc atatttaact atgtatttga 2068

atztatcgag tgtgattttc tgtctttggc aacatatatt tttgtcatat tgggcatttt 2128

attatgatat catacagtgt ttatggatga tatgatacta tc 2170

<210> SEQ ID NO 22
 <211> LENGTH: 599
 <212> TYPE: PRT
 <213> ORGANISM: Mentha spicata

<400> SEQUENCE: 22

Met Ala Leu Lys Val Leu Ser Val Ala Thr Gln Met Ala Ile Pro Ser
 1 5 10 15

Asn Leu Thr Thr Cys Leu Gln Pro Ser His Phe Lys Ser Ser Pro Lys
 20 25 30

Leu Leu Ser Ser Thr Asn Ser Ser Ser Arg Ser Arg Leu Arg Val Tyr
 35 40 45

Cys Ser Ser Ser Gln Leu Thr Thr Glu Arg Arg Ser Gly Asn Tyr Asn
 50 55 60

Pro Ser Arg Trp Asp Val Asn Phe Ile Gln Ser Leu Leu Ser Asp Tyr
 65 70 75 80

Lys Glu Asp Lys His Val Ile Arg Ala Ser Glu Leu Val Thr Leu Val
 85 90 95

Lys Met Glu Leu Glu Lys Glu Thr Asp Gln Ile Arg Gln Leu Glu Leu
 100 105 110

Ile Asp Asp Leu Gln Arg Met Gly Leu Ser Asp His Phe Gln Asn Glu
 115 120 125

Phe Lys Glu Ile Leu Ser Ser Ile Tyr Leu Asp His His Tyr Tyr Lys
 130 135 140

Asn Pro Phe Pro Lys Glu Glu Arg Asp Leu Tyr Ser Thr Ser Leu Ala
 145 150 155 160

Phe Arg Leu Leu Arg Glu His Gly Phe Gln Val Ala Gln Glu Val Phe
 165 170 175

Asp Ser Phe Lys Asn Glu Glu Gly Glu Phe Lys Glu Ser Leu Ser Asp
 180 185 190

Asp Thr Arg Gly Leu Leu Gln Leu Tyr Glu Ala Ser Phe Leu Leu Thr
 195 200 205

Glu Gly Glu Thr Thr Leu Glu Ser Ala Arg Glu Phe Ala Thr Lys Phe
 210 215 220

-continued

Leu Glu Glu Lys Val Asn Glu Gly Gly Val Asp Gly Asp Leu Leu Thr
 225 230 235 240
 Arg Ile Ala Tyr Ser Leu Asp Ile Pro Leu His Trp Arg Ile Lys Arg
 245 250 255
 Pro Asn Ala Pro Val Trp Ile Glu Trp Tyr Arg Lys Arg Pro Asp Met
 260 265 270
 Asn Pro Val Val Leu Glu Leu Ala Ile Leu Asp Leu Asn Ile Val Gln
 275 280 285
 Ala Gln Phe Gln Glu Glu Leu Lys Glu Ser Phe Arg Trp Trp Arg Asn
 290 295 300
 Thr Gly Phe Val Glu Lys Leu Pro Phe Ala Arg Asp Arg Leu Val Glu
 305 310 315 320
 Cys Tyr Phe Trp Asn Thr Gly Ile Ile Glu Pro Arg Gln His Ala Ser
 325 330 335
 Ala Arg Ile Met Met Gly Lys Val Asn Ala Leu Ile Thr Val Ile Asp
 340 345 350
 Asp Ile Tyr Asp Val Tyr Gly Thr Leu Glu Glu Leu Glu Gln Phe Thr
 355 360 365
 Asp Leu Ile Arg Arg Trp Asp Ile Asn Ser Ile Asp Gln Leu Pro Asp
 370 375 380
 Tyr Met Gln Leu Cys Phe Leu Ala Leu Asn Asn Phe Val Asp Asp Thr
 385 390 395 400
 Ser Tyr Asp Val Met Lys Glu Lys Gly Val Asn Val Ile Pro Tyr Leu
 405 410 415
 Arg Gln Ser Trp Val Asp Leu Ala Asp Lys Tyr Met Val Glu Ala Arg
 420 425 430
 Trp Phe Tyr Gly Gly His Lys Pro Ser Leu Glu Glu Tyr Leu Glu Asn
 435 440 445
 Ser Trp Gln Ser Ile Ser Gly Pro Cys Met Leu Thr His Ile Phe Phe
 450 455 460
 Arg Val Thr Asp Ser Phe Thr Lys Glu Thr Val Asp Ser Leu Tyr Lys
 465 470 475 480
 Tyr His Asp Leu Val Arg Trp Ser Ser Phe Val Leu Arg Leu Ala Asp
 485 490 495
 Asp Leu Gly Thr Ser Val Glu Glu Val Ser Arg Gly Asp Val Pro Lys
 500 505 510
 Ser Leu Gln Cys Tyr Met Ser Asp Tyr Asn Ala Ser Glu Ala Glu Ala
 515 520 525
 Arg Lys His Val Lys Trp Leu Ile Ala Glu Val Trp Lys Lys Met Asn
 530 535 540
 Ala Glu Arg Val Ser Lys Asp Ser Pro Phe Gly Lys Asp Phe Ile Gly
 545 550 555 560
 Cys Ala Val Asp Leu Gly Arg Met Ala Gln Leu Met Tyr His Asn Gly
 565 570 575
 Asp Gly His Gly Thr Gln His Pro Ile Ile His Gln Gln Met Thr Arg
 580 585 590
 Thr Leu Phe Glu Pro Phe Ala
 595

<210> SEQ ID NO 23

<211> LENGTH: 1967

<212> TYPE: DNA

<213> ORGANISM: Salvia officinalis

<220> FEATURE:

<221> NAME/KEY: CDS

-continued

<222> LOCATION: (13)...(1785)

<223> OTHER INFORMATION: 1,8-cineole synthase

<400> SEQUENCE: 23

```

gatcaccaca ag atg tcg agt ctt ata atg caa gtt gtt att cct aag cca      51
          Met Ser Ser Leu Ile Met Gln Val Val Ile Pro Lys Pro
            1             5             10

gcc aaa att ttt cac aat aac tta ttc agc gtg att tca aaa cga cat      99
Ala Lys Ile Phe His Asn Asn Leu Phe Ser Val Ile Ser Lys Arg His
      15             20             25

cgt ttc agt act aca atc acc act cgt ggt ggc agg tgg gca cat tgc      147
Arg Phe Ser Thr Thr Ile Thr Arg Gly Gly Arg Trp Ala His Cys
      30             35             40             45

tca cta caa atg ggt aat gag atc caa act gga cga cga act gga ggc      195
Ser Leu Gln Met Gly Asn Glu Ile Gln Thr Gly Arg Arg Thr Gly Gly
            50             55             60

tac cag cct acc ctt tgg gat ttc agc acc att caa ttg ttc gac tct      243
Tyr Gln Pro Thr Leu Trp Asp Phe Ser Thr Ile Gln Leu Phe Asp Ser
            65             70             75

gag tat aag gaa gag aag cac ttg atg agg gcc gca ggt atg ata gcc      291
Glu Tyr Lys Glu Glu Lys His Leu Met Arg Ala Ala Gly Met Ile Ala
            80             85             90

caa gtg aat atg ttg ttg cag gaa gaa gta gat tcg att caa cgg ttg      339
Gln Val Asn Met Leu Leu Gln Glu Glu Val Asp Ser Ile Gln Arg Leu
            95             100            105

gag ttg att gat gac cta cga agg ctg ggt ata tct tgc cat ttt gac      387
Glu Leu Ile Asp Asp Leu Arg Arg Leu Gly Ile Ser Cys His Phe Asp
      110            115            120            125

cgc gag atc gtt gaa ata tta aac tca aaa tat tat acc aac aat gag      435
Arg Glu Ile Val Glu Ile Leu Asn Ser Lys Tyr Tyr Thr Asn Asn Glu
            130            135            140

ata gat gaa agt gat cta tac tca aca gcc ctt aga ttc aag ctc cta      483
Ile Asp Glu Ser Asp Leu Tyr Ser Thr Ala Leu Arg Phe Lys Leu Leu
            145            150            155

aga caa tac gat ttt agc gtc tct caa gag gta ttt gat tgt ttc aag      531
Arg Gln Tyr Asp Phe Ser Val Ser Gln Glu Val Phe Asp Cys Phe Lys
            160            165            170

aat gac aag ggt act gat ttc aag cca agc cta gtc gat gat act aga      579
Asn Asp Lys Gly Thr Asp Phe Lys Pro Ser Leu Val Asp Asp Thr Arg
            175            180            185

gga ttg tta caa ttg tac gaa gct tcg ttt tta tca gca caa ggc gaa      627
Gly Leu Leu Gln Leu Tyr Glu Ala Ser Phe Leu Ser Ala Gln Gly Glu
      190             195             200             205

gaa acc cta cat ctt gcc aga gat ttt gct act aaa ttt ctg cat aaa      675
Glu Thr Leu His Leu Ala Arg Asp Phe Ala Thr Lys Phe Leu His Lys
            210            215            220

aga gta cta gtt gat aaa gac att aat ctc tta tca tca att gaa cgt      723
Arg Val Leu Val Asp Lys Asp Ile Asn Leu Leu Ser Ser Ile Glu Arg
            225            230            235

cgc ttg gag ttg cct act cat tgg agg gtt caa atg ccc aac gca aga      771
Ala Leu Glu Leu Pro Thr His Trp Arg Val Gln Met Pro Asn Ala Arg
            240            245            250

tcc ttc att gat gct tat aag agg aga ccc gac atg aat ccg act gtg      819
Ser Phe Ile Asp Ala Tyr Lys Arg Arg Pro Asp Met Asn Pro Thr Val
            255            260            265

cta gaa cta gct aaa ttg gac ttc aat atg gtt caa gca caa ttt caa      867
Leu Glu Leu Ala Lys Leu Asp Phe Asn Met Val Gln Ala Gln Phe Gln
      270             275             280             285

caa gag ctc aaa gag gcc tct agg tgg tgg aat agt acg ggt ctt gtc      915
Gln Glu Leu Lys Glu Ala Ser Arg Trp Trp Asn Ser Thr Gly Leu Val

```

-continued

290		295		300		
cac gag ctt ccc ttt gtg aga gat agg att gtg gaa tgc tac tac tgg						963
His Glu Leu Pro Phe Val Arg Asp Arg Ile Val Glu Cys Tyr Tyr Trp	305		310		315	
acg aca gga gtg gtt gag cgt cgt gaa cat gga tac gag agg ata atg						1011
Thr Thr Gly Val Val Glu Arg Arg Glu His Gly Tyr Glu Arg Ile Met	320		325		330	
ctc acc aaa ata aat gct ctt gtt aca aca ata gac gat gtc ttt gat						1059
Leu Thr Lys Ile Asn Ala Leu Val Thr Thr Ile Asp Asp Val Phe Asp	335		340		345	
att tat ggt acg ctt gaa gag cta caa cta ttc aca act gct att caa						1107
Ile Tyr Gly Thr Leu Glu Glu Leu Gln Leu Phe Thr Thr Ala Ile Gln	350		355		360	365
aga tgg gat att gaa tca atg aag caa ctc cct cct tac atg caa ata						1155
Arg Trp Asp Ile Glu Ser Met Lys Gln Leu Pro Pro Tyr Met Gln Ile	370		375		380	
tgt tat ctt gct ctc ttc aac ttt gtg aat gag atg gct tat gat act						1203
Cys Tyr Leu Ala Leu Phe Asn Phe Val Asn Glu Met Ala Tyr Asp Thr	385		390		395	
ctt agg gat aaa ggt ttc aac tcc acc cca tat cta cga aaa gcg tgg						1251
Leu Arg Asp Lys Gly Phe Asn Ser Thr Pro Tyr Leu Arg Lys Ala Trp	400		405		410	
gtt gat ttg gtt gag tca tat cta ata gag gca aag tgg tac tac atg						1299
Val Asp Leu Val Glu Ser Tyr Leu Ile Glu Ala Lys Trp Tyr Tyr Met	415		420		425	
gga cat aaa cct agt ttg gaa gaa tat atg aag aat agt tgg ata tca						1347
Gly His Lys Pro Ser Leu Glu Glu Tyr Met Lys Asn Ser Trp Ile Ser	430		435		440	445
atc gga ggc atc ccc att cta tct cat cta ttt ttc cgg cta aca gat						1395
Ile Gly Gly Ile Pro Ile Leu Ser His Leu Phe Phe Arg Leu Thr Asp	450		455		460	
tcg att gag gaa gag gat gct gag agt atg cat aaa tac cat gat att						1443
Ser Ile Glu Glu Glu Asp Ala Glu Ser Met His Lys Tyr His Asp Ile	465		470		475	
gtt cgt gca tcg tgt act att cta agg ctt gct gat gat atg gga aca						1491
Val Arg Ala Ser Cys Thr Ile Leu Arg Leu Ala Asp Asp Met Gly Thr	480		485		490	
tcg ctg gat gag gtg gag aga ggc gac gtg ccc aaa tca gtt cag tgc						1539
Ser Leu Asp Glu Val Glu Arg Gly Asp Val Pro Lys Ser Val Gln Cys	495		500		505	
tac atg aat gag aag aat gct tcg gaa gaa gaa gcg cga gag cat gtg						1587
Tyr Met Asn Glu Lys Asn Ala Ser Glu Glu Glu Ala Arg Glu His Val	510		515		520	525
cga tca ctc ata gac caa aca tgg aag atg atg aac aag gaa atg atg						1635
Arg Ser Leu Ile Asp Gln Thr Trp Lys Met Met Asn Lys Glu Met Met	530		535		540	
acg tca tca ttt tcc aaa tat ttt gta caa gtt tct gct aat ctt gca						1683
Thr Ser Ser Phe Ser Lys Tyr Phe Val Gln Val Ser Ala Asn Leu Ala	545		550		555	
aga atg gcg caa tgg ata tac cag cat gaa tct gat gga ttt ggc atg						1731
Arg Met Ala Gln Trp Ile Tyr Gln His Glu Ser Asp Gly Phe Gly Met	560		565		570	
caa cat tca ttg gtg aac aaa atg ctc aga ggg ttg ttg ttc gac cgc						1779
Gln His Ser Leu Val Asn Lys Met Leu Arg Gly Leu Leu Phe Asp Arg	575		580		585	
tat gag taa ctaatcttcg cccgggttcc aaatgaatca atctgtttgtg						1828
Tyr Glu	590					
ttgtctgttcc acctgatatc aataataatt agacaaatgt ttctgtacgg gtggccaac						1888

-continued

cgtcaggccc atttcgctca tgttcataat aaataataaa actgttaatc aataacaaaa 1948
 aaaaaaaaaa aaaaaaaaaa 1967

<210> SEQ ID NO 24
 <211> LENGTH: 591
 <212> TYPE: PRT
 <213> ORGANISM: *Salvia officinalis*

<400> SEQUENCE: 24

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

-continued

Thr Leu Glu Glu Leu Gln Leu Phe Thr Thr Ala Ile Gln Arg Trp Asp
 355 360 365
 Ile Glu Ser Met Lys Gln Leu Pro Pro Tyr Met Gln Ile Cys Tyr Leu
 370 375 380
 Ala Leu Phe Asn Phe Val Asn Glu Met Ala Tyr Asp Thr Leu Arg Asp
 385 390 395 400
 Lys Gly Phe Asn Ser Thr Pro Tyr Leu Arg Lys Ala Trp Val Asp Leu
 405 410 415
 Val Glu Ser Tyr Leu Ile Glu Ala Lys Trp Tyr Tyr Met Gly His Lys
 420 425 430
 Pro Ser Leu Glu Glu Tyr Met Lys Asn Ser Trp Ile Ser Ile Gly Gly
 435 440 445
 Ile Pro Ile Leu Ser His Leu Phe Phe Arg Leu Thr Asp Ser Ile Glu
 450 455 460
 Glu Glu Asp Ala Glu Ser Met His Lys Tyr His Asp Ile Val Arg Ala
 465 470 475 480
 Ser Cys Thr Ile Leu Arg Leu Ala Asp Asp Met Gly Thr Ser Leu Asp
 485 490 495
 Glu Val Glu Arg Gly Asp Val Pro Lys Ser Val Gln Cys Tyr Met Asn
 500 505 510
 Glu Lys Asn Ala Ser Glu Glu Glu Ala Arg Glu His Val Arg Ser Leu
 515 520 525
 Ile Asp Gln Thr Trp Lys Met Met Asn Lys Glu Met Met Thr Ser Ser
 530 535 540
 Phe Ser Lys Tyr Phe Val Gln Val Ser Ala Asn Leu Ala Arg Met Ala
 545 550 555 560
 Gln Trp Ile Tyr Gln His Glu Ser Asp Gly Phe Gly Met Gln His Ser
 565 570 575
 Leu Val Asn Lys Met Leu Arg Gly Leu Leu Phe Asp Arg Tyr Glu
 580 585 590

<210> SEQ ID NO 25

<211> LENGTH: 2024

<212> TYPE: DNA

<213> ORGANISM: Salvia officinalis

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (11)...(1804)

<223> OTHER INFORMATION: (+)-bornyl diphosphate synthase

<400> SEQUENCE: 25

gatcacaaaa atg tct atc att agc atg aac gta tcg atc ctt agc aag 49
 Met Ser Ile Ile Ser Met Asn Val Ser Ile Leu Ser Lys
 1 5 10
 cca cta aat tgc ctc cac aac ttg gag agg aga cct tca aaa gcc ttg 97
 Pro Leu Asn Cys Leu His Asn Leu Glu Arg Arg Pro Ser Lys Ala Leu
 15 20 25
 ctt gtc cct tgc act gca ccc acc gct cgc ctc cgg gca tct tgc tcc 145
 Leu Val Pro Cys Thr Ala Pro Thr Ala Arg Leu Arg Ala Ser Cys Ser
 30 35 40 45
 tca aaa cta caa gaa gct cat caa atc cga cga tct gga aac tac caa 193
 Ser Lys Leu Gln Glu Ala His Gln Ile Arg Arg Ser Gly Asn Tyr Gln
 50 55 60
 cct gcc ctt tgg gat tcc aat tac att cag tct ctc aat act cca tat 241
 Pro Ala Leu Trp Asp Ser Asn Tyr Ile Gln Ser Leu Asn Thr Pro Tyr
 65 70 75
 acg gag gag agg cac ttg gat aga aaa gca gag ctg att gtg caa gtg 289

-continued

```

gat gca gca tat gat att ctc aaa gaa cat ggt ttc ttt tgt ctc caa 1249
Asp Ala Ala Tyr Asp Ile Leu Lys Glu His Gly Phe Phe Cys Leu Gln
      400                      405                      410

tat ctc cgg aaa tcg gtg gta gat ttg gtt gaa gca tat ttt cac gag 1297
Tyr Leu Arg Lys Ser Val Val Asp Leu Val Glu Ala Tyr Phe His Glu
      415                      420                      425

gca aag tgg tac cac agc ggt tat aca cca agc ctg gat gaa tat ctc 1345
Ala Lys Trp Tyr His Ser Gly Tyr Thr Pro Ser Leu Asp Glu Tyr Leu
      430                      435                      440                      445

aac atc gcc aag att tca gtg gcg tct cct gca ata ata tcc cca acc 1393
Asn Ile Ala Lys Ile Ser Val Ala Ser Pro Ala Ile Ile Ser Pro Thr
      450                      455                      460

tat ttc aca ttc gca aac gcg tct cat gac aca gca gtc atc gac agc 1441
Tyr Phe Thr Phe Ala Asn Ala Ser His Asp Thr Ala Val Ile Asp Ser
      465                      470                      475

ttg tac caa tat cat gac ata ctt tgc cta gca gga att att ttg agg 1489
Leu Tyr Gln Tyr His Asp Ile Leu Cys Leu Ala Gly Ile Ile Leu Arg
      480                      485                      490

ctt ccc gac gat ctt ggg aca tca tat ttt gag ctg gcg aga ggc gac 1537
Leu Pro Asp Asp Leu Gly Thr Ser Tyr Phe Glu Leu Ala Arg Gly Asp
      495                      500                      505

gtg cgg aaa aca atc cag tgc tac atg aag gaa aca aat gct agt gag 1585
Val Pro Lys Thr Ile Gln Cys Tyr Met Lys Glu Thr Asn Ala Ser Glu
      510                      515                      520                      525

gag gag gcg gtg gag cac gtg aag ttt ctg ata agg gag gcg tgg aag 1633
Glu Glu Ala Val Glu His Val Lys Phe Leu Ile Arg Glu Ala Trp Lys
      530                      535                      540

gat atg aac acg gcc ata gca gcc ggt tat ccg ttt ccg gat ggt atg 1681
Asp Met Asn Thr Ala Ile Ala Ala Gly Tyr Pro Phe Pro Asp Gly Met
      545                      550                      555

gtg gcg ggc gca gct aat att ggg cgc gtg gcg cag ttt att tat ctc 1729
Val Ala Gly Ala Ala Asn Ile Gly Arg Val Ala Gln Phe Ile Tyr Leu
      560                      565                      570

cac gga gat ggg ttt ggc gtg caa cac tcg aaa acg tac gag cat atc 1777
His Gly Asp Gly Phe Gly Val Gln His Ser Lys Thr Tyr Glu His Ile
      575                      580                      585

gcc ggc cta ctg ttc gag cct tat gca tga acaaatggga gactgcttga 1827
Ala Gly Leu Leu Phe Glu Pro Tyr Ala
      590                      595

tatatatataa tttggcacac caataattgc atgttatata tgttgaaaa taagtgtctg 1887

gttgagatgt catgtggtgt attatctaaa taattcaagg ttgocctggt tatgtagccg 1947

gtggtgcaac tacctcccat tcaaatcaat taaatctaaa cagtcgagtc aagctcgagc 2007

tcgaggaaaa aaaaaaa 2024
    
```

<210> SEQ ID NO 26
 <211> LENGTH: 598
 <212> TYPE: PRT
 <213> ORGANISM: Salvia officinalis

<400> SEQUENCE: 26

```

Met Ser Ile Ile Ser Met Asn Val Ser Ile Leu Ser Lys Pro Leu Asn
 1           5           10           15

Cys Leu His Asn Leu Glu Arg Arg Pro Ser Lys Ala Leu Leu Val Pro
      20           25           30

Cys Thr Ala Pro Thr Ala Arg Leu Arg Ala Ser Cys Ser Ser Lys Leu
      35           40           45

Gln Glu Ala His Gln Ile Arg Arg Ser Gly Asn Tyr Gln Pro Ala Leu
      50           55           60
    
```

-continued

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

-continued

Tyr His Asp Ile Leu Cys Leu Ala Gly Ile Ile Leu Arg Leu Pro Asp
 485 490 495

Asp Leu Gly Thr Ser Tyr Phe Glu Leu Ala Arg Gly Asp Val Pro Lys
 500 505 510

Thr Ile Gln Cys Tyr Met Lys Glu Thr Asn Ala Ser Glu Glu Glu Ala
 515 520 525

Val Glu His Val Lys Phe Leu Ile Arg Glu Ala Trp Lys Asp Met Asn
 530 535 540

Thr Ala Ile Ala Ala Gly Tyr Pro Phe Pro Asp Gly Met Val Ala Gly
 545 550 555 560

Ala Ala Asn Ile Gly Arg Val Ala Gln Phe Ile Tyr Leu His Gly Asp
 565 570 575

Gly Phe Gly Val Gln His Ser Lys Thr Tyr Glu His Ile Ala Gly Leu
 580 585 590

Leu Phe Glu Pro Tyr Ala
 595

<210> SEQ ID NO 27
 <211> LENGTH: 1959
 <212> TYPE: DNA
 <213> ORGANISM: Mentha x piperita
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (71)...(1720)
 <223> OTHER INFORMATION: (E)-B-farnesene synthase

<400> SEQUENCE: 27

aaactctgca atttcatata taacatcata aaatcagaga gagagacaga gagtttgttg 60

tagtgaaaaa atg gct aca aac ggc gtc gta att agt tgc tta agg gaa 109
 Met Ala Thr Asn Gly Val Val Ile Ser Cys Leu Arg Glu
 1 5 10

gta agg cca cct atg acg aag cat gcg cca agc atg tgg act gat acc 157
 Val Arg Pro Pro Met Thr Lys His Ala Pro Ser Met Trp Thr Asp Thr
 15 20 25

ttt tct aac ttt tct ctt gac gat aag gaa caa caa aag tgc tca gaa 205
 Phe Ser Asn Phe Ser Leu Asp Asp Lys Glu Gln Gln Lys Cys Ser Glu
 30 35 40 45

acc atc gaa gca ctt aag caa gaa gca aga ggc atg ctt atg gct gca 253
 Thr Ile Glu Ala Leu Lys Gln Glu Ala Arg Gly Met Leu Met Ala Ala
 50 55 60

acc act cct ctc caa caa atg aca cta atc gac act ctc gag cgt ttg 301
 Thr Thr Pro Leu Gln Gln Met Thr Leu Ile Asp Thr Leu Glu Arg Leu
 65 70 75

gga ttg tct ttc cat ttt gag acg gag atc gaa tac aaa atc gaa cta 349
 Gly Leu Ser Phe His Phe Glu Thr Glu Ile Glu Tyr Lys Ile Glu Leu
 80 85 90

atc aac gct gca gaa gac gac ggc ttt gat ttg ttc gct act gct ctt 397
 Ile Asn Ala Ala Glu Asp Asp Gly Phe Asp Leu Phe Ala Thr Ala Leu
 95 100 105

cgt ttc cgt ttg ctc aga caa cat caa cgc cac gtt tct tgt gat gtt 445
 Arg Phe Arg Leu Leu Arg Gln His Gln Arg His Val Ser Cys Asp Val
 110 115 120 125

ttc gac aag ttc atc gac aaa gat ggc aag ttc gaa gaa tcc ctt agc 493
 Phe Asp Lys Phe Ile Asp Lys Asp Gly Lys Phe Glu Glu Ser Leu Ser
 130 135 140

aat aat gtt gaa ggc cta tta agc ttg tat gaa gca gct cat gtt ggg 541
 Asn Asn Val Glu Gly Leu Leu Ser Leu Tyr Glu Ala Ala His Val Gly
 145 150 155

ttt cgc gaa gaa aga ata tta caa gag gct gta aat ttt acg agg cat 589

-continued

tct aag ttt gaa gga ttg gtt gag gaa aca tgg aag gat ata aac aag	1549
Ser Lys Phe Glu Gly Leu Val Glu Glu Thr Trp Lys Asp Ile Asn Lys	
480 485 490	
gaa ttc ata gcc aca act aat tat aat gtg ggt aga gaa att gcc atc	1597
Glu Phe Ile Ala Thr Thr Asn Tyr Asn Val Gly Arg Glu Ile Ala Ile	
495 500 505	
aca ttc ctc aac tac gct cgg ata tgt gaa gcc agt tac agc aaa act	1645
Thr Phe Leu Asn Tyr Ala Arg Ile Cys Glu Ala Ser Tyr Ser Lys Thr	
510 515 520 525	
gac gga gac gct tat tca gat cct aat gtt gcc aag gca aat gtc gtt	1693
Asp Gly Asp Ala Tyr Ser Asp Pro Asn Val Ala Lys Ala Asn Val Val	
530 535 540	
gct ctc ttt gtt gat gcc ata gtc ttt tga ttgcataat caaagaccct	1743
Ala Leu Phe Val Asp Ala Ile Val Phe	
545 550	
ataattataa ttatatgtgt ttaagaaact aataagcttg ctttatgtat agttgtcaat	1803
tgaataataa tgtattaatt agtagagtta agaagttata aagaataaag aggagctggt	1863
agacgtaaac aagaataat gtgtcaaaat aacttcaact ttttcaagaa taaagaattg	1923
gaagagacca atatatacaa aaaaaaaaa aaaaaa	1959
<210> SEQ ID NO 28	
<211> LENGTH: 550	
<212> TYPE: PRT	
<213> ORGANISM: Mentha x piperita	
<400> SEQUENCE: 28	
Met Ala Thr Asn Gly Val Val Ile Ser Cys Leu Arg Glu Val Arg Pro	
1 5 10 15	
Pro Met Thr Lys His Ala Pro Ser Met Trp Thr Asp Thr Phe Ser Asn	
20 25 30	
Phe Ser Leu Asp Asp Lys Glu Gln Gln Lys Cys Ser Glu Thr Ile Glu	
35 40 45	
Ala Leu Lys Gln Glu Ala Arg Gly Met Leu Met Ala Ala Thr Thr Pro	
50 55 60	
Leu Gln Gln Met Thr Leu Ile Asp Thr Leu Glu Arg Leu Gly Leu Ser	
65 70 75 80	
Phe His Phe Glu Thr Glu Ile Glu Tyr Lys Ile Glu Leu Ile Asn Ala	
85 90 95	
Ala Glu Asp Asp Gly Phe Asp Leu Phe Ala Thr Ala Leu Arg Phe Arg	
100 105 110	
Leu Leu Arg Gln His Gln Arg His Val Ser Cys Asp Val Phe Asp Lys	
115 120 125	
Phe Ile Asp Lys Asp Gly Lys Phe Glu Glu Ser Leu Ser Asn Asn Val	
130 135 140	
Glu Gly Leu Leu Ser Leu Tyr Glu Ala Ala His Val Gly Phe Arg Glu	
145 150 155 160	
Glu Arg Ile Leu Gln Glu Ala Val Asn Phe Thr Arg His His Leu Glu	
165 170 175	
Gly Ala Glu Leu Asp Gln Ser Pro Leu Leu Ile Arg Glu Lys Val Lys	
180 185 190	
Arg Ala Leu Glu His Pro Leu His Arg Asp Phe Pro Ile Val Tyr Ala	
195 200 205	
Arg Leu Phe Ile Ser Ile Tyr Glu Lys Asp Asp Ser Arg Asp Glu Leu	
210 215 220	
Leu Leu Lys Leu Ser Lys Val Asn Phe Lys Phe Met Gln Asn Leu Tyr	

-continued

```

225             230             235             240
Lys Glu Glu Leu Ser  Gln Leu Ser Arg Trp Trp Asn Thr Trp Asn Leu
             245             250             255
Lys Ser Lys Leu Pro  Tyr Ala Arg Asp Arg Val Val Glu Ala Tyr Val
             260             265             270
Trp Gly Val Gly Tyr His Tyr Glu Pro Gln Tyr Ser Tyr Val Arg Met
             275             280             285
Gly Leu Ala Lys Gly Val Leu Ile Cys Gly Ile Met Asp Asp Thr Tyr
             290             295             300
Asp Asn Tyr Ala Thr  Leu Asn Glu Ala Gln Leu Phe Thr Gln Val Leu
             305             310             315
Asp Lys Trp Asp Arg Asp Glu Ala Glu Arg Leu Pro Glu Tyr Met Lys
             325             330             335
Ile Val Tyr Arg Phe Ile Leu Ser Ile Tyr Glu Asn Tyr Glu Arg Asp
             340             345             350
Ala Ala Lys Leu Gly Lys Ser Phe Ala Ala Pro Tyr Phe Lys Glu Thr
             355             360             365
Val Lys Gln Leu Ala Arg Ala Phe Asn Glu Glu Gln Lys Trp Val Met
             370             375             380
Glu Arg Gln Leu Pro  Ser Phe Gln Asp Tyr Val Lys Asn Ser Glu Lys
             385             390             395
Thr Ser Cys Ile Tyr Thr Met Phe Ala Ser Ile Ile Pro Gly Leu Lys
             405             410             415
Ser Val Thr Gln Glu Thr Ile Asp Trp Ile Lys Ser Glu Pro Thr Leu
             420             425             430
Ala Thr Ser Thr Ala Met Ile Gly Arg Tyr Trp Asn Asp Thr Ser Ser
             435             440             445
Gln Leu Arg Glu Ser Lys Gly Gly Glu Met Leu Thr Ala Leu Asp Phe
             450             455             460
His Met Lys Glu Tyr Gly Leu Thr Lys Glu Glu Ala Ala Ser Lys Phe
             465             470             475
Glu Gly Leu Val Glu Glu Thr Trp Lys Asp Ile Asn Lys Glu Phe Ile
             485             490             495
Ala Thr Thr Asn Tyr Asn Val Gly Arg Glu Ile Ala Ile Thr Phe Leu
             500             505             510
Asn Tyr Ala Arg Ile Cys Glu Ala Ser Tyr Ser Lys Thr Asp Gly Asp
             515             520             525
Ala Tyr Ser Asp Pro Asn Val Ala Lys Ala Asn Val Val Ala Leu Phe
             530             535             540
Val Asp Ala Ile Val Phe
545             550

```

```

<210> SEQ ID NO 29
<211> LENGTH: 2196
<212> TYPE: DNA
<213> ORGANISM: Abies grandis
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (69)...(1949)
<223> OTHER INFORMATION: myrcene synthase

```

<400> SEQUENCE: 29

```

tgccggcagc aggttatctt gagcttcctc catataggcc aacacatata ataatcaagg   60
gagcaaga atg gct ctg gtt tct atc tca cgc ttg gct tog aaa tct tgc   110
   Met Ala Leu Val Ser Ile Ser Pro Leu Ala Ser Lys Ser Cys
         1             5             10

```

-continued

ctg cgc aag tcg ttg atc agt tca att cat gaa cat aag cct ccc tat	158
Leu Arg Lys Ser Leu Ile Ser Ser Ile His Glu His Lys Pro Pro Tyr	
15 20 25 30	
aga aca atc cca aat ctt gga atg cgt agg cga ggg aaa tct gtc acg	206
Arg Thr Ile Pro Asn Leu Gly Met Arg Arg Arg Gly Lys Ser Val Thr	
35 40 45	
cct tcc atg agc atc agt ttg gcc acc gct gca cct gat gat ggt gta	254
Pro Ser Met Ser Ile Ser Leu Ala Thr Ala Ala Pro Asp Asp Gly Val	
50 55 60	
caa aga cgc ata ggt gac tac cat tcc aat atc tgg gac gat gat ttc	302
Gln Arg Arg Ile Gly Asp Tyr His Ser Asn Ile Trp Asp Asp Asp Phe	
65 70 75	
ata cag tct cta tca acg cct tat ggg gaa ccc tct tac cag gaa cgt	350
Ile Gln Ser Leu Ser Thr Pro Tyr Gly Glu Pro Ser Tyr Gln Glu Arg	
80 85 90	
gct gag aga tta att gtg gag gta aag aag ata ttc aat tca atg tac	398
Ala Glu Arg Leu Ile Val Glu Val Lys Lys Ile Phe Asn Ser Met Tyr	
95 100 105 110	
ctg gat gat gga aga tta atg agt tcc ttt aat gat ctc atg caa cgc	446
Leu Asp Asp Gly Arg Leu Met Ser Ser Phe Asn Asp Leu Met Gln Arg	
115 120 125	
ctt tgg ata gtc gat agc gtt gaa cgt ttg ggg ata gct aga cat ttc	494
Leu Trp Ile Val Asp Ser Val Glu Arg Leu Gly Ile Ala Arg His Phe	
130 135 140	
aag aac gag ata aca tca gct ctg gat tat gtt ttc cgt tac tgg gag	542
Lys Asn Glu Ile Thr Ser Ala Leu Asp Tyr Val Phe Arg Tyr Trp Glu	
145 150 155	
gaa aac ggc att gga tgt ggg aga gac agt att gtt act gat ctc aac	590
Glu Asn Gly Ile Gly Cys Gly Arg Asp Ser Ile Val Thr Asp Leu Asn	
160 165 170	
tca act gcg ttg ggg ttt cga act ctt cga tta cac ggg tac act gta	638
Ser Thr Ala Leu Gly Phe Arg Thr Leu Arg Leu His Gly Tyr Thr Val	
175 180 185 190	
tct cca gag gtt tta aaa gct ttt caa gat caa aat gga cag ttt gta	686
Ser Pro Glu Val Leu Lys Ala Phe Gln Asp Gln Asn Gly Gln Phe Val	
195 200 205	
tgc tcc ccc ggt cag aca gag ggt gag atc aga agc gtt ctt aac tta	734
Cys Ser Pro Gly Gln Thr Glu Gly Ile Arg Ser Val Leu Asn Leu	
210 215 220	
tat cgg gct tcc ctc att gcc ttc cct ggt gag aaa gtt atg gaa gaa	782
Tyr Arg Ala Ser Leu Ile Ala Phe Pro Gly Glu Lys Val Met Glu Glu	
225 230 235	
gct gaa atc ttc tcc aca aga tat ttg aaa gaa gct cta caa aag att	830
Ala Glu Ile Phe Ser Thr Arg Tyr Leu Lys Glu Ala Leu Gln Lys Ile	
240 245 250	
cca gtc tcc gct ctt tca caa gag ata aag ttt gtt atg gaa tat ggc	878
Pro Val Ser Ala Leu Ser Gln Glu Ile Lys Phe Val Met Glu Tyr Gly	
255 260 265 270	
tgg cac aca aat ttg cca aga ttg gaa gca aga aat tac ata gac aca	926
Trp His Thr Asn Leu Pro Arg Leu Glu Ala Arg Asn Tyr Ile Asp Thr	
275 280 285	
ctt gag aaa gac acc agt gca tgg ctc aat aaa aat gct ggg aag aag	974
Leu Glu Lys Asp Thr Ser Ala Trp Leu Asn Lys Asn Ala Gly Lys Lys	
290 295 300	
ctt tta gaa ctt gca aaa ttg gag ttc aat ata ttt aac tcc tta caa	1022
Leu Leu Glu Leu Ala Lys Leu Glu Phe Asn Ile Phe Asn Ser Leu Gln	
305 310 315	
caa aag gaa tta caa tat ctt ttg aga tgg tgg aaa gag tcg gat ttg	1070
Gln Lys Glu Leu Gln Tyr Leu Leu Arg Trp Trp Lys Glu Ser Asp Leu	

-continued

320	325	330	
cct aaa ttg aca ttt Pro Lys Leu Thr Phe 335	gct cgg cat cgt Ala Arg His Arg 340	cat gtg gaa ttc tac act ttg His Val Glu Phe Tyr Thr Leu 345	1118
gcc tct tgt att gcc att Ala Ser Cys Ile Ala Ile 355	gac cca aaa cat Asp Pro Lys His 360	tct gca ttc aga cta ggc Ser Ala Phe Arg Leu Gly 365	1166
ttc gcc aaa atg tgt cat Phe Ala Lys Met Cys His 370	ctt gtc aca gtt ttg gac Leu Val Thr Val Leu Asp 375	gat att tac gac Asp Ile Tyr Asp 380	1214
act ttt gga acg att gac Thr Phe Gly Thr Ile Asp 385	gag gct ctt gaa ctc Glu Leu Glu Leu Phe Thr 390	tca tct gca att aag Ser Ala Ile Lys 395	1262
aga tgg aat tca tca gag Arg Trp Asn Ser Ser Glu 400	ata gaa cac ctt cca Ile Glu His Leu Pro Glu 405	gaa tat atg aaa tgt Tyr Met Lys Cys 410	1310
gtg tac atg gtc gtg ttt Val Tyr Met Val Val Phe 415	gaa act gta aat gaa Glu Thr Val Asn Glu Leu 420	ctg aca cga gag gcg Thr Arg Glu Ala 425	1358
gag aag act caa ggg aga Glu Lys Thr Gln Gly Arg 435	aac act ctc aac tat Asn Thr Leu Asn Tyr Val 440	ggt cga aag gct tgg Arg Lys Ala Trp 445	1406
gag gct tat ttt gat tca Glu Ala Tyr Phe Asp Ser 450	tat atg gaa gaa gca Tyr Met Glu Glu Ala Lys 455	atc tct aat Trp Ile Ser Asn 460	1454
ggt tat ctg cca atg ttt Gly Tyr Leu Pro Met Phe 465	gaa gag tac cat gag Glu Tyr His Glu Asn Gly 470	aaa gtg agc Lys Val Ser 475	1502
tct gca tat cgc gta gca Ser Ala Tyr Arg Val Ala 480	aca ttg caa ccc atc ctc Thr Leu Gln Pro Ile Leu 485	act ttg aat gca Thr Leu Asn Ala 490	1550
tgg ctt cct gat tac atc Trp Leu Pro Asp Tyr Ile 495	ttg aag gga att gat Leu Lys Gly Ile Asp Phe 500	cca tcc agg ttc Pro Ser Arg Phe 505	1598
aat gat ttg gca tcg toc Asn Asp Leu Ala Ser Ser 515	ttc ctt cgg cta cga Leu Arg Leu Arg Gly Asp 520	gac aca cgc tgc Thr Arg Cys 525	1646
tac aag gcc gat agg gat Tyr Lys Ala Asp Arg Asp 530	cgt ggt gaa gaa gct Arg Gly Glu Glu Ala Ser 535	tcg tgt ata tca tgt Cys Ile Ser Cys 540	1694
tat atg aaa gac aat cct Tyr Met Lys Asp Asn Pro 545	gga tca acc gaa gaa Gly Ser Thr Glu Glu Asp 550	gac ctc aat cat Ala Leu Asn His 555	1742
atc aat gcc atg gtc aat Ile Asn Ala Met Val Asn 560	gac ata atc aaa gaa Asp Ile Ile Lys Glu Leu 565	tta aat tgg gaa ctt Asn Trp Glu Leu 570	1790
cta aga tcc aac gac aat Leu Arg Ser Asn Asp Asn 575	att cca atg ctg gcc Ile Pro Met Leu Ala Lys 580	aaa cat gct ttt Lys Lys His Ala Phe 585	1838
gac ata aca aga gct ctc Asp Ile Thr Arg Ala Leu 595	cac cat ctc tac ata His His Leu Tyr Ile Tyr 600	tat cga gat ggc ttt Arg Asp Gly Phe 605	1886
agt gtt gcc aac aag gaa Ser Val Ala Asn Lys Glu 610	aca aaa aaa ttg gtt Thr Lys Lys Leu Val Met 615	atg gaa aca ctc ctt Glu Thr Leu Leu 620	1934
gaa tct atg ctt ttt taa Glu Ser Met Leu Phe 625	ctataacat atocataata 625	ataagctcat 625	1982
aatgctaaat tattgcctt 625	atgacatagt ttatgtatgt 625	acttgtgtga attcaatcat 625	2042

-continued

atcgtgtggg tatgattaaa aagctagagc ttactaggtt agtaacatgg tgataaaagt 2102
 tataaaatgt gagttataga gataccctatg ttgaataatg aattacaaaa agagaaattt 2162
 atgtagaata agattggaag cttttcaatt gttt 2196

<210> SEQ ID NO 30

<211> LENGTH: 627

<212> TYPE: PRT

<213> ORGANISM: *Abies grandis*

<400> SEQUENCE: 30

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

-continued

tca ttc tcc ctc gac aat cag att gct gga aaa tat gct caa gag atc Ser Phe Ser Leu Asp Asn Gln Ile Ala Gly Lys Tyr Ala Gln Glu Ile 35 40 45	203
gaa act ttg aag gaa caa tca aga att ata tta tct gca tct tct cga Glu Thr Leu Lys Glu Gln Ser Arg Ile Ile Leu Ser Ala Ser Ser Arg 50 55 60 65	251
aga aca ttg gct gag aaa ttg gat ctg ata gac att gtt gag cgc ctt Arg Thr Leu Ala Glu Lys Leu Asp Leu Ile Asp Ile Val Glu Arg Leu 70 75 80	299
ggc att gct tat cat ttt gaa aaa caa ata gat gat atg ttg gat caa Gly Ile Ala Tyr His Phe Glu Lys Gln Ile Asp Asp Met Leu Asp Gln 85 90 95	347
ttt tac aaa gca gat cct aac ttt gag gct cac gag tac aat gat tta Phe Tyr Lys Ala Asp Pro Asn Phe Glu Ala His Glu Tyr Asn Asp Leu 100 105 110	395
caa act tta tcc gtt caa ttt cga cta ttg aga caa cat ggt tac aat Gln Thr Leu Ser Val Gln Phe Arg Leu Leu Arg Gln His Gly Tyr Asn 115 120 125	443
atc tcc cca aaa ctt ttt att aga ttc caa gat gca aaa ggc aaa ttt Ile Ser Pro Lys Leu Phe Ile Arg Phe Gln Asp Ala Lys Gly Lys Phe 130 135 140 145	491
aaa gaa tct ctt tgt aac gac atc aag ggt ctt ttg aac tta tac gaa Lys Glu Ser Leu Cys Asn Asp Ile Lys Gly Leu Leu Asn Leu Tyr Glu 150 155 160	539
gcc tcg cat gta agg act cat gga gaa gat att ttg gaa gag gca ctt Ala Ser His Val Arg Thr His Gly Glu Asp Ile Leu Glu Glu Ala Leu 165 170 175	587
gct ttc tct act gct cat ctt gaa tct gca gct cca cat ttg aag tca Ala Phe Ser Thr Ala His Leu Glu Ser Ala Ala Pro His Leu Lys Ser 180 185 190	635
cct ctg agt aag caa gtg aca cat gcc ctt gag caa tct ctc cat aag Pro Leu Ser Lys Gln Val Thr His Ala Leu Glu Gln Ser Leu His Lys 195 200 205	683
agc att cca aga gtt gag aca cgc tac ttc atc tct atc tac gaa gag Ser Ile Pro Arg Val Glu Thr Arg Tyr Phe Ile Ser Ile Tyr Glu Glu 210 215 220 225	731
gag gaa cag aag aat gat gtg ttg ctt caa ttt gca aaa ctg gac ttc Glu Glu Gln Lys Asn Asp Val Leu Leu Gln Phe Ala Lys Leu Asp Phe 230 235 240	779
aac tta ctt cag atg ttg cac aaa caa gaa ctt agt gaa gta tca agg Asn Leu Leu Gln Met Leu His Lys Gln Glu Leu Ser Glu Val Ser Arg 245 250 255	827
tgg tgg aaa gat ttg gat ttt gtg aca aca ctt cca tat gct agg gat Trp Trp Lys Asp Leu Asp Phe Val Thr Thr Leu Pro Tyr Ala Arg Asp 260 265 270	875
aga gca gtg gaa tgc tac ttt tgg acg atg ggg gtg tat gct gaa cct Arg Ala Val Glu Cys Tyr Phe Trp Thr Met Gly Val Tyr Ala Glu Pro 275 280 285	923
caa tac tct cag gct cgt gtc atg ctt gct aag act ata gca atg att Gln Tyr Ser Gln Ala Arg Val Met Leu Ala Lys Thr Ile Ala Met Ile 290 295 300 305	971
tct ata gta gat gac aca ttc gat gct tat ggc att gtc aaa gaa ctt Ser Ile Val Asp Asp Thr Phe Asp Ala Tyr Gly Ile Val Lys Glu Leu 310 315 320	1019
gag atc tac acc gat gcc ata cag agg tgg gat att agc caa att gat Glu Ile Tyr Thr Asp Ala Ile Gln Arg Trp Asp Ile Ser Gln Ile Asp 325 330 335	1067
cgg ctc cct gat tac atg aaa atc agt tac aaa gca ctt tta gat ctc Arg Leu Pro Asp Tyr Met Lys Ile Ser Tyr Lys Ala Leu Leu Asp Leu 1115	1115

-continued

340	345	350	
tac aat gat tat gaa atg gag ttg tcc aag gat ggt aga tct gat gtt			1163
Tyr Asn Asp Tyr Glu Met Glu Leu Ser Lys Asp Gly Arg Ser Asp Val			
355	360	365	
gtt cac tac gcg aaa gaa aga atg aaa gaa atc gtg aga aac tat ttt			1211
Val His Tyr Ala Lys Glu Arg Met Lys Glu Ile Val Arg Asn Tyr Phe			
370	375	380	385
gtg gaa gca aaa tgg ttc att gaa gga tat atg ccg cca gtc tct gag			1259
Val Glu Ala Lys Trp Phe Ile Glu Gly Tyr Met Pro Pro Val Ser Glu			
	390	395	400
tat ctt agc aat gca tta gct acc agc act tat tac ttg ctt acg act			1307
Tyr Leu Ser Asn Ala Leu Ala Thr Ser Thr Tyr Tyr Leu Leu Thr Thr			
	405	410	415
aca tct tat ttg ggc atg aag tct gct aac aag caa gat ttt gaa tgg			1355
Thr Ser Tyr Leu Gly Met Lys Ser Ala Asn Lys Gln Asp Phe Glu Trp			
	420	425	430
ttg gcc aag aac cct aaa att ctt gag gct aat gtg acg tta tgc cga			1403
Leu Ala Lys Asn Pro Lys Ile Leu Glu Ala Asn Val Thr Leu Cys Arg			
	435	440	445
gtc ata gat gac ata gcc acc tat gag gtt gag aag ggt aga ggt cag			1451
Val Ile Asp Asp Ile Ala Thr Tyr Glu Val Glu Lys Gly Arg Gly Gln			
	450	455	460
att gcc act gga att gaa tgt tac atg aga gat tat ggt gta tcc aca			1499
Ile Ala Thr Gly Ile Glu Cys Tyr Met Arg Asp Tyr Gly Val Ser Thr			
	470	475	480
gaa aag gcc atg gaa aaa ttc caa gaa atg gct gag aca gca tgg aag			1547
Glu Lys Ala Met Glu Lys Phe Gln Glu Met Ala Glu Thr Ala Trp Lys			
	485	490	495
gat gta aat gaa gga atc ctt cga cca act ccc gtc tct aca gag att			1595
Asp Val Asn Glu Gly Ile Leu Arg Pro Thr Pro Val Ser Thr Glu Ile			
	500	505	510
ctc act cgc att ctc aat ctt gct cgc att atc gat gtt act tat aag			1643
Leu Thr Arg Ile Leu Asn Leu Ala Arg Ile Ile Asp Val Thr Tyr Lys			
	515	520	525
cac aat caa gat gga tac act cat ccg gaa aaa gta cta aaa cct cat			1691
His Asn Gln Asp Gly Tyr Thr His Pro Glu Lys Val Leu Lys Pro His			
	530	535	540
att att gcg ttg ttg gtg gac tct att gaa att taa atcatcgatt			1737
Ile Ile Ala Leu Leu Val Asp Ser Ile Glu Ile			
	550	555	
gttttgata tctgggagca cttgcttccc atccoctaaa attataagta ttgattgat			1797
gccttggttg tatctatgct gctagcgcct agctaagata ggagttgctg gagatacatg			1857
ttatagtgca gtgcagttaa ttccttaatt ttttttgta tcattattga cattttaaat			1917
atatatatat atatcactgc tttttat			1944

<210> SEQ ID NO 32

<211> LENGTH: 556

<212> TYPE: PRT

<213> ORGANISM: Solanum tuberosum

<400> SEQUENCE: 32

Met Thr Pro Ala Ala Val Val Met Ser Asn Tyr Gly Glu Glu Glu Ile
1 5 10 15Val Arg Pro Ile Ala Asp Phe Ser Pro Ser Leu Trp Gly Asp Arg Phe
20 25 30His Ser Phe Ser Leu Asp Asn Gln Ile Ala Gly Lys Tyr Ala Gln Glu
35 40 45

-continued

Ile Glu Thr Leu Lys Glu Gln Ser Arg Ile Ile Leu Ser Ala Ser Ser
50 55 60

Arg Arg Thr Leu Ala Glu Lys Leu Asp Leu Ile Asp Ile Val Glu Arg
65 70 75 80

Leu Gly Ile Ala Tyr His Phe Glu Lys Gln Ile Asp Asp Met Leu Asp
85 90 95

Gln Phe Tyr Lys Ala Asp Pro Asn Phe Glu Ala His Glu Tyr Asn Asp
100 105 110

Leu Gln Thr Leu Ser Val Gln Phe Arg Leu Leu Arg Gln His Gly Tyr
115 120 125

Asn Ile Ser Pro Lys Leu Phe Ile Arg Phe Gln Asp Ala Lys Gly Lys
130 135 140

Phe Lys Glu Ser Leu Cys Asn Asp Ile Lys Gly Leu Leu Asn Leu Tyr
145 150 155 160

Glu Ala Ser His Val Arg Thr His Gly Glu Asp Ile Leu Glu Glu Ala
165 170 175

Leu Ala Phe Ser Thr Ala His Leu Glu Ser Ala Ala Pro His Leu Lys
180 185 190

Ser Pro Leu Ser Lys Gln Val Thr His Ala Leu Glu Gln Ser Leu His
195 200 205

Lys Ser Ile Pro Arg Val Glu Thr Arg Tyr Phe Ile Ser Ile Tyr Glu
210 215 220

Glu Glu Glu Gln Lys Asn Asp Val Leu Leu Gln Phe Ala Lys Leu Asp
225 230 235 240

Phe Asn Leu Leu Gln Met Leu His Lys Gln Glu Leu Ser Glu Val Ser
245 250 255

Arg Trp Trp Lys Asp Leu Asp Phe Val Thr Thr Leu Pro Tyr Ala Arg
260 265 270

Asp Arg Ala Val Glu Cys Tyr Phe Trp Thr Met Gly Val Tyr Ala Glu
275 280 285

Pro Gln Tyr Ser Gln Ala Arg Val Met Leu Ala Lys Thr Ile Ala Met
290 295 300

Ile Ser Ile Val Asp Asp Thr Phe Asp Ala Tyr Gly Ile Val Lys Glu
305 310 315 320

Leu Glu Ile Tyr Thr Asp Ala Ile Gln Arg Trp Asp Ile Ser Gln Ile
325 330 335

Asp Arg Leu Pro Asp Tyr Met Lys Ile Ser Tyr Lys Ala Leu Leu Asp
340 345 350

Leu Tyr Asn Asp Tyr Glu Met Glu Leu Ser Lys Asp Gly Arg Ser Asp
355 360 365

Val Val His Tyr Ala Lys Glu Arg Met Lys Glu Ile Val Arg Asn Tyr
370 375 380

Phe Val Glu Ala Lys Trp Phe Ile Glu Gly Tyr Met Pro Pro Val Ser
385 390 395 400

Glu Tyr Leu Ser Asn Ala Leu Ala Thr Ser Thr Tyr Tyr Leu Leu Thr
405 410 415

Thr Thr Ser Tyr Leu Gly Met Lys Ser Ala Asn Lys Gln Asp Phe Glu
420 425 430

Trp Leu Ala Lys Asn Pro Lys Ile Leu Glu Ala Asn Val Thr Leu Cys
435 440 445

Arg Val Ile Asp Asp Ile Ala Thr Tyr Glu Val Glu Lys Gly Arg Gly
450 455 460

Gln Ile Ala Thr Gly Ile Glu Cys Tyr Met Arg Asp Tyr Gly Val Ser

-continued

465		470		475		480									
Thr	Glu	Lys	Ala	Met	Glu	Lys	Phe	Gln	Glu	Met	Ala	Glu	Thr	Ala	Trp
				485					490					495	
Lys	Asp	Val	Asn	Glu	Gly	Ile	Leu	Arg	Pro	Thr	Pro	Val	Ser	Thr	Glu
			500					505					510		
Ile	Leu	Thr	Arg	Ile	Leu	Asn	Leu	Ala	Arg	Ile	Ile	Asp	Val	Thr	Tyr
		515					520					525			
Lys	His	Asn	Gln	Asp	Gly	Tyr	Thr	His	Pro	Glu	Lys	Val	Leu	Lys	Pro
	530						535				540				
His	Ile	Ile	Ala	Leu	Leu	Val	Asp	Ser	Ile	Glu	Ile				
545					550					555					

<210> SEQ ID NO 33
 <211> LENGTH: 3950
 <212> TYPE: DNA
 <213> ORGANISM: Gossypium arboreum
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (1457)...(1579)
 <223> OTHER INFORMATION: cadinene synthase
 <221> NAME/KEY: CDS
 <222> LOCATION: (1670)...(1939)
 <221> NAME/KEY: CDS
 <222> LOCATION: (2092)...(2466)
 <221> NAME/KEY: CDS
 <222> LOCATION: (2559)...(2774)
 <221> NAME/KEY: CDS
 <222> LOCATION: (2963)...(3103)
 <221> NAME/KEY: CDS
 <222> LOCATION: (3206)...(3454)
 <221> NAME/KEY: CDS
 <222> LOCATION: (3596)...(3886)

<400> SEQUENCE: 33

aatttaactt ttattaattt aaaatttaaa gatttcaaag gggttctaaa atggaattt	60
ttcgatttta agggaattgt gccagcccct agtttcgccc ttgtttgtag tgctttattt	120
taaaaaagta aatataatag aatatgtata tatatatata tataaaccaa agtgaagat	180
gaaaatttat ataatgacg gctgcaagct tcaagctcac aataaatga ttctttacca	240
tcaagaaaca ttgggtgctt atacagagaa aagaaaaact ttggctctcc tcgtagctaa	300
tattttaaca atttaatttt tatataataa attttaaca attatttcat attttttaaa	360
tatattcatg ttgaatgtag cagtatatag ttatattagt tatgctcata aattttggat	420
gcattagatt ttcccttatgt aatttgataa caatgattat ttttttact tctaacaaat	480
aattaaatat tttttgttg attcgataaa tatcattatt ttttaaatga tttaaaatat	540
aaaaataata atagattcga ccgaacgctc accctattga gtgagtatat caattattag	600
aatttaatta aaaaaggaaa ccaaatatag cgggcttaat ttgtttaat attaatttat	660
gtgtggaaat tcaacttaaa acagagtcca tggctgctaa catattatat attaaacat	720
ttcctattaa taaatttatg aacgagagtt acatccttct aaattcattt tacttagagg	780
cggagtataa tttttatgt agtagttatt cttttactat ataaataaat aaataaaatt	840
ttaatcgctt gtgtattatg attgattcag ctgaatcaa gttggaataa tattttaatt	900
tgggatccca attaatgag attggtttga ttttggttg taaatatttt ttattaattt	960
tagataaatt attggaagtt ggagtcaaaa ttgaccgtct cagctaatta tacaataat	1020
aataatatag agaaatgggt atattgctca aactcactt ttactacgtc agcaatagtc	1080
agacagactg ctaagtaaac aatgtacact caattcgagt caaacaatc ctttatccca	1140

-continued

agattctaaa ataattgtgtt tgaggcacca attttgaagg atagaaagtg gaacaaaaca	1200
aaaggatatt aaaaaacaag gaaattttctc actgtatttg catatttttc tccttccagt	1260
ataattaaaa tacgtgcaat ttacgttgta ctttggtagc tcctatctta tacctataaa	1320
tacatgcaac aattgcacac atcgtctcat ccaaaacctg tgttttaaac actaaacagt	1380
aagcaaaaggc agcaaatata tctttgaatt atttgcttcc aaaaccctac acttttcctt	1440
caacacatcc tagaaa atg gct tca caa gct tct caa gtt ctt gct tca ccc	1492
Met Ala Ser Gln Ala Ser Gln Val Leu Ala Ser Pro	
1 5 10	
cat ccc gcc att tca tcc gaa aat cga ccc aag gct gat ttt cat ccc	1540
His Pro Ala Ile Ser Ser Glu Asn Arg Pro Lys Ala Asp Phe His Pro	
15 20 25	
ggg att tgg ggt gat atg ttc atc atc tgt cct gat acg gtaatctata	1589
Gly Ile Trp Gly Asp Met Phe Ile Ile Cys Pro Asp Thr	
30 35 40	
atTTTTTct tactttctct tttatcgatt ttttaagtttt ttggagattt catggaaaag	1649
cattatacgt acttgagcag gat atc gat gct gca act gaa tta caa tat gaa	1702
Asp Ile Asp Ala Ala Thr Glu Leu Gln Tyr Glu	
45 50	
gaa tta aaa gca caa gtg agg aag atg att atg gaa cct gtt gat gat	1750
Glu Leu Lys Ala Gln Val Arg Lys Met Ile Met Glu Pro Val Asp Asp	
55 60 65	
tca aac caa aag ttg ccc ttc att gat gct gtt caa aga tta ggt gtg	1798
Ser Asn Gln Lys Leu Pro Phe Ile Asp Ala Val Gln Arg Leu Gly Val	
70 75 80	
agt tat cat ttt gag aaa gag att gaa gat gaa cta gag aat att tac	1846
Ser Tyr His Phe Glu Lys Glu Ile Glu Asp Glu Leu Glu Asn Ile Tyr	
85 90 95 100	
cgt gac acc aac aac aat gat gcg gac acc gat ctc tac act aca gct	1894
Arg Asp Thr Asn Asn Asn Asp Ala Asp Thr Asp Leu Tyr Thr Thr Ala	
105 110 115	
ctt cga ttc cgg tta ctt aga gag cat ggc ttc gat att tct tgt	1939
Leu Arg Phe Arg Leu Leu Arg Glu His Gly Phe Asp Ile Ser Cys	
120 125 130	
ggtaattaag tcttaaactt tcataactct tcttatccat ttatcaatta atattatcaa	1999
actttacatt aataatcadc tgtacaatac ttcaatatat atatatttat tgatgaaact	2059
aatgtttgat gatgattttg ggtgcttgac ca gat gca ttc aac aag ttc aaa	2112
Asp Ala Phe Asn Lys Phe Lys	
135	
gat gag gca ggg aac ttc aag gca tca ttg aca agt gat gtg caa ggg	2160
Asp Glu Ala Gly Asn Phe Lys Ala Ser Leu Thr Ser Asp Val Gln Gly	
140 145 150	
ttg ttg gaa ctt tat gaa gct tcc tat atg agg gtc cat ggg gaa gat	2208
Leu Leu Glu Leu Tyr Glu Ala Ser Tyr Met Arg Val His Gly Glu Asp	
155 160 165 170	
ata ctt gat gaa gcc att tct ttc acc act gct caa ctt aca ctt gct	2256
Ile Leu Asp Glu Ala Ile Ser Phe Thr Thr Ala Gln Leu Thr Leu Ala	
175 180 185	
cta cca act tta cac cat cct tta tcg gaa cag gtc ggc cat gcc tta	2304
Leu Pro Thr Leu His His Pro Leu Ser Glu Gln Val Gly His Ala Leu	
190 195 200	
aag cag tct atc cga agg ggc ttg cca agg gtt gag gcc cgg aat ttc	2352
Lys Gln Ser Ile Arg Arg Gly Leu Pro Arg Val Glu Ala Arg Asn Phe	
205 210 215	
att tcg ata tac caa gat tta gaa tcc cat aac aaa tcg ttg ctt caa	2400
Ile Ser Ile Tyr Gln Asp Leu Glu Ser His Asn Lys Ser Leu Leu Gln	
220 225 230	

-continued

ttt gca aag att gat ttc aac ttg ttg cag ctt ttg cat agg aaa gag	2448
Phe Ala Lys Ile Asp Phe Asn Leu Leu Gln Leu Leu His Arg Lys Glu	
235 240 245 250	
cta agt gag atc tgc agg taagtgtttg gagatcttta aagctatgaa	2496
Leu Ser Glu Ile Cys Arg	
255	
gtctaataact atttcaattg atcacacgac tgttgctgac attttatgat gcttttttta	2556
gg tgg tgg aaa gat tta gac ttt aca aga aaa cta cca ttt gca aga	2603
Trp Trp Lys Asp Leu Asp Phe Thr Arg Lys Leu Pro Phe Ala Arg	
260 265 270	
gat aga gtg gtt gaa ggc tat ttt tgg ata atg gga gtt tac ttt gaa	2651
Asp Arg Val Val Glu Gly Tyr Phe Trp Ile Met Gly Val Tyr Phe Glu	
275 280 285	
ccc caa tac tct ctt ggt aga aag atg ttg aca aaa gtc ata gca atg	2699
Pro Gln Tyr Ser Leu Gly Arg Lys Met Leu Thr Lys Val Ile Ala Met	
290 295 300	
gct tcc att gtt gat gat act tat gat tca tat gca acc tat gat gaa	2747
Ala Ser Ile Val Asp Asp Thr Tyr Asp Ser Tyr Ala Thr Tyr Asp Glu	
305 310 315	
ctc att ccc tat aca aat gca att gaa ggtgagattt tttttccttt	2794
Leu Ile Pro Tyr Thr Asn Ala Ile Glu	
320 325	
cctccaaaaa aaaaaaagt ttttgagatc cccaagaat aggggaaaat atatgttttt	2854
aaacgttagg atattcactc caacttgacg ttgctcatat tttaatggtg atagtatgaa	2914
ctaaccaggc taagttttag attcaaatta accctgaaat tgtgtttt agg tgg gat	2971
Arg Trp Asp	
330	
att aaa tgc atg aac caa ctc ccg aat tac atg aaa ata agc tac aag	3019
Ile Lys Cys Met Asn Gln Leu Pro Asn Tyr Met Lys Ile Ser Tyr Lys	
335 340 345	
gca cta tta gat gtt tat gaa gaa atg gaa cag ctg ttg gca aat caa	3067
Ala Leu Leu Asp Val Tyr Glu Glu Met Glu Gln Leu Leu Ala Asn Gln	
350 355 360	
ggg aga cag tac cga gtt gag tat gcg aaa aag gcg gtatgtaatg	3113
Gly Arg Gln Tyr Arg Val Glu Tyr Ala Lys Lys Ala	
365 370 375	
atacaatagt atgatatgct ttaatcataa acgtataaaa tttgaaaatt acattagcaa	3173
tttgcttact tttttatgcc tttaatcctc ag atg ata cgt ctt gtt caa gct	3226
Met Ile Arg Leu Val Gln Ala	
380	
tac ctt ttg gag gcc aaa tgg act cat caa aat tat aaa cca acc ttt	3274
Tyr Leu Leu Glu Ala Lys Trp Thr His Gln Asn Tyr Lys Pro Thr Phe	
385 390 395	
gag gaa ttt aga gat aat gca ttg cca acc tct ggc tat gcc atg ctt	3322
Glu Glu Phe Arg Asp Asn Ala Leu Pro Thr Ser Gly Tyr Ala Met Leu	
400 405 410	
gct ata acg gcg ttt gtc ggc atg ggc gaa gtt ata acc cct gag acc	3370
Ala Ile Thr Ala Phe Val Gly Met Gly Glu Val Ile Thr Pro Glu Thr	
415 420 425 430	
ttc aaa tgg gcc gcc agt gac ccc aag atc att aag gct tcc acc att	3418
Phe Lys Trp Ala Ala Ser Asp Pro Lys Ile Ile Lys Ala Ser Thr Ile	
435 440 445	
att tgc agg ttc atg gac gat att gct gaa cat aag gtatactata	3464
Ile Cys Arg Phe Met Asp Asp Ile Ala Glu His Lys	
450 455	
tattcatatt caagaattct aaaaatcgat tatggtatat atatgcactt aaatctatat	3524

-continued

```

catagaattg taaggcttct agggtttgca tttgctaagt taattaatat acatggttca 3584
tatgggtgca g ttc aac cat agg aga gaa gac gat tgc tca gcg atc gaa 3634
      Phe Asn His Arg Arg Glu Asp Asp Cys Ser Ala Ile Glu
      460                465                470

tgt tac atg aaa caa tat ggg gtg aca gcg cag gaa gca tac aat gaa 3682
Cys Tyr Met Lys Gln Tyr Gly Val Thr Ala Gln Glu Ala Tyr Asn Glu
      475                480                485

ttc aac aaa cac att gag agt tca tgg aaa gat gta aat gaa gag ttc 3730
Phe Asn Lys His Ile Glu Ser Ser Trp Lys Asp Val Asn Glu Glu Phe
      490                495                500

ttg aaa ccg aca gaa atg ccg aca ccc gtt ctt tgt cgt agc ctc aac 3778
Leu Lys Pro Thr Glu Met Pro Thr Pro Val Leu Cys Arg Ser Leu Asn
      505                510                515

ctt gct agg gtt atg gat gta ctt tac aga gaa ggt gac ggt tat aca 3826
Leu Ala Arg Val Met Asp Val Leu Tyr Arg Glu Gly Asp Gly Tyr Thr
      520                525                530                535

cat gtt ggg aaa gct gct aaa ggt ggg atc act tca tta ttg att gat 3874
His Val Gly Lys Ala Ala Lys Gly Gly Ile Thr Ser Leu Leu Ile Asp
      540                545                550

cca ata caa att tga aattcaacat tggcttaaga tttactatga gataaaatta 3929
Pro Ile Gln Ile
      555

ataaggtttg tacaatgaag g 3950
    
```

```

<210> SEQ ID NO 34
<211> LENGTH: 41
<212> TYPE: PRT
<213> ORGANISM: Gossypium arboreum
    
```

<400> SEQUENCE: 34

```

Met Ala Ser Gln Ala Ser Gln Val Leu Ala Ser Pro His Pro Ala Ile
 1                5                10                15

Ser Ser Glu Asn Arg Pro Lys Ala Asp Phe His Pro Gly Ile Trp Gly
      20                25                30

Asp Met Phe Ile Ile Cys Pro Asp Thr
      35                40
    
```

```

<210> SEQ ID NO 35
<211> LENGTH: 90
<212> TYPE: PRT
<213> ORGANISM: Gossypium arboreum
    
```

<400> SEQUENCE: 35

```

Asp Ile Asp Ala Ala Thr Glu Leu Gln Tyr Glu Glu Leu Lys Ala Gln
 1                5                10                15

Val Arg Lys Met Ile Met Glu Pro Val Asp Asp Ser Asn Gln Lys Leu
      20                25                30

Pro Phe Ile Asp Ala Val Gln Arg Leu Gly Val Ser Tyr His Phe Glu
      35                40                45

Lys Glu Ile Glu Asp Glu Leu Glu Asn Ile Tyr Arg Asp Thr Asn Asn
      50                55                60

Asn Asp Ala Asp Thr Asp Leu Tyr Thr Thr Ala Leu Arg Phe Arg Leu
      65                70                75                80

Leu Arg Glu His Gly Phe Asp Ile Ser Cys
      85                90
    
```

```

<210> SEQ ID NO 36
<211> LENGTH: 125
<212> TYPE: PRT
    
```

-continued

<213> ORGANISM: *Gossypium arboreum*

<400> SEQUENCE: 36

Asp Ala Phe Asn Lys Phe Lys Asp Glu Ala Gly Asn Phe Lys Ala Ser
 1 5 10 15
 Leu Thr Ser Asp Val Gln Gly Leu Leu Glu Leu Tyr Glu Ala Ser Tyr
 20 25 30
 Met Arg Val His Gly Glu Asp Ile Leu Asp Glu Ala Ile Ser Phe Thr
 35 40 45
 Thr Ala Gln Leu Thr Leu Ala Leu Pro Thr Leu His His Pro Leu Ser
 50 55 60
 Glu Gln Val Gly His Ala Leu Lys Gln Ser Ile Arg Arg Gly Leu Pro
 65 70 75 80
 Arg Val Glu Ala Arg Asn Phe Ile Ser Ile Tyr Gln Asp Leu Glu Ser
 85 90 95
 His Asn Lys Ser Leu Leu Gln Phe Ala Lys Ile Asp Phe Asn Leu Leu
 100 105 110
 Gln Leu Leu His Arg Lys Glu Leu Ser Glu Ile Cys Arg
 115 120 125

<210> SEQ ID NO 37

<211> LENGTH: 72

<212> TYPE: PRT

<213> ORGANISM: *Gossypium arboreum*

<400> SEQUENCE: 37

Trp Trp Lys Asp Leu Asp Phe Thr Arg Lys Leu Pro Phe Ala Arg Asp
 1 5 10 15
 Arg Val Val Glu Gly Tyr Phe Trp Ile Met Gly Val Tyr Phe Glu Pro
 20 25 30
 Gln Tyr Ser Leu Gly Arg Lys Met Leu Thr Lys Val Ile Ala Met Ala
 35 40 45
 Ser Ile Val Asp Asp Thr Tyr Asp Ser Tyr Ala Thr Tyr Asp Glu Leu
 50 55 60
 Ile Pro Tyr Thr Asn Ala Ile Glu
 65 70

<210> SEQ ID NO 38

<211> LENGTH: 47

<212> TYPE: PRT

<213> ORGANISM: *Gossypium arboreum*

<400> SEQUENCE: 38

Arg Trp Asp Ile Lys Cys Met Asn Gln Leu Pro Asn Tyr Met Lys Ile
 1 5 10 15
 Ser Tyr Lys Ala Leu Leu Asp Val Tyr Glu Glu Met Glu Gln Leu Leu
 20 25 30
 Ala Asn Gln Gly Arg Gln Tyr Arg Val Glu Tyr Ala Lys Lys Ala
 35 40 45

<210> SEQ ID NO 39

<211> LENGTH: 83

<212> TYPE: PRT

<213> ORGANISM: *Gossypium arboreum*

<400> SEQUENCE: 39

Met Ile Arg Leu Val Gln Ala Tyr Leu Leu Glu Ala Lys Trp Thr His
 1 5 10 15

-continued

Gln Asn Tyr Lys Pro Thr Phe Glu Glu Phe Arg Asp Asn Ala Leu Pro
 20 25 30
 Thr Ser Gly Tyr Ala Met Leu Ala Ile Thr Ala Phe Val Gly Met Gly
 35 40 45
 Glu Val Ile Thr Pro Glu Thr Phe Lys Trp Ala Ala Ser Asp Pro Lys
 50 55 60
 Ile Ile Lys Ala Ser Thr Ile Ile Cys Arg Phe Met Asp Asp Ile Ala
 65 70 75 80
 Glu His Lys

<210> SEQ ID NO 40
 <211> LENGTH: 97
 <212> TYPE: PRT
 <213> ORGANISM: Gossypium arboreum

<400> SEQUENCE: 40

Phe Asn His Arg Arg Glu Asp Asp Cys Ser Ala Ile Glu Cys Tyr Met
 1 5 10 15
 Lys Gln Tyr Gly Val Thr Ala Gln Glu Ala Tyr Asn Glu Phe Asn Lys
 20 25 30
 His Ile Glu Ser Ser Trp Lys Asp Val Asn Glu Glu Phe Leu Lys Pro
 35 40 45
 Thr Glu Met Pro Thr Pro Val Leu Cys Arg Ser Leu Asn Leu Ala Arg
 50 55 60
 Val Met Asp Val Leu Tyr Arg Glu Gly Asp Gly Tyr Thr His Val Gly
 65 70 75 80
 Lys Ala Ala Lys Gly Gly Ile Thr Ser Leu Leu Ile Asp Pro Ile Gln
 85 90 95
 Ile

<210> SEQ ID NO 41
 <211> LENGTH: 1994
 <212> TYPE: DNA
 <213> ORGANISM: Ricinus communis
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (67)...(1869)
 <223> OTHER INFORMATION: casbene synthase

<400> SEQUENCE: 41

actcagcagc gcgctctcct accccaatta gcacagaaga tttggtgggt cctctccttg 60
 tgaaac atg gca ttg cca tca gct gct atg caa tcc aac cct gaa aag 108
 Met Ala Leu Pro Ser Ala Ala Met Gln Ser Asn Pro Glu Lys
 1 5 10
 ctt aac tta ttt cac aga ttg tca agc tta ccc acc act agc ttg gaa 156
 Leu Asn Leu Phe His Arg Leu Ser Ser Leu Pro Thr Thr Ser Leu Glu
 15 20 25 30
 tat ggc aat aat cgc ttc cct ttc ttt tcc tca tct gcc aag tca cac 204
 Tyr Gly Asn Asn Arg Phe Pro Phe Phe Ser Ser Ser Ala Lys Ser His
 35 40 45
 ttt aaa aaa cca act caa gca tgt tta tcc tca aca acc cac caa gaa 252
 Phe Lys Lys Pro Thr Gln Ala Cys Leu Ser Ser Thr Thr His Gln Glu
 50 55 60
 gtt cgt cca tta gca tac ttt cct cct act gtc tgg ggc aat cgc ttt 300
 Val Arg Pro Leu Ala Tyr Phe Pro Pro Thr Val Trp Gly Asn Arg Phe
 65 70 75
 gct tcc ttg acc ttc aat cca tcg gaa ttt gaa tcg tat gat gaa cgg 348
 Ala Ser Leu Thr Phe Asn Pro Ser Glu Phe Glu Ser Tyr Asp Glu Arg
 80 85 90

-continued

gta att gtg ctg aag aaa aaa gtt aag gac ata tta att tca tct aca Val Ile Val Leu Lys Lys Lys Val Lys Asp Ile Leu Ile Ser Ser Thr 95 100 105 110	396
agt gat tca gtg gag acc gtt att tta atc gac tta tta tgt cgg ctt Ser Asp Ser Val Glu Thr Val Ile Leu Ile Asp Leu Leu Cys Arg Leu 115 120 125	444
ggc gta tca tat cac ttt gaa aat gat att gaa gag cta cta agt aaa Gly Val Ser Tyr His Phe Glu Asn Asp Ile Glu Glu Leu Leu Ser Lys 130 135 140	492
atc ttc aac tcc cag cct gac ctt gtc gat gaa aaa gaa tgt gat ctc Ile Phe Asn Ser Gln Pro Asp Leu Val Asp Glu Lys Glu Cys Asp Leu 145 150 155	540
tac act gcg gca att gta ttc cga gtt ttc aga cag cat ggt ttt aaa Tyr Thr Ala Ala Ile Val Phe Arg Val Phe Arg Gln His Gly Phe Lys 160 165 170	588
atg tct tcg gat gtg ttt agc aaa ttc aag gac agt gat ggt aag ttc Met Ser Ser Asp Val Phe Ser Lys Phe Lys Asp Ser Asp Gly Lys Phe 175 180 185 190	636
aag gaa tcc cta cgg ggt gat gct aag ggt atg ctc agc ctt ttt gaa Lys Glu Ser Leu Arg Gly Asp Ala Lys Gly Met Leu Ser Leu Phe Glu 195 200 205	684
gct tcc cat cta agt gtg cat gga gaa gac att ctt gaa gaa gcc ttt Ala Ser His Leu Ser Val His Gly Glu Asp Ile Leu Glu Glu Ala Phe 210 215 220	732
gct ttc acc aag gat tac tta cag tcc tct gca gtt gag tta ttc cct Ala Phe Thr Lys Asp Tyr Leu Gln Ser Ser Ala Val Glu Leu Phe Pro 225 230 235	780
aat ctc aaa agg cat ata acg aac gcc cta gag cag cct ttc cac agt Asn Leu Lys Arg His Ile Thr Asn Ala Leu Glu Gln Pro Phe His Ser 240 245 250	828
ggc gtg cgg agg cta gag gcc agg aaa ttc atc gat cta tac gaa gct Gly Val Pro Arg Leu Glu Ala Arg Lys Phe Ile Asp Leu Tyr Glu Ala 255 260 265 270	876
gat att gaa tgc cgg aat gaa act ctg ctc gag ttt gca aag ttg gat Asp Ile Glu Cys Arg Asn Glu Thr Leu Leu Glu Phe Ala Lys Leu Asp 275 280 285	924
tat aat aga gtt cag tta ttg cac caa caa gag ctg tgc cag ttc tca Tyr Asn Arg Val Gln Leu Leu His Gln Gln Glu Leu Cys Gln Phe Ser 290 295 300	972
aag tgg tgg aaa gac ctg aat ctt gct tcg gat att cct tat gca aga Lys Trp Trp Lys Asp Leu Asn Leu Ala Ser Asp Ile Pro Tyr Ala Arg 305 310 315	1020
gac aga atg gca gag att ttc ttt tgg gca gtc gcg atg tac ttt gag Asp Arg Met Ala Glu Ile Phe Phe Trp Ala Val Ala Met Tyr Phe Glu 320 325 330	1068
cct gac tat gca cac acc cga atg att att gcg aag gtt gta ttg ctt Pro Asp Tyr Ala His Thr Arg Met Ile Ile Ala Lys Val Val Leu Leu 335 340 345 350	1116
ata tca cta ata gat gat aca att gat gcg tat gca aca atg gag gaa Ile Ser Leu Ile Asp Asp Thr Ile Asp Ala Tyr Ala Thr Met Glu Glu 355 360 365	1164
act cat att ctt gct gaa gca gtc gca agg tgg gac atg agc tgc ctc Thr His Ile Leu Ala Glu Ala Val Arg Trp Asp Met Ser Cys Leu 370 375 380	1212
gag aag ctg cca gat tac atg aaa gtt att tat aaa cta ttg cta aac Glu Lys Leu Pro Asp Tyr Met Lys Val Ile Tyr Lys Leu Leu Leu Asn 385 390 395	1260
acc ttc tct gaa ttc gag aaa gaa ttg acg gcg gaa ggc aag tcc tac Thr Phe Ser Glu Phe Glu Lys Glu Leu Thr Ala Glu Gly Lys Ser Tyr	1308

-continued

400	405	410	
agc gtc aaa tac gga	agg gaa gcg ttt	caa gaa cta gtg aga ggt tac	1356
Ser Val Lys Tyr Gly	Arg Glu Ala Phe	Gln Glu Leu Val Arg Gly Tyr	
415	420	425 430	
tac ctg gag gct gta	tgg cgc gac gag	ggt aaa ata cca tcg ttc gat	1404
Tyr Leu Glu Ala Val	Trp Arg Asp Glu	Gly Lys Ile Pro Ser Phe Asp	
435	440	445	
gac tac ttg tat aat	gga tcc atg acc	acc gga ttg cct ctc gtc tca	1452
Asp Tyr Leu Tyr Asn	Gly Ser Met Thr	Thr Gly Leu Pro Leu Val Ser	
450	455	460	
aca gct tct ttc atg	gga gtt caa gaa	att aca ggt ctc aac gaa ttc	1500
Thr Ala Ser Phe Met	Gly Val Gln Glu	Ile Thr Gly Leu Asn Glu Phe	
465	470	475	
caa tgg ctg gaa act	aat ccc aaa tta	agt tat gct tcc ggt gca ttc	1548
Gln Trp Leu Glu Thr	Asn Pro Lys Leu	Ser Tyr Ala Ser Gly Ala Phe	
480	485	490	
atc cga ctt gtc aac	gac tta act tct	cat gtg act gaa caa caa aga	1596
Ile Arg Leu Val Asn	Asp Leu Thr Ser	His Val Thr Glu Gln Gln Arg	
495	500	505 510	
gga cac gtt gca tct	tgc atc gac tgc	tat atg aac caa cat gga gtt	1644
Gly His Val Ala Ser	Cys Ile Asp Cys	Tyr Met Asn Gln His Gly Val	
515	520	525	
tcc aaa gac gaa gca	gtc aaa ata ctt	caa aaa atg gct aca gat tgt	1692
Ser Lys Asp Glu Ala	Val Lys Ile Leu	Gln Lys Met Ala Thr Asp Cys	
530	535	540	
tgg aaa gaa att aat	gaa gaa tgt atg	agg cag agt caa gtg tca gtg	1740
Trp Lys Glu Ile Asn	Glu Glu Cys Met	Arg Gln Ser Gln Val Ser Val	
545	550	555	
ggt cac cta atg aga	ata gtt aat ctg	gca cgt ctt acg gat gtg agt	1788
Gly His Leu Met Arg	Ile Val Asn Leu	Ala Arg Leu Thr Asp Val Ser	
560	565	570	
tac aag tat gga gac	ggt tac act gat	tcc cag caa ttg aaa caa ttt	1836
Tyr Lys Tyr Gly Asp	Gly Tyr Thr Asp	Ser Gln Gln Leu Lys Gln Phe	
575	580	585 590	
gtt aag gga ttg ttc	ggt gat cca att	tct att tgaactcaat aattcctttt	1889
Val Lys Gly Leu Phe	Val Asp Pro Ile	Ser Ile	
595	600		
ttcattttgt acttcaataa	gttataaatg acccgtgcac	tagcgggtgt gattattgta	1949
tttaaatgac cttttaaatt	aatatatgaa tcaagaattt	tatag	1994

<210> SEQ ID NO 42
 <211> LENGTH: 601
 <212> TYPE: PRT
 <213> ORGANISM: Ricinus communis

<400> SEQUENCE: 42

Met Ala Leu Pro Ser Ala Ala Met Gln Ser Asn Pro Glu Lys Leu Asn	1	5	10	15
Leu Phe His Arg Leu Ser Ser Leu Pro Thr Thr Ser Leu Glu Tyr Gly	20	25	30	
Asn Asn Arg Phe Pro Phe Phe Ser Ser Ser Ala Lys Ser His Phe Lys	35	40	45	
Lys Pro Thr Gln Ala Cys Leu Ser Ser Thr Thr His Gln Glu Val Arg	50	55	60	
Pro Leu Ala Tyr Phe Pro Pro Thr Val Trp Gly Asn Arg Phe Ala Ser	65	70	75	80
Leu Thr Phe Asn Pro Ser Glu Phe Glu Ser Tyr Asp Glu Arg Val Ile	85	90	95	

-continued

Val Leu Lys Lys Lys Val Lys Asp Ile Leu Ile Ser Ser Thr Ser Asp
 100 105 110
 Ser Val Glu Thr Val Ile Leu Ile Asp Leu Leu Cys Arg Leu Gly Val
 115 120 125
 Ser Tyr His Phe Glu Asn Asp Ile Glu Glu Leu Leu Ser Lys Ile Phe
 130 135 140
 Asn Ser Gln Pro Asp Leu Val Asp Glu Lys Glu Cys Asp Leu Tyr Thr
 145 150 155 160
 Ala Ala Ile Val Phe Arg Val Phe Arg Gln His Gly Phe Lys Met Ser
 165 170 175
 Ser Asp Val Phe Ser Lys Phe Lys Asp Ser Asp Gly Lys Phe Lys Glu
 180 185 190
 Ser Leu Arg Gly Asp Ala Lys Gly Met Leu Ser Leu Phe Glu Ala Ser
 195 200 205
 His Leu Ser Val His Gly Glu Asp Ile Leu Glu Glu Ala Phe Ala Phe
 210 215 220
 Thr Lys Asp Tyr Leu Gln Ser Ser Ala Val Glu Leu Phe Pro Asn Leu
 225 230 235 240
 Lys Arg His Ile Thr Asn Ala Leu Glu Gln Pro Phe His Ser Gly Val
 245 250 255
 Pro Arg Leu Glu Ala Arg Lys Phe Ile Asp Leu Tyr Glu Ala Asp Ile
 260 265 270
 Glu Cys Arg Asn Glu Thr Leu Leu Glu Phe Ala Lys Leu Asp Tyr Asn
 275 280 285
 Arg Val Gln Leu Leu His Gln Gln Glu Leu Cys Gln Phe Ser Lys Trp
 290 295 300
 Trp Lys Asp Leu Asn Leu Ala Ser Asp Ile Pro Tyr Ala Arg Asp Arg
 305 310 315 320
 Met Ala Glu Ile Phe Phe Trp Ala Val Ala Met Tyr Phe Glu Pro Asp
 325 330 335
 Tyr Ala His Thr Arg Met Ile Ile Ala Lys Val Val Leu Leu Ile Ser
 340 345 350
 Leu Ile Asp Asp Thr Ile Asp Ala Tyr Ala Thr Met Glu Glu Thr His
 355 360 365
 Ile Leu Ala Glu Ala Val Ala Arg Trp Asp Met Ser Cys Leu Glu Lys
 370 375 380
 Leu Pro Asp Tyr Met Lys Val Ile Tyr Lys Leu Leu Leu Asn Thr Phe
 385 390 395 400
 Ser Glu Phe Glu Lys Glu Leu Thr Ala Glu Gly Lys Ser Tyr Ser Val
 405 410 415
 Lys Tyr Gly Arg Glu Ala Phe Gln Glu Leu Val Arg Gly Tyr Tyr Leu
 420 425 430
 Glu Ala Val Trp Arg Asp Glu Gly Lys Ile Pro Ser Phe Asp Asp Tyr
 435 440 445
 Leu Tyr Asn Gly Ser Met Thr Thr Gly Leu Pro Leu Val Ser Thr Ala
 450 455 460
 Ser Phe Met Gly Val Gln Glu Ile Thr Gly Leu Asn Glu Phe Gln Trp
 465 470 475 480
 Leu Glu Thr Asn Pro Lys Leu Ser Tyr Ala Ser Gly Ala Phe Ile Arg
 485 490 495
 Leu Val Asn Asp Leu Thr Ser His Val Thr Glu Gln Gln Arg Gly His
 500 505 510

-continued

Val Ala Ser Cys Ile Asp Cys Tyr Met Asn Gln His Gly Val Ser Lys
 515 520 525

Asp Glu Ala Val Lys Ile Leu Gln Lys Met Ala Thr Asp Cys Trp Lys
 530 535 540

Glu Ile Asn Glu Glu Cys Met Arg Gln Ser Gln Val Ser Val Gly His
 545 550 555 560

Leu Met Arg Ile Val Asn Leu Ala Arg Leu Thr Asp Val Ser Tyr Lys
 565 570 575

Tyr Gly Asp Gly Tyr Thr Asp Ser Gln Gln Leu Lys Gln Phe Val Lys
 580 585 590

Gly Leu Phe Val Asp Pro Ile Ser Ile
 595 600

<210> SEQ ID NO 43
 <211> LENGTH: 2700
 <212> TYPE: DNA
 <213> ORGANISM: Taxus brevifolia
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (22)...(2607)
 <223> OTHER INFORMATION: taxadiene synthase

<400> SEQUENCE: 43

ttccctgctgctctctggaga a atg gct cag ctc tca ttt aat gca gcg ctg 51
 Met Ala Gln Leu Ser Phe Asn Ala Ala Leu 10
 1 5

aag atg aac gca ttg ggg aac aag gca atc cac gat cca acg aat tgc 99
 Lys Met Asn Ala Leu Gly Asn Lys Ala Ile His Asp Pro Thr Asn Cys 25
 15 20

aga gcc aaa tct gag cgc caa atg atg tgg gtt tgc tcc aga tca ggg 147
 Arg Ala Lys Ser Glu Arg Gln Met Met Trp Val Cys Ser Arg Ser Gly 40
 30 35

cga acc aga gta aaa atg tcg aga gga agt ggt ggt cct ggt cct gtc 195
 Arg Thr Arg Val Lys Met Ser Arg Gly Ser Gly Gly Pro Gly Pro Val 55
 45 50

gta atg atg agc agc agc act ggc act agc aag gtg gtt tcc gag act 243
 Val Met Met Ser Ser Ser Thr Gly Thr Ser Lys Val Val Ser Glu Thr 70
 60 65

tcc agt acc att gtg gat gat atc cct cga ctc tcc gcc aat tat cat 291
 Ser Ser Thr Ile Val Asp Asp Ile Pro Arg Leu Ser Ala Asn Tyr His 90
 75 80 85

ggc gat ctg tgg cac cac aat gtt ata caa act ctg gag aca ccg ttt 339
 Gly Asp Leu Trp His His Asn Val Ile Gln Thr Leu Glu Thr Pro Phe 105
 95 100

cgt gag agt tct act tac caa gaa cgg gca gat gag ctg gtt gtg aaa 387
 Arg Glu Ser Ser Thr Tyr Gln Glu Arg Ala Asp Glu Leu Val Val Lys 120
 110 115

att aaa gat atg ttc aat gcg ctc gga gac gga gat atc agt ccg tct 435
 Ile Lys Asp Met Phe Asn Ala Leu Gly Asp Gly Asp Ile Ser Pro Ser 135
 125 130

gca tac gac act gcg tgg gtg gcg agg ctg gcg acc att tcc tct gat 483
 Ala Tyr Asp Thr Ala Trp Val Ala Arg Leu Ala Thr Ile Ser Ser Asp 150
 140 145 150

gga tct gag aag cca cgg ttt cct cag gcc ctc aac tgg gtt ttc aac 531
 Gly Ser Glu Lys Pro Arg Phe Pro Gln Ala Leu Asn Trp Val Phe Asn 170
 155 160 165

aac cag ctc cag gat gga tcg tgg ggt atc gaa tcg cac ttt agt tta 579
 Asn Gln Leu Gln Asp Gly Ser Trp Gly Ile Glu Ser His Phe Ser Leu 185
 175 180

tgc gat cga ttg ctt aac acg acc aat tct gtt atc gcc ctc tcg gtt 627

-continued

Cys Asp Arg Leu Leu Asn Thr Thr Asn Ser Val Ile Ala Leu Ser Val	
190 195 200	
tgg aaa aca ggg cac agc caa gta caa caa ggt gct gag ttt att gca	675
Trp Lys Thr Gly His Ser Gln Val Gln Gln Gly Ala Glu Phe Ile Ala	
205 210 215	
gag aat cta aga tta ctc aat gag gaa gat gag ttg tcc ccg gat ttc	723
Glu Asn Leu Arg Leu Leu Asn Glu Glu Asp Glu Leu Ser Pro Asp Phe	
220 225 230	
caa ata atc ttt cct gct ctg ctg caa aag gca aaa gcg ttg ggg atc	771
Gln Ile Ile Phe Pro Ala Leu Leu Gln Lys Ala Lys Ala Leu Gly Ile	
235 240 245 250	
aat ctt cct tac gat ctt cca ttt atc aaa tat ttg tcg aca aca cgg	819
Asn Leu Pro Tyr Asp Leu Pro Phe Ile Lys Tyr Leu Ser Thr Thr Arg	
255 260 265	
gaa gcc agg ctt aca gat gtt tct gcg gca gca gac aat att cca gcc	867
Glu Ala Arg Leu Thr Asp Val Ser Ala Ala Ala Asp Asn Ile Pro Ala	
270 275 280	
aac atg ttg aat gcg ttg gaa ggt ctc gag gaa gtt att gac tgg aac	915
Asn Met Leu Asn Ala Leu Glu Gly Leu Glu Glu Val Ile Asp Trp Asn	
285 290 295	
aag att atg agg ttt caa agt aaa gat gga tct ttc ctg agc tcc cct	963
Lys Ile Met Arg Phe Gln Ser Lys Asp Gly Ser Phe Leu Ser Ser Pro	
300 305 310	
gcc tcc act gcc tgt gta ctg atg aat aca ggg gac gaa aaa tgt ttc	1011
Ala Ser Thr Ala Cys Val Leu Met Asn Thr Gly Asp Glu Lys Cys Phe	
315 320 325 330	
act ttt ctc aac aat ctg ctc gac aaa ttc ggc ggc tgc gtg ccc tgt	1059
Thr Phe Leu Asn Asn Leu Leu Asp Lys Phe Gly Gly Cys Val Pro Cys	
335 340 345	
atg tat tcc atc gat ctg ctg gaa cgc ctt tcg ctg gtt gat aac att	1107
Met Tyr Ser Ile Asp Leu Leu Glu Arg Leu Ser Leu Val Asp Asn Ile	
350 355 360	
gag cat ctc gga atc ggt cgc cat ttc aaa caa gaa atc aaa gga gct	1155
Glu His Leu Gly Ile Gly Arg His Phe Lys Gln Glu Ile Lys Gly Ala	
365 370 375	
ctt gat tat gtc tac aga cat tgg agt gaa agg ggc atc ggt tgg ggc	1203
Leu Asp Tyr Val Tyr Arg His Trp Ser Glu Arg Gly Ile Gly Trp Gly	
380 385 390	
aga gac agc ctt gtt cca gat ctc aac acc aca gcc ctc ggc ctg cga	1251
Arg Asp Ser Leu Val Pro Asp Leu Asn Thr Thr Ala Leu Gly Leu Arg	
395 400 405 410	
act ctt cgc atg cac gga tac aat gtt tct tca gac gtt ttg aat aat	1299
Thr Leu Arg Met His Gly Tyr Asn Val Ser Ser Asp Val Leu Asn Asn	
415 420 425	
ttc aaa gat gaa aac ggg cgg ttc ttc tcc tct gcg ggc caa acc cat	1347
Phe Lys Asp Glu Asn Gly Arg Phe Phe Ser Ser Ala Gly Gln Thr His	
430 435 440	
gtc gaa ttg aga agc gtg gtg aat ctt ttc aga gct tcc gac ctt gca	1395
Val Glu Leu Arg Ser Val Val Asn Leu Phe Arg Ala Ser Asp Leu Ala	
445 450 455	
ttt cct gac gaa aga gct atg gac gat gct aga aaa ttt gca gaa cca	1443
Phe Pro Asp Glu Arg Ala Met Asp Asp Ala Arg Lys Phe Ala Glu Pro	
460 465 470	
tat ctt aga gag gca ctt gca acg aaa atc tca acc aat aca aaa cta	1491
Tyr Leu Arg Glu Ala Leu Ala Thr Lys Ile Ser Thr Asn Thr Lys Leu	
475 480 485 490	
ttc aaa gag att gag tac gtg gtg gag tac cct tgg cac atg agt atc	1539
Phe Lys Glu Ile Glu Tyr Val Val Glu Tyr Pro Trp His Met Ser Ile	
495 500 505	

-continued

cca cgc tta gaa gcc aga agt tat att gat tca tat gac gac aat tat	1587
Pro Arg Leu Glu Ala Arg Ser Tyr Ile Asp Ser Tyr Asp Asp Asn Tyr	
510 515 520	
gta tgg cag agg aag act cta tat aga atg cca tct ttg agt aat tca	1635
Val Trp Gln Arg Lys Thr Leu Tyr Arg Met Pro Ser Leu Ser Asn Ser	
525 530 535	
aaa tgt tta gaa ttg gca aaa ttg gac ttc aat atc gta caa tct ttg	1683
Lys Cys Leu Glu Leu Ala Lys Leu Asp Phe Asn Ile Val Gln Ser Leu	
540 545 550	
cat caa gag gag ttg aag ctt cta aca aga tgg tgg aag gaa tcc ggc	1731
His Gln Glu Glu Leu Lys Leu Leu Thr Arg Trp Trp Lys Glu Ser Gly	
555 560 565 570	
atg gca gat ata aat ttc act cga cac cga gtg gcg gag gtt tat ttt	1779
Met Ala Asp Ile Asn Phe Thr Arg His Arg Val Ala Glu Val Tyr Phe	
575 580 585	
tca tca gct aca ttt gaa ccc gaa tat tct gcc act aga att gcc ttc	1827
Ser Ser Ala Thr Phe Glu Pro Glu Tyr Ser Ala Thr Arg Ile Ala Phe	
590 595 600	
aca aaa att ggt tgt tta caa gtc ctt ttt gat gat atg gct gac atc	1875
Thr Lys Ile Gly Cys Leu Gln Val Leu Phe Asp Asp Met Ala Asp Ile	
605 610 615	
ttt gca aca cta gat gaa ttg aaa agt ttc act gag gga gta aag aga	1923
Phe Ala Thr Leu Asp Glu Leu Lys Ser Phe Thr Glu Gly Val Lys Arg	
620 625 630	
tgg gat aca tct ttg cta cat gag att cca gag tgt atg caa act tgc	1971
Trp Asp Thr Ser Leu Leu His Glu Ile Pro Glu Cys Met Gln Thr Cys	
635 640 645 650	
ttt aaa gtt tgg ttc aaa tta atg gaa gaa gta aat aat gat gtg gtt	2019
Phe Lys Val Trp Phe Lys Leu Met Glu Glu Val Asn Asn Asp Val Val	
655 660 665	
aag gta caa gga cgt gac atg ctc gct cac ata aga aaa ccc tgg gag	2067
Lys Val Gln Gly Arg Asp Met Leu Ala His Ile Arg Lys Pro Trp Glu	
670 675 680	
ttg tac ttc aat tgt tat gta caa gaa agg gag tgg ctt gaa gcc ggg	2115
Leu Tyr Phe Asn Cys Tyr Val Gln Glu Arg Glu Trp Leu Glu Ala Gly	
685 690 695	
tat ata cca act ttt gaa gag tac tta aag act tat gct ata tca gta	2163
Tyr Ile Pro Thr Phe Glu Glu Tyr Leu Lys Thr Tyr Ala Ile Ser Val	
700 705 710	
ggc ctt gga ccg tgt acc cta caa cca ata cta cta atg ggt gag ctt	2211
Gly Leu Gly Pro Cys Thr Leu Gln Pro Ile Leu Leu Met Gly Glu Leu	
715 720 725 730	
gtg aaa gat gat gtt gtt gag aaa gtg cac tat ccc tca aat atg ttt	2259
Val Lys Asp Asp Val Val Glu Lys Val His Tyr Pro Ser Asn Met Phe	
735 740 745	
gag ctt gta tcc ttg agc tgg cga cta aca aac gac acc aaa aca tat	2307
Glu Leu Val Ser Leu Ser Trp Arg Leu Thr Asn Asp Thr Lys Thr Tyr	
750 755 760	
cag gct gaa aag gct cga gga caa caa gcc tca ggc ata gca tgc tat	2355
Gln Ala Glu Lys Ala Arg Gly Gln Gln Ala Ser Gly Ile Ala Cys Tyr	
765 770 775	
atg aag gat aat cca gga gca act gag gaa gat gcc att aag cac ata	2403
Met Lys Asp Asn Pro Gly Ala Thr Glu Glu Asp Ala Ile Lys His Ile	
780 785 790	
tgt cgt gtt gtt gat cgg gcc ttg aaa gaa gca agc ttt gaa tat ttc	2451
Cys Arg Val Val Asp Arg Ala Leu Lys Glu Ala Ser Phe Glu Tyr Phe	
795 800 805 810	
aaa cca tcc aat gat atc cca atg ggt tgc aag tcc ttt att ttt aac	2499
Lys Pro Ser Asn Asp Ile Pro Met Gly Cys Lys Ser Phe Ile Phe Asn	
815 820 825	

-continued

```

ctt aga ttg tgt gtc caa atc ttt tac aag ttt ata gat ggg tac gga      2547
Leu Arg Leu Cys Val Gln Ile Phe Tyr Lys Phe Ile Asp Gly Tyr Gly
      830                      835                      840

atc gcc aat gag gag att aag gac tat ata aga aaa gtt tat att gat      2595
Ile Ala Asn Glu Glu Ile Lys Asp Tyr Ile Arg Lys Val Tyr Ile Asp
      845                      850                      855

cca att caa gta tga tatatcatgt aaaacctctt tticcatgata aattgactta      2650
Pro Ile Gln Val
      860

ttattgtatt ggcaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa      2700

```

```

<210> SEQ ID NO 44
<211> LENGTH: 862
<212> TYPE: PRT
<213> ORGANISM: Taxus brevifolia

```

```

<400> SEQUENCE: 44

```

```

Met Ala Gln Leu Ser Phe Asn Ala Ala Leu Lys Met Asn Ala Leu Gly
  1                      5                      10                      15

Asn Lys Ala Ile His Asp Pro Thr Asn Cys Arg Ala Lys Ser Glu Arg
      20                      25                      30

Gln Met Met Trp Val Cys Ser Arg Ser Gly Arg Thr Arg Val Lys Met
      35                      40                      45

Ser Arg Gly Ser Gly Gly Pro Gly Pro Val Val Met Met Ser Ser Ser
      50                      55                      60

Thr Gly Thr Ser Lys Val Val Ser Glu Thr Ser Ser Thr Ile Val Asp
      65                      70                      75                      80

Asp Ile Pro Arg Leu Ser Ala Asn Tyr His Gly Asp Leu Trp His His
      85                      90                      95

Asn Val Ile Gln Thr Leu Glu Thr Pro Phe Arg Glu Ser Ser Thr Tyr
      100                      105                      110

Gln Glu Arg Ala Asp Glu Leu Val Val Lys Ile Lys Asp Met Phe Asn
      115                      120                      125

Ala Leu Gly Asp Gly Asp Ile Ser Pro Ser Ala Tyr Asp Thr Ala Trp
      130                      135                      140

Val Ala Arg Leu Ala Thr Ile Ser Ser Asp Gly Ser Glu Lys Pro Arg
      145                      150                      155                      160

Phe Pro Gln Ala Leu Asn Trp Val Phe Asn Asn Gln Leu Gln Asp Gly
      165                      170                      175

Ser Trp Gly Ile Glu Ser His Phe Ser Leu Cys Asp Arg Leu Leu Asn
      180                      185                      190

Thr Thr Asn Ser Val Ile Ala Leu Ser Val Trp Lys Thr Gly His Ser
      195                      200                      205

Gln Val Gln Gln Gly Ala Glu Phe Ile Ala Glu Asn Leu Arg Leu Leu
      210                      215                      220

Asn Glu Glu Asp Glu Leu Ser Pro Asp Phe Gln Ile Ile Phe Pro Ala
      225                      230                      235                      240

Leu Leu Gln Lys Ala Lys Ala Leu Gly Ile Asn Leu Pro Tyr Asp Leu
      245                      250                      255

Pro Phe Ile Lys Tyr Leu Ser Thr Thr Arg Glu Ala Arg Leu Thr Asp
      260                      265                      270

Val Ser Ala Ala Ala Asp Asn Ile Pro Ala Asn Met Leu Asn Ala Leu
      275                      280                      285

Glu Gly Leu Glu Glu Val Ile Asp Trp Asn Lys Ile Met Arg Phe Gln
      290                      295                      300

```

-continued

Ser Lys Asp Gly Ser Phe Leu Ser Ser Pro Ala Ser Thr Ala Cys Val
 305 310 315 320
 Leu Met Asn Thr Gly Asp Glu Lys Cys Phe Thr Phe Leu Asn Asn Leu
 325 330 335
 Leu Asp Lys Phe Gly Gly Cys Val Pro Cys Met Tyr Ser Ile Asp Leu
 340 345 350
 Leu Glu Arg Leu Ser Leu Val Asp Asn Ile Glu His Leu Gly Ile Gly
 355 360 365
 Arg His Phe Lys Gln Glu Ile Lys Gly Ala Leu Asp Tyr Val Tyr Arg
 370 375 380
 His Trp Ser Glu Arg Gly Ile Gly Trp Gly Arg Asp Ser Leu Val Pro
 385 390 395 400
 Asp Leu Asn Thr Thr Ala Leu Gly Leu Arg Thr Leu Arg Met His Gly
 405 410 415
 Tyr Asn Val Ser Ser Asp Val Leu Asn Asn Phe Lys Asp Glu Asn Gly
 420 425 430
 Arg Phe Phe Ser Ser Ala Gly Gln Thr His Val Glu Leu Arg Ser Val
 435 440 445
 Val Asn Leu Phe Arg Ala Ser Asp Leu Ala Phe Pro Asp Glu Arg Ala
 450 455 460
 Met Asp Asp Ala Arg Lys Phe Ala Glu Pro Tyr Leu Arg Glu Ala Leu
 465 470 475 480
 Ala Thr Lys Ile Ser Thr Asn Thr Lys Leu Phe Lys Glu Ile Glu Tyr
 485 490 495
 Val Val Glu Tyr Pro Trp His Met Ser Ile Pro Arg Leu Glu Ala Arg
 500 505 510
 Ser Tyr Ile Asp Ser Tyr Asp Asp Asn Tyr Val Trp Gln Arg Lys Thr
 515 520 525
 Leu Tyr Arg Met Pro Ser Leu Ser Asn Ser Lys Cys Leu Glu Leu Ala
 530 535 540
 Lys Leu Asp Phe Asn Ile Val Gln Ser Leu His Gln Glu Glu Leu Lys
 545 550 555 560
 Leu Leu Thr Arg Trp Trp Lys Glu Ser Gly Met Ala Asp Ile Asn Phe
 565 570 575
 Thr Arg His Arg Val Ala Glu Val Tyr Phe Ser Ser Ala Thr Phe Glu
 580 585 590
 Pro Glu Tyr Ser Ala Thr Arg Ile Ala Phe Thr Lys Ile Gly Cys Leu
 595 600 605
 Gln Val Leu Phe Asp Asp Met Ala Asp Ile Phe Ala Thr Leu Asp Glu
 610 615 620
 Leu Lys Ser Phe Thr Glu Gly Val Lys Arg Trp Asp Thr Ser Leu Leu
 625 630 635 640
 His Glu Ile Pro Glu Cys Met Gln Thr Cys Phe Lys Val Trp Phe Lys
 645 650 655
 Leu Met Glu Glu Val Asn Asn Asp Val Val Lys Val Gln Gly Arg Asp
 660 665 670
 Met Leu Ala His Ile Arg Lys Pro Trp Glu Leu Tyr Phe Asn Cys Tyr
 675 680 685
 Val Gln Glu Arg Glu Trp Leu Glu Ala Gly Tyr Ile Pro Thr Phe Glu
 690 695 700
 Glu Tyr Leu Lys Thr Tyr Ala Ile Ser Val Gly Leu Gly Pro Cys Thr
 705 710 715 720

-continued

Leu Gln Pro Ile Leu Leu Met Gly Glu Leu Val Lys Asp Asp Val Val
725 730 735

Glu Lys Val His Tyr Pro Ser Asn Met Phe Glu Leu Val Ser Leu Ser
740 745 750

Trp Arg Leu Thr Asn Asp Thr Lys Thr Tyr Gln Ala Glu Lys Ala Arg
755 760 765

Gly Gln Gln Ala Ser Gly Ile Ala Cys Tyr Met Lys Asp Asn Pro Gly
770 775 780

Ala Thr Glu Glu Asp Ala Ile Lys His Ile Cys Arg Val Val Asp Arg
785 790 795 800

Ala Leu Lys Glu Ala Ser Phe Glu Tyr Phe Lys Pro Ser Asn Asp Ile
805 810 815

Pro Met Gly Cys Lys Ser Phe Ile Phe Asn Leu Arg Leu Cys Val Gln
820 825 830

Ile Phe Tyr Lys Phe Ile Asp Gly Tyr Gly Ile Ala Asn Glu Glu Ile
835 840 845

Lys Asp Tyr Ile Arg Lys Val Tyr Ile Asp Pro Ile Gln Val
850 855 860

<210> SEQ ID NO 45
<211> LENGTH: 2424
<212> TYPE: DNA
<213> ORGANISM: Abies grandis
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (2)...(2347)
<223> OTHER INFORMATION: E-alpha-bisabolene synthase

<400> SEQUENCE: 45

g ggt tat gat ctt gtg cat tct ctt aaa tca cct tat att gat tct agt 49
Gly Tyr Asp Leu Val His Ser Leu Lys Ser Pro Tyr Ile Asp Ser Ser
1 5 10 15

tac aga gaa cgc gcg gag gtc ctt gtt agc gag att aaa gtg atg ctt 97
Tyr Arg Glu Arg Ala Glu Val Leu Val Ser Glu Ile Lys Val Met Leu
20 25 30

aat cca gct att aca gga gat gga gaa tca atg att act cca tct gct 145
Asn Pro Ala Ile Thr Gly Asp Gly Glu Ser Met Ile Thr Pro Ser Ala
35 40 45

tat gac aca gca tgg gta gcg agg gtg ccc gcc att gat ggc tct gct 193
Tyr Asp Thr Ala Trp Val Ala Arg Val Pro Ala Ile Asp Gly Ser Ala
50 55 60

cgc ccg caa ttt ccc caa aca gtt gac tgg att ttg aaa aac cag tta 241
Arg Pro Gln Phe Pro Gln Thr Val Asp Trp Ile Leu Lys Asn Gln Leu
65 70 75 80

aaa gat ggt tca tgg gga att cag tcc cac ttt ctg ctg tcc gac cgt 289
Lys Asp Gly Ser Trp Gly Ile Gln Ser His Phe Leu Leu Ser Asp Arg
85 90 95

ctt ctt gcc act ctt tct tgt gtt ctt gtg ctc ctt aaa tgg aac gtt 337
Leu Leu Ala Thr Leu Ser Cys Val Leu Val Leu Leu Lys Trp Asn Val
100 105 110

ggg gat ctg caa gta gag cag gga att gaa ttc ata aag agc aat ctg 385
Gly Asp Leu Gln Val Glu Gln Gly Ile Glu Phe Ile Lys Ser Asn Leu
115 120 125

gaa cta gta aag gat gaa acc gat caa gat agc ttg gta aca gac ttt 433
Glu Leu Val Lys Asp Glu Thr Asp Gln Asp Ser Leu Val Thr Asp Phe
130 135 140

gag atc ata ttt cct tct ctg tta aga gaa gct caa tct ctg cgc ctc 481
Glu Ile Ile Phe Pro Ser Leu Leu Arg Glu Ala Gln Ser Leu Arg Leu
145 150 155 160

-continued

tgg ttt aga gat tcg ggg ttg cca cta ttc acc ttc gct cgg gag agg	1489
Trp Phe Arg Asp Ser Gly Leu Pro Leu Phe Thr Phe Ala Arg Glu Arg	
485 490 495	
ccg ctg gaa ttc tac ttc tta gta gcg gcg ggg acc tat gaa ccc cag	1537
Pro Leu Glu Phe Tyr Phe Leu Val Ala Ala Gly Thr Tyr Glu Pro Gln	
500 505 510	
tat gcc aaa tgc agg ttc ctc ttt aca aaa gtg gca tgc ttg cag act	1585
Tyr Ala Lys Cys Arg Phe Leu Phe Thr Lys Val Ala Cys Leu Gln Thr	
515 520 525	
gtt ctg gac gat atg tat gac act tat gga acc cta gat gaa ttg aag	1633
Val Leu Asp Asp Met Tyr Asp Thr Tyr Gly Thr Leu Asp Glu Leu Lys	
530 535 540	
cta ttc act gag gct gtg aga aga tgg gac ctc tcc ttt aca gaa aac	1681
Leu Phe Thr Glu Ala Val Arg Arg Trp Asp Leu Ser Phe Thr Glu Asn	
545 550 555 560	
ctt cca gac tat atg aaa cta tgt tac caa atc tat tat gac ata gtt	1729
Leu Pro Asp Tyr Met Lys Leu Cys Tyr Gln Ile Tyr Tyr Asp Ile Val	
565 570 575	
cac gag gtg gct tgg gag gca gag aag gaa cag ggg cgt gaa ttg gtc	1777
His Glu Val Ala Trp Glu Ala Glu Lys Glu Gln Gly Arg Glu Leu Val	
580 585 590	
agc ttt ttc aga aag gga tgg gag gat tat ctt ctg ggt tat tat gaa	1825
Ser Phe Phe Arg Lys Gly Trp Glu Asp Tyr Leu Leu Gly Tyr Tyr Glu	
595 600 605	
gaa gct gaa tgg tta gct gct gag tat gtg cct acc ttg gac gag tac	1873
Glu Ala Glu Trp Leu Ala Ala Glu Tyr Val Pro Thr Leu Asp Glu Tyr	
610 615 620	
ata aag aat gga atc aca tct atc ggc caa cgt ata ctt ctg ttg agt	1921
Ile Lys Asn Gly Ile Thr Ser Ile Gly Gln Arg Ile Leu Leu Leu Ser	
625 630 635 640	
gga gtg ttg ata atg gat ggg caa ctc ctt tcg caa gag gca tta gag	1969
Gly Val Leu Ile Met Asp Gly Gln Leu Leu Ser Gln Glu Ala Leu Glu	
645 650 655	
aaa gta gat tat cca gga aga cgt gtt ctc aca gag ctg aat agc ctc	2017
Lys Val Asp Tyr Pro Gly Arg Arg Val Leu Thr Glu Leu Asn Ser Leu	
660 665 670	
att tcc cgc ctg gcg gat gac acg aag aca tat aaa gct gag aag gct	2065
Ile Ser Arg Leu Ala Asp Asp Thr Lys Thr Tyr Lys Ala Glu Lys Ala	
675 680 685	
cgt gga gaa ttg gcg tcc agc att gaa tgt tac atg aaa gac cat cct	2113
Arg Gly Glu Leu Ala Ser Ser Ile Glu Cys Tyr Met Lys Asp His Pro	
690 695 700	
gaa tgt aca gag gaa gag gct ctc gat cac atc tat agc att ctg gag	2161
Glu Cys Thr Glu Glu Glu Ala Leu Asp His Ile Tyr Ser Ile Leu Glu	
705 710 715 720	
ccg gcg gtg aag gaa ctg aca aga gag ttt ctg aag ccc gac gac gtc	2209
Pro Ala Val Lys Glu Leu Thr Arg Glu Phe Leu Lys Pro Asp Asp Val	
725 730 735	
cca ttc gcc tgc aag aag atg ctt ttc gag gag aca aga gtg acg atg	2257
Pro Phe Ala Cys Lys Lys Met Leu Phe Glu Glu Thr Arg Val Thr Met	
740 745 750	
gtg ata ttc aag gat gga gat gga ttc ggt gtt tcc aaa tta gaa gtc	2305
Val Ile Phe Lys Asp Gly Asp Gly Phe Gly Val Ser Lys Leu Glu Val	
755 760 765	
aaa gat cat atc aaa gag tgt ctc att gaa ccg ctg cca ctg taa	2350
Lys Asp His Ile Lys Glu Cys Leu Ile Glu Pro Leu Pro Leu	
770 775 780	
tcaaaatagt tgcaataata attgaaataa tgtcaactat gtttcacaaa aaaaaaaaaa	2410

-continued

aaaaaaaaaa aaaa

2424

<210> SEQ ID NO 46

<211> LENGTH: 782

<212> TYPE: PRT

<213> ORGANISM: Abies grandis

<400> SEQUENCE: 46

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

-continued

Ala Phe Pro Gly Glu Asn Ile Leu Asp Glu Ala Lys Ser Phe Ala Thr
 370 375 380

Lys Tyr Leu Arg Glu Ala Leu Glu Lys Ser Glu Thr Ser Ser Ala Trp
 385 390 395 400

Asn Asn Lys Gln Asn Leu Ser Gln Glu Ile Lys Tyr Ala Leu Lys Thr
 405 410 415

Ser Trp His Ala Ser Val Pro Arg Val Glu Ala Lys Arg Tyr Cys Gln
 420 425 430

Val Tyr Arg Pro Asp Tyr Ala Arg Ile Ala Lys Cys Val Tyr Lys Leu
 435 440 445

Pro Tyr Val Asn Asn Glu Lys Phe Leu Glu Leu Gly Lys Leu Asp Phe
 450 455 460

Asn Ile Ile Gln Ser Ile His Gln Glu Glu Met Lys Asn Val Thr Ser
 465 470 475 480

Trp Phe Arg Asp Ser Gly Leu Pro Leu Phe Thr Phe Ala Arg Glu Arg
 485 490 495

Pro Leu Glu Phe Tyr Phe Leu Val Ala Ala Gly Thr Tyr Glu Pro Gln
 500 505 510

Tyr Ala Lys Cys Arg Phe Leu Phe Thr Lys Val Ala Cys Leu Gln Thr
 515 520 525

Val Leu Asp Asp Met Tyr Asp Thr Tyr Gly Thr Leu Asp Glu Leu Lys
 530 535 540

Leu Phe Thr Glu Ala Val Arg Arg Trp Asp Leu Ser Phe Thr Glu Asn
 545 550 555 560

Leu Pro Asp Tyr Met Lys Leu Cys Tyr Gln Ile Tyr Tyr Asp Ile Val
 565 570 575

His Glu Val Ala Trp Glu Ala Glu Lys Glu Gln Gly Arg Glu Leu Val
 580 585 590

Ser Phe Phe Arg Lys Gly Trp Glu Asp Tyr Leu Leu Gly Tyr Tyr Glu
 595 600 605

Glu Ala Glu Trp Leu Ala Ala Glu Tyr Val Pro Thr Leu Asp Glu Tyr
 610 615 620

Ile Lys Asn Gly Ile Thr Ser Ile Gly Gln Arg Ile Leu Leu Leu Ser
 625 630 635 640

Gly Val Leu Ile Met Asp Gly Gln Leu Leu Ser Gln Glu Ala Leu Glu
 645 650 655

Lys Val Asp Tyr Pro Gly Arg Arg Val Leu Thr Glu Leu Asn Ser Leu
 660 665 670

Ile Ser Arg Leu Ala Asp Asp Thr Lys Thr Tyr Lys Ala Glu Lys Ala
 675 680 685

Arg Gly Glu Leu Ala Ser Ser Ile Glu Cys Tyr Met Lys Asp His Pro
 690 695 700

Glu Cys Thr Glu Glu Glu Ala Leu Asp His Ile Tyr Ser Ile Leu Glu
 705 710 715 720

Pro Ala Val Lys Glu Leu Thr Arg Glu Phe Leu Lys Pro Asp Asp Val
 725 730 735

Pro Phe Ala Cys Lys Lys Met Leu Phe Glu Glu Thr Arg Val Thr Met
 740 745 750

Val Ile Phe Lys Asp Gly Asp Gly Phe Gly Val Ser Lys Leu Glu Val
 755 760 765

Lys Asp His Ile Lys Glu Cys Leu Ile Glu Pro Leu Pro Leu
 770 775 780

-continued

<210> SEQ ID NO 47
 <211> LENGTH: 1865
 <212> TYPE: DNA
 <213> ORGANISM: *Abies grandis*
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (1)...(1743)
 <223> OTHER INFORMATION: d-selinene synthase

<400> SEQUENCE: 47

atg gct gag att tct gaa tct tcc atc cct cga cgc aca ggg aat cat	48
Met Ala Glu Ile Ser Glu Ser Ser Ile Pro Arg Arg Thr Gly Asn His	
1 5 10 15	
cac gga aat gtg tgg gac gat gac ctg ata cac tct ctg aac tcg ccc	96
His Gly Asn Val Trp Asp Asp Asp Leu Ile His Ser Leu Asn Ser Pro	
20 25 30	
tat ggg gca cct gca tat tat gag ctg ctt caa aag ctt att cag gag	144
Tyr Gly Ala Pro Ala Tyr Tyr Glu Leu Leu Gln Lys Leu Ile Gln Glu	
35 40 45	
atc aag cat tta ctt ttg act gaa atg gaa atg gat gat ggc gat cat	192
Ile Lys His Leu Leu Leu Thr Glu Met Glu Met Asp Asp Gly Asp His	
50 55 60	
gat tta atc aaa cgt ctt cag atc gtt gac act ttg gaa tgc ctg gga	240
Asp Leu Ile Lys Arg Leu Gln Ile Val Asp Thr Leu Glu Cys Leu Gly	
65 70 75 80	
atc gat aga cat ttt gaa cac gaa ata caa aca gct gct tta gat tac	288
Ile Asp Arg His Phe Glu His Glu Ile Gln Thr Ala Ala Leu Asp Tyr	
85 90 95	
gtt tac aga tgg tgg aac gaa aaa ggt atc ggg gag gga tca aga gat	336
Val Tyr Arg Trp Trp Asn Glu Lys Gly Ile Gly Glu Gly Ser Arg Asp	
100 105 110	
tcc ttc agc aaa gat ctg aac gct acg gct tta gga ttt cgc gct ctg	384
Ser Phe Ser Lys Asp Leu Asn Ala Thr Ala Leu Gly Phe Arg Ala Leu	
115 120 125	
cga ctg cat cga tat aac gta tcg tca ggt gtg ttg aag aat ttc aag	432
Arg Leu His Arg Tyr Asn Val Ser Ser Gly Val Leu Lys Asn Phe Lys	
130 135 140	
gat gaa aac ggg aag ttc ttc tgc aac ttt act ggt gaa gaa gga aga	480
Asp Glu Asn Gly Lys Phe Phe Cys Asn Phe Thr Gly Glu Glu Gly Arg	
145 150 155 160	
gga gat aaa caa gtg aga agc atg ttg tcg tta ctt cga gct tca gag	528
Gly Asp Lys Gln Val Arg Ser Met Leu Ser Leu Leu Arg Ala Ser Glu	
165 170 175	
att tcg ttt ccc gga gaa aaa gtg atg gaa gag gcc aag gca ttc aca	576
Ile Ser Phe Pro Gly Glu Lys Val Met Glu Glu Ala Lys Ala Phe Thr	
180 185 190	
aga gaa tat cta aac caa gtt tta gct gga cac ggg gat gtg act gac	624
Arg Glu Tyr Leu Asn Gln Val Leu Ala Gly His Gly Asp Val Thr Asp	
195 200 205	
gtg gat caa agc ctt ttg aga gag gtg aag tac gca ttg gag ttt cca	672
Val Asp Gln Ser Leu Leu Arg Glu Val Lys Tyr Ala Leu Glu Phe Pro	
210 215 220	
tgg cat tgc agt gtg cgg aga tgg gag gca agg agc ttt ctg gaa ata	720
Trp His Cys Ser Val Pro Arg Trp Glu Ala Arg Ser Phe Leu Glu Ile	
225 230 235 240	
tat gga cac aac cat tcg tgg ctg aag tcg aat atc aac caa aaa atg	768
Tyr Gly His Asn His Ser Trp Leu Lys Ser Asn Ile Asn Gln Lys Met	
245 250 255	
ttg aag tta gcc aaa ttg gac ttc aat att ctg caa tgc aaa cat cac	816
Leu Lys Leu Ala Lys Leu Asp Phe Asn Ile Leu Gln Cys Lys His His	
260 265 270	

-continued

aag gag ata cag ttt att aca agg tgg tgg aga gac tcg ggt ata tcg Lys Glu Ile Gln Phe Ile Thr Arg Trp Trp Arg Asp Ser Gly Ile Ser 275 280 285	864
cag ctg aat ttc tat cga aag cga cac gtg gaa tat tat tct tgg gtt Gln Leu Asn Phe Tyr Arg Lys Arg His Val Glu Tyr Tyr Ser Trp Val 290 295 300	912
ggt atg tgc att ttt gag cca gag ttc tct gaa agt aga att gcc ttc Val Met Cys Ile Phe Glu Pro Glu Phe Ser Glu Ser Arg Ile Ala Phe 305 310 315 320	960
gcc aaa act gct atc ctg tgt act gtt cta gat gac ctc tat gat acg Ala Lys Thr Ala Ile Leu Cys Thr Val Leu Asp Asp Leu Tyr Asp Thr 325 330 335	1008
cac gca aca ttg cat gaa atc aaa atc atg aca gag gga gtg aga cga His Ala Thr Leu His Glu Ile Lys Ile Met Thr Glu Gly Val Arg Arg 340 345 350	1056
tgg gat ctt tcg ttg aca gat gac ctc cca gac tac att aaa att gca Trp Asp Leu Ser Leu Thr Asp Asp Leu Pro Asp Tyr Ile Lys Ile Ala 355 360 365	1104
ttc cag ttc ttc ttc aat aca gtg aat gaa ttg ata gtt gaa atc gtg Phe Gln Phe Phe Phe Asn Thr Val Asn Glu Leu Ile Val Glu Ile Val 370 375 380	1152
aaa cgg caa ggg cgg gat atg aca acc ata gtt aaa gat tgc tgg aag Lys Arg Gln Gly Arg Asp Met Thr Thr Ile Val Lys Asp Cys Trp Lys 385 390 395 400	1200
cga tac att gag tct tat ctg caa gaa gcg gaa tgg ata gca act gga Arg Tyr Ile Glu Ser Tyr Leu Gln Glu Ala Glu Trp Ile Ala Thr Gly 405 410 415	1248
cat att ccc act ttt aac gaa tac ata aag aac ggc atg gct agc tca His Ile Pro Thr Phe Asn Glu Tyr Ile Lys Asn Gly Met Ala Ser Ser 420 425 430	1296
ggg atg tgt att cta aat ttg aat cca ctt ctc ttg ttg gat aaa ctt Gly Met Cys Ile Leu Asn Leu Asn Pro Leu Leu Leu Leu Asp Lys Leu 435 440 445	1344
ctc ccc gac aac att ctg gag caa ata cat tct cca tcc aag atc ctg Leu Pro Asp Asn Ile Leu Glu Gln Ile His Ser Pro Ser Lys Ile Leu 450 455 460	1392
gac ctc tta gaa ttg acg ggc aga atc gcc gat gac tta aaa gat ttc Asp Leu Leu Glu Leu Thr Gly Arg Ile Ala Asp Asp Leu Lys Asp Phe 465 470 475 480	1440
gag gac gag aag gaa cgc ggg gag atg gct tca tct tta cag tgt tat Glu Asp Glu Lys Glu Arg Gly Glu Met Ala Ser Ser Leu Gln Cys Tyr 485 490 495	1488
atg aaa gaa aat cct gaa tct aca gtg gaa aat gct tta aat cac ata Met Lys Glu Asn Pro Glu Ser Thr Val Glu Asn Ala Leu Asn His Ile 500 505 510	1536
aaa ggc atc ctt aat cgt tcc ctt gag gaa ttt aat tgg gag ttt atg Lys Gly Ile Leu Asn Arg Ser Leu Glu Glu Phe Asn Trp Glu Phe Met 515 520 525	1584
aag cag gat agt gtc cca atg tgt tgc aag aaa ttc act ttc aat ata Lys Gln Asp Ser Val Pro Met Cys Cys Lys Lys Phe Thr Phe Asn Ile 530 535 540	1632
ggt cga gga ctt caa ttc atc tac aaa tac aga gac ggc tta tac att Gly Arg Gly Leu Gln Phe Ile Tyr Lys Tyr Arg Asp Gly Leu Tyr Ile 545 550 555 560	1680
tct gac aag gaa gta aag gac cag ata ttc aaa att cta gtc cac caa Ser Asp Lys Glu Val Lys Asp Gln Ile Phe Lys Ile Leu Val His Gln 565 570 575	1728
ggt cca atg gag gaa tag tgatggtcctt gggtgtagtt gtctattatg Val Pro Met Glu Glu	1776

-continued

580

gtatattgca ttgacattta tgcttaaagg tgtttcttaa acgtttaggg cggaccgtta 1836
 aataagttgg caataattaa tatctcgag 1865

<210> SEQ ID NO 48
 <211> LENGTH: 581
 <212> TYPE: PRT
 <213> ORGANISM: Abies grandis

<400> SEQUENCE: 48

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

-continued

Trp Asp Leu Ser Leu Thr Asp Asp Leu Pro Asp Tyr Ile Lys Ile Ala
 355 360 365
 Phe Gln Phe Phe Phe Asn Thr Val Asn Glu Leu Ile Val Glu Ile Val
 370 375 380
 Lys Arg Gln Gly Arg Asp Met Thr Thr Ile Val Lys Asp Cys Trp Lys
 385 390 395 400
 Arg Tyr Ile Glu Ser Tyr Leu Gln Glu Ala Glu Trp Ile Ala Thr Gly
 405 410 415
 His Ile Pro Thr Phe Asn Glu Tyr Ile Lys Asn Gly Met Ala Ser Ser
 420 425 430
 Gly Met Cys Ile Leu Asn Leu Asn Pro Leu Leu Leu Asp Lys Leu
 435 440 445
 Leu Pro Asp Asn Ile Leu Glu Gln Ile His Ser Pro Ser Lys Ile Leu
 450 455 460
 Asp Leu Leu Glu Leu Thr Gly Arg Ile Ala Asp Asp Leu Lys Asp Phe
 465 470 475 480
 Glu Asp Glu Lys Glu Arg Gly Glu Met Ala Ser Ser Leu Gln Cys Tyr
 485 490 495
 Met Lys Glu Asn Pro Glu Ser Thr Val Glu Asn Ala Leu Asn His Ile
 500 505 510
 Lys Gly Ile Leu Asn Arg Ser Leu Glu Glu Phe Asn Trp Glu Phe Met
 515 520 525
 Lys Gln Asp Ser Val Pro Met Cys Cys Lys Lys Phe Thr Phe Asn Ile
 530 535 540
 Gly Arg Gly Leu Gln Phe Ile Tyr Lys Tyr Arg Asp Gly Leu Tyr Ile
 545 550 555 560
 Ser Asp Lys Glu Val Lys Asp Gln Ile Phe Lys Ile Leu Val His Gln
 565 570 575
 Val Pro Met Glu Glu
 580

<210> SEQ ID NO 49
 <211> LENGTH: 1785
 <212> TYPE: DNA
 <213> ORGANISM: Abies grandis
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (4)...(1782)
 <223> OTHER INFORMATION: gamma-humulene synthase

<400> SEQUENCE: 49

tcc atg gct cag att tct gaa tct gta tca ccc tct acc gat ttg aag 48
 Met Ala Gln Ile Ser Glu Ser Val Ser Pro Ser Thr Asp Leu Lys
 1 5 10 15
 agc acc gaa tct tcc att acc tct aat cga cat gga aat atg tgg gag 96
 Ser Thr Glu Ser Ser Ile Thr Ser Asn Arg His Gly Asn Met Trp Glu
 20 25 30
 gac gat cgc ata cag tct ctc aac tca cct tat ggg gca cct gca tat 144
 Asp Asp Arg Ile Gln Ser Leu Asn Ser Pro Tyr Gly Ala Pro Ala Tyr
 35 40 45
 caa gaa cgc agc gaa aag ctt att gaa gag atc aaa ctt tta ttt ttg 192
 Gln Glu Arg Ser Glu Lys Leu Ile Glu Glu Ile Lys Leu Leu Phe Leu
 50 55 60
 agt gac atg gac gat agc tgc aat gat agc gat cgt gat tta atc aaa 240
 Ser Asp Met Asp Asp Ser Cys Asn Asp Ser Asp Arg Asp Leu Ile Lys
 65 70 75
 cgt ctt gag atc gtt gat act gtc gag tgt ctg gga att gat cga cat 288

-continued

```

ggg caa gat atg gcg gcc tac ata aga aaa aat gca tgg gag cga tac 1248
Gly Gln Asp Met Ala Ala Tyr Ile Arg Lys Asn Ala Trp Glu Arg Tyr
400 405 410 415

ctt gaa gct tat ctg caa gat gcg gaa tgg ata gcc act gga cat gtc 1296
Leu Glu Ala Tyr Leu Gln Asp Ala Glu Trp Ile Ala Thr Gly His Val
420 425 430

ccc acc ttt gat gag tac ttg aat aat ggc aca cca aac act ggg atg 1344
Pro Thr Phe Asp Glu Tyr Leu Asn Asn Gly Thr Pro Asn Thr Gly Met
435 440 445

tgt gta ttg aat ttg att ccg ctt ctg tta atg ggt gaa cat tta cca 1392
Cys Val Leu Asn Leu Ile Pro Leu Leu Leu Met Gly Glu His Leu Pro
450 455 460

atc gac att ctg gag caa ata ttc ttg ccc tcc agg ttc cac cat ctc 1440
Ile Asp Ile Leu Glu Gln Ile Phe Leu Pro Ser Arg Phe His His Leu
465 470 475

att gaa ttg gct tcc agg ctc gtc gat gac gcg aga gat ttc cag gcg 1488
Ile Glu Leu Ala Ser Arg Leu Val Asp Asp Ala Arg Asp Phe Gln Ala
480 485 490 495

gag aag gat cat ggg gat tta tcg tgt att gag tgt tat tta aaa gat 1536
Glu Lys Asp His Gly Asp Leu Ser Cys Ile Glu Cys Tyr Leu Lys Asp
500 505 510

cat cct gag tct aca gta gaa gat gct tta aat cat gtt aat ggc ctc 1584
His Pro Glu Ser Thr Val Glu Asp Ala Leu Asn His Val Asn Gly Leu
515 520 525

ctt ggc aat tgc ctt ctg gaa atg aat tgg aag ttc tta aag aag cag 1632
Leu Gly Asn Cys Leu Leu Glu Met Asn Trp Lys Phe Leu Lys Lys Gln
530 535 540

gac agt gtg cca ctc tcg tgt aag aag tac agc ttc cat gta ttg gca 1680
Asp Ser Val Pro Leu Ser Cys Lys Lys Tyr Ser Phe His Val Leu Ala
545 550 555

cga agc atc caa ttc atg tac aat caa ggc gat ggc ttc tcc att tcg 1728
Arg Ser Ile Gln Phe Met Tyr Asn Gln Gly Asp Gly Phe Ser Ile Ser
560 565 570 575

aac aaa gtg atc aag gat caa gtg cag aaa gtt ctt att gtc ccc gtg 1776
Asn Lys Val Ile Lys Asp Gln Val Gln Lys Val Leu Ile Val Pro Val
580 585 590

cct att tga 1785
Pro Ile

```

<210> SEQ ID NO 50

<211> LENGTH: 593

<212> TYPE: PRT

<213> ORGANISM: Abies grandis

<400> SEQUENCE: 50

```

Met Ala Gln Ile Ser Glu Ser Val Ser Pro Ser Thr Asp Leu Lys Ser
 1 5 10 15

Thr Glu Ser Ser Ile Thr Ser Asn Arg His Gly Asn Met Trp Glu Asp
 20 25 30

Asp Arg Ile Gln Ser Leu Asn Ser Pro Tyr Gly Ala Pro Ala Tyr Gln
 35 40 45

Glu Arg Ser Glu Lys Leu Ile Glu Glu Ile Lys Leu Leu Phe Leu Ser
 50 55 60

Asp Met Asp Asp Ser Cys Asn Asp Ser Asp Arg Asp Leu Ile Lys Arg
 65 70 75 80

Leu Glu Ile Val Asp Thr Val Glu Cys Leu Gly Ile Asp Arg His Phe
 85 90 95

Gln Pro Glu Ile Lys Leu Ala Leu Asp Tyr Val Tyr Arg Cys Trp Asn
 100 105 110

```

-continued

Glu Arg Gly Ile Gly Glu Gly Ser Arg Asp Ser Leu Lys Lys Asp Leu
 115 120 125
 Asn Ala Thr Ala Leu Gly Phe Arg Ala Leu Arg Leu His Arg Tyr Asn
 130 135 140
 Val Ser Ser Gly Val Leu Glu Asn Phe Arg Asp Asp Asn Gly Gln Phe
 145 150 155 160
 Phe Cys Gly Ser Thr Val Glu Glu Glu Gly Ala Glu Ala Tyr Asn Lys
 165 170 175
 His Val Arg Cys Met Leu Ser Leu Ser Arg Ala Ser Asn Ile Leu Phe
 180 185 190
 Pro Gly Glu Lys Val Met Glu Glu Ala Lys Ala Phe Thr Thr Asn Tyr
 195 200 205
 Leu Lys Lys Val Leu Ala Gly Arg Glu Ala Thr His Val Asp Glu Ser
 210 215 220
 Leu Leu Gly Glu Val Lys Tyr Ala Leu Glu Phe Pro Trp His Cys Ser
 225 230 235 240
 Val Gln Arg Trp Glu Ala Arg Ser Phe Ile Glu Ile Phe Gly Gln Ile
 245 250 255
 Asp Ser Glu Leu Lys Ser Asn Leu Ser Lys Lys Met Leu Glu Leu Ala
 260 265 270
 Lys Leu Asp Phe Asn Ile Leu Gln Cys Thr His Gln Lys Glu Leu Gln
 275 280 285
 Ile Ile Ser Arg Trp Phe Ala Asp Ser Ser Ile Ala Ser Leu Asn Phe
 290 295 300
 Tyr Arg Lys Cys Tyr Val Glu Phe Tyr Phe Trp Met Ala Ala Ala Ile
 305 310 315 320
 Ser Glu Pro Glu Phe Ser Gly Ser Arg Val Ala Phe Thr Lys Ile Ala
 325 330 335
 Ile Leu Met Thr Met Leu Asp Asp Leu Tyr Asp Thr His Gly Thr Leu
 340 345 350
 Asp Gln Leu Lys Ile Phe Thr Glu Gly Val Arg Arg Trp Asp Val Ser
 355 360 365
 Leu Val Glu Gly Leu Pro Asp Phe Met Lys Ile Ala Phe Glu Phe Trp
 370 375 380
 Leu Lys Thr Ser Asn Glu Leu Ile Ala Glu Ala Val Lys Ala Gln Gly
 385 390 395 400
 Gln Asp Met Ala Ala Tyr Ile Arg Lys Asn Ala Trp Glu Arg Tyr Leu
 405 410 415
 Glu Ala Tyr Leu Gln Asp Ala Glu Trp Ile Ala Thr Gly His Val Pro
 420 425 430
 Thr Phe Asp Glu Tyr Leu Asn Asn Gly Thr Pro Asn Thr Gly Met Cys
 435 440 445
 Val Leu Asn Leu Ile Pro Leu Leu Leu Met Gly Glu His Leu Pro Ile
 450 455 460
 Asp Ile Leu Glu Gln Ile Phe Leu Pro Ser Arg Phe His His Leu Ile
 465 470 475 480
 Glu Leu Ala Ser Arg Leu Val Asp Asp Ala Arg Asp Phe Gln Ala Glu
 485 490 495
 Lys Asp His Gly Asp Leu Ser Cys Ile Glu Cys Tyr Leu Lys Asp His
 500 505 510
 Pro Glu Ser Thr Val Glu Asp Ala Leu Asn His Val Asn Gly Leu Leu
 515 520 525

-continued

Gly Asn Cys Leu Leu Glu Met Asn Trp Lys Phe Leu Lys Lys Gln Asp
 530 535 540
 Ser Val Pro Leu Ser Cys Lys Lys Tyr Ser Phe His Val Leu Ala Arg
 545 550 555 560
 Ser Ile Gln Phe Met Tyr Asn Gln Gly Asp Gly Phe Ser Ile Ser Asn
 565 570 575
 Lys Val Ile Lys Asp Gln Val Gln Lys Val Leu Ile Val Pro Val Pro
 580 585 590

Ile

<210> SEQ ID NO 51
 <211> LENGTH: 2024
 <212> TYPE: DNA
 <213> ORGANISM: Lycopersicon esculentum
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (32)...(1675)
 <223> OTHER INFORMATION: VFNT germacrene C synthase

<400> SEQUENCE: 51

aaaaaaagcc aaaccttaga acaacaagc a atg gct gct tct tct gct gat 52
 Met Ala Ala Ser Ser Ala Asp
 1 5
 aag tgt cgc ccc ttg gct aat ttt cac cca tct gtt tgg gga tat cat 100
 Lys Cys Arg Pro Leu Ala Asn Phe His Pro Ser Val Trp Gly Tyr His
 10 15 20
 ttc ctt tct tat act cat gaa att act aat caa gaa aaa gtt gaa gtt 148
 Phe Leu Ser Tyr Thr His Glu Ile Thr Asn Gln Glu Lys Val Glu Val
 25 30 35
 gat gag tac aaa gag aca att aga aaa atg ctg gtg gaa act tgc gac 196
 Asp Glu Tyr Lys Glu Thr Ile Arg Lys Met Leu Val Glu Thr Cys Asp
 40 45 50 55
 aat agc act caa aag ctt gtg ttg ata gac gcg atg caa cga ttg gga 244
 Asn Ser Thr Gln Lys Leu Val Leu Ile Asp Ala Met Gln Arg Leu Gly
 60 65 70
 gtg gct tat cat ttc gat aat gaa att gaa aca tcc att caa aac att 292
 Val Ala Tyr His Phe Asp Asn Glu Ile Glu Thr Ser Ile Gln Asn Ile
 75 80 85
 ttt gat gca tcg tcc aaa cag aat gat aat gac aac aac ctt tac gtt 340
 Phe Asp Ala Ser Ser Lys Gln Asn Asp Asn Asp Asn Asn Leu Tyr Val
 90 95 100
 gtg tct ctt cgt ttt cga ctt gtg agg caa caa ggc cat tac atg tct 388
 Val Ser Leu Arg Phe Arg Leu Val Arg Gln Gln Gly His Tyr Met Ser
 105 110 115
 tca gat gtg ttc aag caa ttc acc aac caa gat ggg aaa ttc aag gaa 436
 Ser Asp Val Phe Lys Gln Phe Thr Asn Gln Asp Gly Lys Phe Lys Glu
 120 125 130 135
 aca ctt act aat gat gtc caa gga tta ttg agt ttg tat gaa gca tca 484
 Thr Leu Thr Asn Asp Val Gln Gly Leu Leu Ser Leu Tyr Glu Ala Ser
 140 145 150
 cat ctg aga gtg cgt aat gag gag att ctt gaa gaa gct ctt aca ttt 532
 His Leu Arg Val Arg Asn Glu Glu Ile Leu Glu Glu Ala Leu Thr Phe
 155 160 165
 acc acc act cat ctc gag tct att gtc tcc aac ttg agc aat aat aat 580
 Thr Thr Thr His Leu Glu Ser Ile Val Ser Asn Leu Ser Asn Asn Asn
 170 175 180
 aac tct ctt aag gtt gaa gtt ggt gaa gcc tta act cag cct att cgc 628
 Asn Ser Leu Lys Val Glu Val Gly Glu Ala Leu Thr Gln Pro Ile Arg
 185 190 195
 atg act tta cca agg atg gga gct aga aaa tac ata tcc att tac gaa 676

-continued

Met Thr Leu Pro Arg Met Gly Ala Arg Lys Tyr Ile Ser Ile Tyr Glu 200 205 210 215	
aac aat gat gca cac cac cat ttg ctt ttg aaa ttt gct aaa ttg gat Asn Asn Asp Ala His His His Leu Leu Leu Lys Phe Ala Lys Leu Asp 220 225 230	724
ttt aac atg ctg caa aag ttt cac caa aga gag ctt agt gat ctt aca Phe Asn Met Leu Gln Lys Phe His Gln Arg Glu Leu Ser Asp Leu Thr 235 240 245	772
agg tgg tgg aaa gat ttg gat ttt gca aat aaa tat cca tat gca aga Arg Trp Trp Lys Asp Leu Asp Phe Ala Asn Lys Tyr Pro Tyr Ala Arg 250 255 260	820
gac agg ttg gtt gag tgt tac ttc tgg ata tta gga gtg tat ttt gag Asp Arg Leu Val Glu Cys Tyr Phe Trp Ile Leu Gly Val Tyr Phe Glu 265 270 275	868
cca aaa tat agt cgt gcg aga aaa atg atg aca aaa gta ctc aac ctg Pro Lys Tyr Ser Arg Ala Arg Lys Met Met Thr Lys Val Leu Asn Leu 280 285 290 295	916
acc tcc att att gag gac act ttt gat gct tat gca acc ttt gag gaa Thr Ser Ile Ile Asp Asp Thr Phe Asp Ala Tyr Ala Thr Phe Asp Glu 300 305 310	964
ctt gtg act ttc aat gat gca atc cag aga tgg gat gct aat gca att Leu Val Thr Phe Asn Asp Ala Ile Gln Arg Trp Asp Ala Asn Ala Ile 315 320 325	1012
gat tca ata caa cca tat atg aga cct gct tat caa gct ctt cta gac Asp Ser Ile Gln Pro Tyr Met Arg Pro Ala Tyr Gln Ala Leu Leu Asp 330 335 340	1060
att tac agt gaa atg gaa caa gtg ttg tcc aaa gaa ggt aaa ctg gac Ile Tyr Ser Glu Met Glu Gln Val Leu Ser Lys Glu Gly Lys Leu Asp 345 350 355	1108
cgt gta tac tat gca aaa aat gag atg aaa aag ttg gtg aga gcc tat Arg Val Tyr Tyr Ala Lys Asn Glu Met Lys Lys Leu Val Arg Ala Tyr 360 365 370 375	1156
ttt aag gaa acc caa tgg ttg aat gat tgt gac cat att cca aaa tat Phe Lys Glu Thr Gln Trp Leu Asn Asp Cys Asp His Ile Pro Lys Tyr 380 385 390	1204
gag gaa caa gtg gag aat gca atc gta agt gct ggc tat atg atg ata Glu Glu Gln Val Glu Asn Ala Ile Val Ser Ala Gly Tyr Met Met Ile 395 400 405	1252
tca aca act tgc ttg gtc ggt ata gaa gaa ttt ata tcc cac gag act Ser Thr Thr Cys Leu Val Gly Ile Glu Glu Phe Ile Ser His Glu Thr 410 415 420	1300
ttt gaa tgg ttg atg aat gag tct gtg att gtt cga gct tcc gca ttg Phe Glu Trp Leu Met Asn Glu Ser Val Ile Val Arg Ala Ser Ala Leu 425 430 435	1348
att gcc aga gca atg aac gat att gtt gga cat gaa gat gaa caa gaa Ile Ala Arg Ala Met Asn Asp Ile Val Gly His Glu Asp Glu Gln Glu 440 445 450 455	1396
aga gga cat gta gct tca ctt att gaa tgt tac atg aaa gat tat gga Arg Gly His Val Ala Ser Leu Ile Glu Cys Tyr Met Lys Asp Tyr Gly 460 465 470	1444
gct tca aag caa gag act tac att aag ttc ctg aaa gag gtc acc aat Ala Ser Lys Gln Glu Thr Tyr Ile Lys Phe Leu Lys Glu Val Thr Asn 475 480 485	1492
gca tgg aag gac ata aac aaa caa ttc tcc cgt cca act gaa gta cca Ala Trp Lys Asp Ile Asn Lys Gln Phe Ser Arg Pro Thr Glu Val Pro 490 495 500	1540
atg ttt gtc ctt gaa cga gtt cta aat ttg aca cgt gtg gct gac acg Met Phe Val Leu Glu Arg Val Leu Asn Leu Thr Arg Val Ala Asp Thr 505 510 515	1588

-continued

tta tat aag gag aaa gat aca tat tca acc gcc aaa gga aaa ctt aaa	1636
Leu Tyr Lys Glu Lys Asp Thr Tyr Ser Thr Ala Lys Gly Lys Leu Lys	
520 525 530 535	
aac atg att aat cca ata cta att gaa tct gtc aaa ata taa	1678
Asn Met Ile Asn Pro Ile Leu Ile Glu Ser Val Lys Ile	
540 545	
atataatgct gaaattgcac cttcatcatc caactattca cagcaaaata aggcatataa	1738
taaattggaag actcacaaca tatgagttgt taattcctgg gatgtttgaa ataaacaata	1798
attgttttta ttaatttgc taagccaaag tgaatatatac aacacttgag ttgtatataa	1858
tcatgtttta tctcatttcc agcttgtgag tttggattat tatattgtta attatcatca	1918
ctttataatg tactgtaatc gtattgtatt tgtattgtag tgttgcata ataaaatttg	1978
aataaaatat atttttgttt caattccaaa aaaaaaaaa aaaaaa	2024

<210> SEQ ID NO 52
 <211> LENGTH: 548
 <212> TYPE: PRT
 <213> ORGANISM: Lycopersicon esculentum

<400> SEQUENCE: 52

Met Ala Ala Ser Ser Ala Asp Lys Cys Arg Pro Leu Ala Asn Phe His	
1 5 10 15	
Pro Ser Val Trp Gly Tyr His Phe Leu Ser Tyr Thr His Glu Ile Thr	
20 25 30	
Asn Gln Glu Lys Val Glu Val Asp Glu Tyr Lys Glu Thr Ile Arg Lys	
35 40 45	
Met Leu Val Glu Thr Cys Asp Asn Ser Thr Gln Lys Leu Val Leu Ile	
50 55 60	
Asp Ala Met Gln Arg Leu Gly Val Ala Tyr His Phe Asp Asn Glu Ile	
65 70 75 80	
Glu Thr Ser Ile Gln Asn Ile Phe Asp Ala Ser Ser Lys Gln Asn Asp	
85 90 95	
Asn Asp Asn Asn Leu Tyr Val Val Ser Leu Arg Phe Arg Leu Val Arg	
100 105 110	
Gln Gln Gly His Tyr Met Ser Ser Asp Val Phe Lys Gln Phe Thr Asn	
115 120 125	
Gln Asp Gly Lys Phe Lys Glu Thr Leu Thr Asn Asp Val Gln Gly Leu	
130 135 140	
Leu Ser Leu Tyr Glu Ala Ser His Leu Arg Val Arg Asn Glu Glu Ile	
145 150 155 160	
Leu Glu Glu Ala Leu Thr Phe Thr Thr Thr His Leu Glu Ser Ile Val	
165 170 175	
Ser Asn Leu Ser Asn Asn Asn Ser Leu Lys Val Glu Val Gly Glu	
180 185 190	
Ala Leu Thr Gln Pro Ile Arg Met Thr Leu Pro Arg Met Gly Ala Arg	
195 200 205	
Lys Tyr Ile Ser Ile Tyr Glu Asn Asn Asp Ala His His His Leu Leu	
210 215 220	
Leu Lys Phe Ala Lys Leu Asp Phe Asn Met Leu Gln Lys Phe His Gln	
225 230 235 240	
Arg Glu Leu Ser Asp Leu Thr Arg Trp Trp Lys Asp Leu Asp Phe Ala	
245 250 255	
Asn Lys Tyr Pro Tyr Ala Arg Asp Arg Leu Val Glu Cys Tyr Phe Trp	
260 265 270	

-continued

Ile Leu Gly Val Tyr Phe Glu Pro Lys Tyr Ser Arg Ala Arg Lys Met
 275 280 285
 Met Thr Lys Val Leu Asn Leu Thr Ser Ile Ile Asp Asp Thr Phe Asp
 290 295 300
 Ala Tyr Ala Thr Phe Asp Glu Leu Val Thr Phe Asn Asp Ala Ile Gln
 305 310 315 320
 Arg Trp Asp Ala Asn Ala Ile Asp Ser Ile Gln Pro Tyr Met Arg Pro
 325 330 335
 Ala Tyr Gln Ala Leu Leu Asp Ile Tyr Ser Glu Met Glu Gln Val Leu
 340 345 350
 Ser Lys Glu Gly Lys Leu Asp Arg Val Tyr Tyr Ala Lys Asn Glu Met
 355 360 365
 Lys Lys Leu Val Arg Ala Tyr Phe Lys Glu Thr Gln Trp Leu Asn Asp
 370 375 380
 Cys Asp His Ile Pro Lys Tyr Glu Glu Gln Val Glu Asn Ala Ile Val
 385 390 395 400
 Ser Ala Gly Tyr Met Met Ile Ser Thr Thr Cys Leu Val Gly Ile Glu
 405 410 415
 Glu Phe Ile Ser His Glu Thr Phe Glu Trp Leu Met Asn Glu Ser Val
 420 425 430
 Ile Val Arg Ala Ser Ala Leu Ile Ala Arg Ala Met Asn Asp Ile Val
 435 440 445
 Gly His Glu Asp Glu Gln Glu Arg Gly His Val Ala Ser Leu Ile Glu
 450 455 460
 Cys Tyr Met Lys Asp Tyr Gly Ala Ser Lys Gln Glu Thr Tyr Ile Lys
 465 470 475 480
 Phe Leu Lys Glu Val Thr Asn Ala Trp Lys Asp Ile Asn Lys Gln Phe
 485 490 495
 Ser Arg Pro Thr Glu Val Pro Met Phe Val Leu Glu Arg Val Leu Asn
 500 505 510
 Leu Thr Arg Val Ala Asp Thr Leu Tyr Lys Glu Lys Asp Thr Tyr Ser
 515 520 525
 Thr Ala Lys Gly Lys Leu Lys Asn Met Ile Asn Pro Ile Leu Ile Glu
 530 535 540
 Ser Val Lys Ile
 545

<210> SEQ ID NO 53
 <211> LENGTH: 1912
 <212> TYPE: DNA
 <213> ORGANISM: Salvia officinalis
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (26)...(1795)
 <223> OTHER INFORMATION: (+)-sabinene syntase

<400> SEQUENCE: 53

agcaatatta caactaacaa taaaa atg tct tcc att agc ata aac ata gct 52
 Met Ser Ser Ile Ser Ile Asn Ile Ala
 1 5
 atg cca ctg aat tcc ctc cac aac ttt gag agg aaa cct tca aaa gca 100
 Met Pro Leu Asn Ser Leu His Asn Phe Glu Arg Lys Pro Ser Lys Ala
 10 15 20 25
 tgg tct acc tct tgc act gca ccc gca gct cgc ctc cgg gca tct tcc 148
 Trp Ser Thr Ser Cys Thr Ala Pro Ala Ala Arg Leu Arg Ala Ser Ser
 30 35 40
 tcc tta caa caa gaa aaa cct cac caa atc cga cgc tct ggg gat tac 196

-continued

Ser	Leu	Gln	Gln	Glu	Lys	Pro	His	Gln	Ile	Arg	Arg	Ser	Gly	Asp	Tyr	
			45					50					55			
caa	ccc	tct	ctt	tgg	gat	ttc	aat	tac	ata	cag	tct	ctc	aac	act	ccg	244
Gln	Pro	Ser	Leu	Trp	Asp	Phe	Asn	Tyr	Ile	Gln	Ser	Leu	Asn	Thr	Pro	
	60						65					70				
tat	aag	gag	cag	aga	cac	ttt	aat	agg	caa	gca	gag	ttg	att	atg	caa	292
Tyr	Lys	Glu	Gln	Arg	His	Phe	Asn	Arg	Gln	Ala	Glu	Leu	Ile	Met	Gln	
	75					80					85					
gtg	agg	atg	ttg	ctc	aag	gta	aag	atg	gag	gca	att	caa	cag	ttg	gag	340
Val	Arg	Met	Leu	Leu	Lys	Val	Lys	Met	Glu	Ala	Ile	Gln	Gln	Leu	Glu	
	90				95					100					105	
ttg	att	gat	gac	ttg	caa	tac	ctg	gga	ctg	tct	tat	ttc	ttt	caa	gat	388
Leu	Ile	Asp	Asp	Leu	Gln	Tyr	Leu	Gly	Leu	Ser	Tyr	Phe	Phe	Gln	Asp	
				110					115					120		
gag	att	aaa	caa	atc	tta	agt	tct	ata	cac	aat	gag	ccc	aga	tat	ttc	436
Glu	Ile	Lys	Ile	Leu	Ser	Ser	Ser	Ile	His	Asn	Glu	Pro	Arg	Tyr	Phe	
			125					130					135			
cac	aat	aat	gat	ttg	tat	ttc	aca	gct	ctt	gga	ttc	aga	atc	ctc	aga	484
His	Asn	Asn	Asp	Leu	Tyr	Phe	Thr	Ala	Leu	Gly	Phe	Arg	Ile	Leu	Arg	
	140						145					150				
caa	cat	ggg	ttt	aat	gtt	tcc	gaa	gat	gta	ttt	gat	tgt	ttc	aaa	att	532
Gln	His	Gly	Phe	Asn	Val	Ser	Glu	Asp	Val	Phe	Asp	Cys	Phe	Lys	Ile	
	155					160					165					
gag	aag	tgc	agt	gat	ttc	aat	gca	aac	ctt	gct	caa	gat	acg	aag	gga	580
Glu	Lys	Cys	Ser	Asp	Phe	Asn	Ala	Asn	Leu	Ala	Gln	Asp	Thr	Lys	Gly	
	170				175					180					185	
atg	tta	caa	ctt	tat	gaa	gca	tct	ttc	ctt	ttg	aga	gaa	ggg	gaa	gat	628
Met	Leu	Gln	Leu	Tyr	Glu	Ala	Ser	Phe	Leu	Leu	Arg	Glu	Gly	Glu	Asp	
				190					195					200		
aca	ttg	gag	cta	gca	aga	cga	ttt	tcc	acc	aga	tct	cta	cga	gaa	aaa	676
Thr	Leu	Glu	Leu	Ala	Arg	Arg	Phe	Ser	Thr	Arg	Ser	Leu	Arg	Glu	Lys	
			205					210					215			
ttt	gat	gaa	ggg	ggg	gat	gaa	att	gat	gaa	gat	cta	tca	tcg	tg	att	724
Phe	Asp	Glu	Gly	Gly	Asp	Glu	Ile	Asp	Glu	Asp	Leu	Ser	Ser	Trp	Ile	
		220					225						230			
cgc	cat	tcc	ttg	gat	ctt	cct	ctt	cat	tgg	agg	gtc	caa	gga	tta	gag	772
Arg	His	Ser	Leu	Asp	Leu	Pro	Leu	His	Trp	Arg	Val	Gln	Gly	Leu	Glu	
	235					240					245					
gca	aga	tgg	ttc	tta	gat	gct	tat	gcg	agg	agg	ccg	gac	atg	aat	cca	820
Ala	Arg	Trp	Phe	Leu	Asp	Ala	Tyr	Ala	Arg	Arg	Pro	Asp	Met	Asn	Pro	
	250				255					260					265	
ctt	att	ttc	aaa	ctc	gcc	aaa	ctc	aac	ttc	aat	att	gtt	cag	gca	aca	868
Leu	Ile	Phe	Lys	Leu	Ala	Lys	Leu	Asn	Phe	Asn	Ile	Val	Gln	Ala	Thr	
			270						275					280		
tat	caa	gaa	gaa	ctg	aaa	gat	atc	tca	agg	tgg	tgg	aat	agt	tcg	tgc	916
Tyr	Gln	Glu	Glu	Leu	Lys	Asp	Ile	Ser	Arg	Trp	Trp	Asn	Ser	Ser	Cys	
			285					290					295			
ctt	gct	gag	aaa	ctc	cca	ttt	gtg	aga	gat	agg	att	gtg	gaa	tgc	ttc	964
Leu	Ala	Glu	Lys	Leu	Pro	Phe	Val	Arg	Asp	Arg	Ile	Val	Glu	Cys	Phe	
		300					305					310				
ttt	tgg	gcc	atc	gcg	gct	ttt	gag	cct	cac	caa	tat	agt	tat	cag	aga	1012
Phe	Trp	Ala	Ile	Ala	Ala	Phe	Glu	Pro	His	Gln	Tyr	Ser	Tyr	Gln	Arg	
		315				320						325				
aaa	atg	gcc	gcc	ggt	att	att	act	ttc	ata	aca	att	atc	gat	gat	ggt	1060
Lys	Met	Ala	Ala	Val	Ile	Ile	Thr	Phe	Ile	Thr	Ile	Ile	Asp	Asp	Val	
	330				335				340						345	
tat	gat	gtg	tat	gga	aca	ata	gaa	gaa	cta	gaa	cta	tta	aca	gat	atg	1108
Tyr	Asp	Val	Tyr	Gly	Thr	Ile	Glu	Glu	Leu	Glu	Leu	Leu	Thr	Asp	Met	
				350					355						360	

-continued

att cgc aga tgg gat aat aaa tca ata agc caa ctt cca tat tat atg Ile Arg Arg Trp Asp Asn Lys Ser Ile Ser Gln Leu Pro Tyr Tyr Met 365 370 375	1156
caa gtg tgc tat ttg gca cta tac aac ttc gtt tct gag cgg gct tac Gln Val Cys Tyr Leu Ala Leu Tyr Asn Phe Val Ser Glu Arg Ala Tyr 380 385 390	1204
gat att cta aaa gat caa cat ttc aac agc atc cca tat tta cag aga Asp Ile Leu Lys Asp Gln His Phe Asn Ser Ile Pro Tyr Leu Gln Arg 395 400 405	1252
tcg tgg gta agt ttg gtt gaa gga tat ctt aag gag gca tac tgg tac Ser Trp Val Ser Leu Val Glu Gly Tyr Leu Lys Glu Ala Tyr Trp Tyr 410 415 420 425	1300
tac aat ggc tat aaa cca agc ttg gaa gaa tat ctc aac aac gcc aag Tyr Asn Gly Tyr Lys Pro Ser Leu Glu Glu Tyr Leu Asn Asn Ala Lys 430 435 440	1348
att tca ata tcg gct cct aca atc ata tcc cag ctt tat ttt aca tta Ile Ser Ile Ser Ala Pro Thr Ile Ile Ser Gln Leu Tyr Phe Thr Leu 445 450 455	1396
gca aac tcg att gat gaa aca gct atc gag agc ttg tac caa tat cat Ala Asn Ser Ile Asp Glu Thr Ala Ile Glu Ser Leu Tyr Gln Tyr His 460 465 470	1444
aac ata ctt tac cta tca gga acc ata tta agg ctt gct gac gat ctt Asn Ile Leu Tyr Leu Ser Gly Thr Ile Leu Arg Leu Ala Asp Asp Leu 475 480 485	1492
ggg aca tca caa cat gag ctg gag aga gga gac gta ccg aaa gca atc Gly Thr Ser Gln His Glu Leu Glu Arg Gly Asp Val Pro Lys Ala Ile 490 495 500 505	1540
cag tgc tac atg aat gac aca aat gct tcg gag aga gag gcg gtg gaa Gln Cys Tyr Met Asn Asp Thr Asn Ala Ser Glu Arg Glu Ala Val Glu 510 515 520	1588
cac gtg aag ttt ctg ata agg gag gcg tgg aag gag atg aac acg gtc His Val Lys Phe Leu Ile Arg Glu Ala Trp Lys Glu Met Asn Thr Val 525 530 535	1636
aca aca gcc agc gat tgt ccg ttt acg gat gat ttg gtt gcg gcc gca Thr Thr Ala Ser Asp Cys Pro Phe Thr Asp Asp Leu Val Ala Ala Ala 540 545 550	1684
gct aat ctt gca agg gcg gct cag ttt ata tat ctc gac ggg gat ggg Ala Asn Leu Ala Arg Ala Ala Gln Phe Ile Tyr Leu Asp Gly Asp Gly 555 560 565	1732
cat ggc gtg caa cac tca gaa ata cat caa cag atg gga ggc ctg cta His Gly Val Gln His Ser Glu Ile His Gln Gln Met Gly Gly Leu Leu 570 575 580 585	1780
ttc cag cct tat gtc tga ataaatcgaa aatccaacct actatgtatc Phe Gln Pro Tyr Val 590	1828
cctcgataat atattcttgg ggtaacatg ttaattaaa gttctaattd aaagagctga	1888
atcgatcctc aaaaaaaaaa aaaa	1912
<210> SEQ ID NO 54 <211> LENGTH: 590 <212> TYPE: PRT <213> ORGANISM: Salvia officinalis <400> SEQUENCE: 54	
Met Ser Ser Ile Ser Ile Asn Ile Ala Met Pro Leu Asn Ser Leu His 1 5 10 15	
Asn Phe Glu Arg Lys Pro Ser Lys Ala Trp Ser Thr Ser Cys Thr Ala 20 25 30	
Pro Ala Ala Arg Leu Arg Ala Ser Ser Ser Leu Gln Gln Glu Lys Pro	

-continued

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

-continued

Ala Ile Glu Ser Leu Tyr Gln Tyr His Asn Ile Leu Tyr Leu Ser Gly
 465 470 475 480

Thr Ile Leu Arg Leu Ala Asp Asp Leu Gly Thr Ser Gln His Glu Leu
 485 490 495

Glu Arg Gly Asp Val Pro Lys Ala Ile Gln Cys Tyr Met Asn Asp Thr
 500 505 510

Asn Ala Ser Glu Arg Glu Ala Val Glu His Val Lys Phe Leu Ile Arg
 515 520 525

Glu Ala Trp Lys Glu Met Asn Thr Val Thr Thr Ala Ser Asp Cys Pro
 530 535 540

Phe Thr Asp Asp Leu Val Ala Ala Ala Ala Asn Leu Ala Arg Ala Ala
 545 550 555 560

Gln Phe Ile Tyr Leu Asp Gly Asp Gly His Gly Val Gln His Ser Glu
 565 570 575

Ile His Gln Gln Met Gly Gly Leu Leu Phe Gln Pro Tyr Val
 580 585 590

<210> SEQ ID NO 55
 <211> LENGTH: 2861
 <212> TYPE: DNA
 <213> ORGANISM: Abies grandis
 <220> FEATURE:
 <221> NAME/KEY: CDS
 <222> LOCATION: (3)...(2606)
 <223> OTHER INFORMATION: abietadiene synthase

<400> SEQUENCE: 55

ag atg gcc atg cct tcc tct tca ttg tca tca cag att ccc act gct 47
 Met Ala Met Pro Ser Ser Leu Ser Ser Gln Ile Pro Thr Ala
 1 5 10 15

gct cat cat cta act gct aac gca caa tcc att ccg cat ttc tcc acg 95
 Ala His His Leu Thr Ala Asn Ala Gln Ser Ile Pro His Phe Ser Thr
 20 25 30

acg ctg aat gct gga agc agt gct agc aaa cgg aga agc ttg tac cta 143
 Thr Leu Asn Ala Gly Ser Ser Ala Ser Lys Arg Arg Ser Leu Tyr Leu
 35 40 45

cga tgg ggt aaa ggt tca aac aag atc att gcc tgt gtt gga gaa ggt 191
 Arg Trp Gly Lys Gly Ser Asn Lys Ile Ile Ala Cys Val Gly Glu Gly
 50 55 60

ggt gca acc tct gtt cct tat cag tct gct gaa aag aat gat tgg ctt 239
 Gly Ala Thr Ser Val Pro Tyr Gln Ser Ala Glu Lys Asn Asp Ser Leu
 65 70 75

tct tct tct aca ttg gtg aaa cga gaa ttt cct cca gga ttt tgg aag 287
 Ser Ser Ser Thr Leu Val Lys Arg Glu Phe Pro Pro Gly Phe Trp Lys
 80 85 90 95

gat gat ctt atc gat tct cta acg tca tct cac aag gtt gca gca tca 335
 Asp Asp Leu Ile Asp Ser Leu Thr Ser Ser His Lys Val Ala Ala Ser
 100 105 110

gac gag aag cgt atc gag aca tta ata tcc gag att aag aat atg ttt 383
 Asp Glu Lys Arg Ile Glu Thr Leu Ile Ser Glu Ile Lys Asn Met Phe
 115 120 125

aga tgt atg ggc tat ggc gaa acg aat ccc tct gca tat gac act gct 431
 Arg Cys Met Gly Tyr Gly Glu Thr Asn Pro Ser Ala Tyr Asp Thr Ala
 130 135 140

tgg gta gca agg att cca gca gtt gat ggc tct gac aac cct cac ttt 479
 Trp Val Ala Arg Ile Pro Ala Val Asp Gly Ser Asp Asn Pro His Phe
 145 150 155

cct gag acg gtt gaa tgg att ctt caa aat cag ttg aaa gat ggg tct 527
 Pro Glu Thr Val Glu Trp Ile Leu Gln Asn Gln Leu Lys Asp Gly Ser

-continued

160	165	170	175	
tgg ggt gaa gga ttc tac ttc ttg gca tat gac aga ata ctg gct aca Trp Gly Glu Gly Phe Tyr Phe Leu Ala Tyr Asp Arg Ile Leu Ala Thr 180 185 190				575
ctt gca tgt att att acc ctt acc ctc tgg cgt act ggg gag aca caa Leu Ala Cys Ile Ile Thr Leu Thr Leu Trp Arg Thr Gly Glu Thr Gln 195 200 205				623
gta cag aaa ggt att gaa ttc ttc agg aca caa gct gga aag atg gaa Val Gln Lys Gly Ile Glu Phe Phe Arg Thr Gln Ala Gly Lys Met Glu 210 215 220				671
gat gaa gct gat agt cat agg cca agt gga ttt gaa ata gta ttt cct Asp Glu Ala Asp Ser His Arg Pro Ser Gly Phe Glu Ile Val Phe Pro 225 230 235				719
gca atg cta aag gaa gct aaa atc tta ggc ttg gat ctg cct tac gat Ala Met Leu Lys Glu Ala Lys Ile Leu Gly Leu Asp Leu Pro Tyr Asp 240 245 250 255				767
ttg cca ttc ctg aaa caa atc atc gaa aag cgg gag gct aag ctt aaa Leu Pro Phe Leu Lys Gln Ile Ile Glu Lys Arg Glu Ala Lys Leu Lys 260 265 270				815
agg att ccc act gat gtt ctc tat gcc ctt cca aca acg tta ttg tat Arg Ile Pro Thr Asp Val Leu Tyr Ala Leu Pro Thr Thr Leu Leu Tyr 275 280 285				863
tct ttg gaa ggt tta caa gaa ata gta gac tgg cag aaa ata atg aaa Ser Leu Glu Gly Leu Gln Glu Ile Val Asp Trp Gln Lys Ile Met Lys 290 295 300				911
ctt caa tcc aag gat gga tca ttt ctc agc tct ccg gca tct aca gcg Leu Gln Ser Lys Asp Gly Ser Phe Leu Ser Ser Pro Ala Ser Thr Ala 305 310 315				959
gct gta ttc atg cgt aca ggg aac aaa aag tgc ttg gat ttc ttg aac Ala Val Phe Met Arg Thr Gly Asn Lys Lys Cys Leu Asp Phe Leu Asn 320 325 330 335				1007
ttt gtc ttg aag aaa ttc gga aac cat gtg cct tgt cac tat ccg ctt Phe Val Leu Lys Lys Phe Gly Asn His Val Pro Cys His Tyr Pro Leu 340 345 350				1055
gat cta ttt gaa cgt ttg tgg gcg gtt gat aca gtt gag ccg cta ggt Asp Leu Phe Glu Arg Leu Trp Ala Val Asp Thr Val Glu Arg Leu Gly 355 360 365				1103
atc gat cgt cat ttc aaa gag gag atc aag gaa gca ttg gat tat gtt Ile Asp Arg His Phe Lys Glu Glu Ile Lys Glu Ala Leu Asp Tyr Val 370 375 380				1151
tac agc cat tgg gac gaa aga ggc att gga tgg gcg aga gag aat cct Tyr Ser His Trp Asp Glu Arg Gly Ile Gly Trp Ala Arg Glu Asn Pro 385 390 395				1199
gtt cct gat att gat gat aca gcc atg ggc ctt cga atc ttg aga tta Val Pro Asp Ile Asp Asp Thr Ala Met Gly Leu Arg Ile Leu Arg Leu 400 405 410 415				1247
cat gga tac aat gta tcc tca gat gtt tta aaa aca ttt aga gat gag His Gly Tyr Asn Val Ser Ser Asp Val Leu Lys Thr Phe Arg Asp Glu 420 425 430				1295
aat ggg gag ttc ttt tgc ttc ttg ggt caa aca cag aga gga gtt aca Asn Gly Glu Phe Phe Cys Phe Leu Gly Gln Thr Gln Arg Gly Val Thr 435 440 445				1343
gac atg tta aac gtc aat cgt tgt tca cat gtt tca ttt ccg gga gaa Asp Met Leu Asn Val Asn Arg Cys Ser His Val Ser Phe Pro Gly Glu 450 455 460				1391
acg atc atg gaa gaa gca aaa ctc tgt acc gaa agg tat ctg agg aat Thr Ile Met Glu Glu Ala Lys Leu Cys Thr Glu Arg Tyr Leu Arg Asn 465 470 475				1439
gct ctg gaa aat gtg gat gcc ttt gac aaa tgg gct ttt aaa aag aat				1487

-continued

Ala	Leu	Glu	Asn	Val	Asp	Ala	Phe	Asp	Lys	Trp	Ala	Phe	Lys	Lys	Asn			
480					485					490					495			
att	cgg	gga	gag	gta	gag	tat	gca	ctc	aaa	tat	ccc	tgg	cat	aag	agt		1535	
Ile	Arg	Gly	Glu	Val	Glu	Tyr	Ala	Leu	Lys	Tyr	Pro	Trp	His	Lys	Ser			
				500					505					510				
atg	cca	agg	ttg	gag	gct	aga	agc	tat	att	gaa	aac	tat	ggg	cca	gat		1583	
Met	Pro	Arg	Leu	Glu	Ala	Arg	Ser	Tyr	Ile	Glu	Asn	Tyr	Gly	Pro	Asp			
				515				520					525					
gat	gtg	tgg	ctt	gga	aaa	act	gta	tat	atg	atg	cca	tac	att	tcg	aat		1631	
Asp	Val	Trp	Leu	Gly	Lys	Thr	Val	Tyr	Met	Met	Pro	Tyr	Ile	Ser	Asn			
				530				535					540					
gaa	aag	tat	tta	gaa	cta	gcg	aaa	ctg	gac	ttc	aat	aag	gtg	cag	tct		1679	
Glu	Lys	Tyr	Leu	Glu	Leu	Ala	Lys	Leu	Asp	Phe	Asn	Lys	Val	Gln	Ser			
	545					550					555							
ata	cac	caa	aca	gag	ctt	caa	gat	ctt	cga	agg	tgg	tgg	aaa	tca	tcc		1727	
Ile	His	Gln	Thr	Glu	Leu	Gln	Asp	Leu	Arg	Arg	Trp	Trp	Lys	Ser	Ser			
					565					570					575			
ggt	ttc	acg	gat	ctg	aat	ttc	act	cgt	gag	cgt	gtg	acg	gaa	ata	tat		1775	
Gly	Phe	Thr	Asp	Leu	Asn	Phe	Thr	Arg	Glu	Arg	Val	Thr	Glu	Ile	Tyr			
				580					585					590				
ttc	tca	ccg	gca	tcc	ttt	atc	ttt	gag	ccc	gag	ttt	tct	aag	tgc	aga		1823	
Phe	Ser	Pro	Ala	Ser	Phe	Ile	Phe	Glu	Pro	Glu	Phe	Ser	Lys	Cys	Arg			
				595				600					605					
gag	gtt	tat	aca	aaa	act	tcc	aat	ttc	act	gtt	att	tta	gat	gat	ctt		1871	
Glu	Val	Tyr	Thr	Lys	Thr	Ser	Asn	Phe	Thr	Val	Ile	Leu	Asp	Asp	Leu			
				610				615					620					
tat	gac	gcc	cat	gga	tct	tta	gac	gat	ctt	aag	ttg	ttc	aca	gaa	tca		1919	
Tyr	Asp	Ala	His	Gly	Ser	Leu	Asp	Asp	Leu	Lys	Leu	Phe	Thr	Glu	Ser			
				625				630				635						
gtc	aaa	aga	tgg	gat	cta	tca	cta	gtg	gac	caa	atg	cca	caa	caa	atg		1967	
Val	Lys	Arg	Trp	Asp	Leu	Ser	Leu	Val	Asp	Gln	Met	Pro	Gln	Gln	Met			
					645					650					655			
aaa	ata	tgt	ttt	gtg	ggg	ttc	tac	aat	act	ttt	aat	gat	ata	gca	aaa		2015	
Lys	Ile	Cys	Phe	Val	Gly	Phe	Tyr	Asn	Thr	Phe	Asn	Asp	Ile	Ala	Lys			
				660					665					670				
gaa	gga	cgt	gag	agg	caa	ggg	cgc	gat	gtg	cta	ggc	tac	att	caa	aat		2063	
Glu	Gly	Arg	Glu	Arg	Gln	Gly	Arg	Asp	Val	Leu	Gly	Tyr	Ile	Gln	Asn			
				675				680					685					
gtt	tgg	aaa	gtc	caa	ctt	gaa	gct	tac	acg	aaa	gaa	gca	gaa	tgg	tct		2111	
Val	Trp	Lys	Val	Gln	Leu	Glu	Ala	Tyr	Thr	Lys	Glu	Ala	Glu	Trp	Ser			
				690				695					700					
gaa	gct	aaa	tat	gtg	cca	tcc	ttc	aat	gaa	tac	ata	gag	aat	gcg	agt		2159	
Glu	Ala	Lys	Tyr	Val	Pro	Ser	Phe	Asn	Glu	Tyr	Ile	Glu	Asn	Ala	Ser			
				705				710				715						
gtg	tca	ata	gca	ttg	gga	aca	gtc	ggt	ctc	att	agt	gct	ctt	ttc	act		2207	
Val	Ser	Ile	Ala	Leu	Gly	Thr	Val	Val	Leu	Ile	Ser	Ala	Leu	Phe	Thr			
				720						730					735			
ggg	gag	ggt	ctt	aca	gat	gaa	gta	ctc	tcc	aaa	att	gat	cgc	gaa	tct		2255	
Gly	Glu	Val	Leu	Thr	Asp	Glu	Val	Leu	Ser	Lys	Ile	Asp	Arg	Glu	Ser			
				740					745					750				
aga	ttt	ctt	caa	ctc	atg	ggc	tta	aca	ggg	cgt	ttg	gtg	aat	gac	acc		2303	
Arg	Phe	Leu	Gln	Leu	Met	Gly	Leu	Thr	Gly	Arg	Leu	Val	Asn	Asp	Thr			
				755				760						765				
aaa	act	tat	cag	gca	gag	aga	ggg	caa	ggg	gag	gtg	gct	tct	gcc	ata		2351	
Lys	Thr	Tyr	Gln	Ala	Glu	Arg	Gly	Gln	Gly	Glu	Val	Ala	Ser	Ala	Ile			
				770				775					780					
caa	tgt	tat	atg	aag	gac	cat	cct	aaa	atc	tct	gaa	gaa	gaa	gct	cta		2399	
Gln	Cys	Tyr	Met	Lys	Asp	His	Pro	Lys	Ile	Ser	Glu	Glu	Glu	Ala	Leu			
				785				790						795				

-continued

caa cat gtc tat agt gtc atg gaa aat gcc ctc gaa gag ttg aat agg 2447
 Gln His Val Tyr Ser Val Met Glu Asn Ala Leu Glu Glu Leu Asn Arg
 800 805 810 815

gag ttt gtg aat aac aaa ata ccg gat att tac aaa aga ctg gtt ttt 2495
 Glu Phe Val Asn Asn Lys Ile Pro Asp Ile Tyr Lys Arg Leu Val Phe
 820 825 830

gaa act gca aga ata atg caa ctc ttt tat atg caa ggg gat ggt ttg 2543
 Glu Thr Ala Arg Ile Met Gln Leu Phe Tyr Met Gln Gly Asp Gly Leu
 835 840 845

aca cta tca cat gat atg gaa att aaa gag cat gtc aaa aat tgc ctc 2591
 Thr Leu Ser His Asp Met Glu Ile Lys Glu His Val Lys Asn Cys Leu
 850 855 860

ttc caa cca gtt gcc tag attaaattat tcagttaaag gccctcatgg 2639
 Phe Gln Pro Val Ala
 865

tattgtgtta acattataat aacagatgct caaaagcttt gagcgggtatt tgtaaaggct 2699

atctttgttt gttgtttgt ttactgccaa ccaaaaagcg ttctaaacc ttggaagaca 2759

tttccatcca agagatggag tctacathtt atttatgaga ttgaattatt tcaagagaat 2819

atactacata tatttaaaag taaaaaaaa aaaaaaaaa aa 2861

<210> SEQ ID NO 56
 <211> LENGTH: 868
 <212> TYPE: PRT
 <213> ORGANISM: Abies grandis

<400> SEQUENCE: 56

Met Ala Met Pro Ser Ser Ser Leu Ser Ser Gln Ile Pro Thr Ala Ala
 1 5 10 15

His His Leu Thr Ala Asn Ala Gln Ser Ile Pro His Phe Ser Thr Thr
 20 25 30

Leu Asn Ala Gly Ser Ser Ala Ser Lys Arg Arg Ser Leu Tyr Leu Arg
 35 40 45

Trp Gly Lys Gly Ser Asn Lys Ile Ile Ala Cys Val Gly Glu Gly Gly
 50 55 60

Ala Thr Ser Val Pro Tyr Gln Ser Ala Glu Lys Asn Asp Ser Leu Ser
 65 70 75 80

Ser Ser Thr Leu Val Lys Arg Glu Phe Pro Pro Gly Phe Trp Lys Asp
 85 90 95

Asp Leu Ile Asp Ser Leu Thr Ser Ser His Lys Val Ala Ala Ser Asp
 100 105 110

Glu Lys Arg Ile Glu Thr Leu Ile Ser Glu Ile Lys Asn Met Phe Arg
 115 120 125

Cys Met Gly Tyr Gly Glu Thr Asn Pro Ser Ala Tyr Asp Thr Ala Trp
 130 135 140

Val Ala Arg Ile Pro Ala Val Asp Gly Ser Asp Asn Pro His Phe Pro
 145 150 155 160

Glu Thr Val Glu Trp Ile Leu Gln Asn Gln Leu Lys Asp Gly Ser Trp
 165 170 175

Gly Glu Gly Phe Tyr Phe Leu Ala Tyr Asp Arg Ile Leu Ala Thr Leu
 180 185 190

Ala Cys Ile Ile Thr Leu Thr Leu Trp Arg Thr Gly Glu Thr Gln Val
 195 200 205

Gln Lys Gly Ile Glu Phe Phe Arg Thr Gln Ala Gly Lys Met Glu Asp
 210 215 220

Glu Ala Asp Ser His Arg Pro Ser Gly Phe Glu Ile Val Phe Pro Ala

-continued

225	230	235	240
Met Leu Lys Glu Ala Lys Ile Leu Gly Leu Asp Leu Pro Tyr Asp Leu 245 250 255			
Pro Phe Leu Lys Gln Ile Ile Glu Lys Arg Glu Ala Lys Leu Lys Arg 260 265 270			
Ile Pro Thr Asp Val Leu Tyr Ala Leu Pro Thr Thr Leu Leu Tyr Ser 275 280 285			
Leu Glu Gly Leu Gln Glu Ile Val Asp Trp Gln Lys Ile Met Lys Leu 290 295 300			
Gln Ser Lys Asp Gly Ser Phe Leu Ser Ser Pro Ala Ser Thr Ala Ala 305 310 315 320			
Val Phe Met Arg Thr Gly Asn Lys Lys Cys Leu Asp Phe Leu Asn Phe 325 330 335			
Val Leu Lys Lys Phe Gly Asn His Val Pro Cys His Tyr Pro Leu Asp 340 345 350			
Leu Phe Glu Arg Leu Trp Ala Val Asp Thr Val Glu Arg Leu Gly Ile 355 360 365			
Asp Arg His Phe Lys Glu Glu Ile Lys Glu Ala Leu Asp Tyr Val Tyr 370 375 380			
Ser His Trp Asp Glu Arg Gly Ile Gly Trp Ala Arg Glu Asn Pro Val 385 390 395 400			
Pro Asp Ile Asp Asp Thr Ala Met Gly Leu Arg Ile Leu Arg Leu His 405 410 415			
Gly Tyr Asn Val Ser Ser Asp Val Leu Lys Thr Phe Arg Asp Glu Asn 420 425 430			
Gly Glu Phe Phe Cys Phe Leu Gly Gln Thr Gln Arg Gly Val Thr Asp 435 440 445			
Met Leu Asn Val Asn Arg Cys Ser His Val Ser Phe Pro Gly Glu Thr 450 455 460			
Ile Met Glu Glu Ala Lys Leu Cys Thr Glu Arg Tyr Leu Arg Asn Ala 465 470 475 480			
Leu Glu Asn Val Asp Ala Phe Asp Lys Trp Ala Phe Lys Lys Asn Ile 485 490 495			
Arg Gly Glu Val Glu Tyr Ala Leu Lys Tyr Pro Trp His Lys Ser Met 500 505 510			
Pro Arg Leu Glu Ala Arg Ser Tyr Ile Glu Asn Tyr Gly Pro Asp Asp 515 520 525			
Val Trp Leu Gly Lys Thr Val Tyr Met Met Pro Tyr Ile Ser Asn Glu 530 535 540			
Lys Tyr Leu Glu Leu Ala Lys Leu Asp Phe Asn Lys Val Gln Ser Ile 545 550 555 560			
His Gln Thr Glu Leu Gln Asp Leu Arg Arg Trp Trp Lys Ser Ser Gly 565 570 575			
Phe Thr Asp Leu Asn Phe Thr Arg Glu Arg Val Thr Glu Ile Tyr Phe 580 585 590			
Ser Pro Ala Ser Phe Ile Phe Glu Pro Glu Phe Ser Lys Cys Arg Glu 595 600 605			
Val Tyr Thr Lys Thr Ser Asn Phe Thr Val Ile Leu Asp Asp Leu Tyr 610 615 620			
Asp Ala His Gly Ser Leu Asp Asp Leu Lys Leu Phe Thr Glu Ser Val 625 630 635 640			
Lys Arg Trp Asp Leu Ser Leu Val Asp Gln Met Pro Gln Gln Met Lys 645 650 655			

-continued

Ile Cys Phe Val Gly Phe Tyr Asn Thr Phe Asn Asp Ile Ala Lys Glu
660 665 670

Gly Arg Glu Arg Gln Gly Arg Asp Val Leu Gly Tyr Ile Gln Asn Val
675 680 685

Trp Lys Val Gln Leu Glu Ala Tyr Thr Lys Glu Ala Glu Trp Ser Glu
690 695 700

Ala Lys Tyr Val Pro Ser Phe Asn Glu Tyr Ile Glu Asn Ala Ser Val
705 710 715 720

Ser Ile Ala Leu Gly Thr Val Val Leu Ile Ser Ala Leu Phe Thr Gly
725 730 735

Glu Val Leu Thr Asp Glu Val Leu Ser Lys Ile Asp Arg Glu Ser Arg
740 745 750

Phe Leu Gln Leu Met Gly Leu Thr Gly Arg Leu Val Asn Asp Thr Lys
755 760 765

Thr Tyr Gln Ala Glu Arg Gly Gln Gly Glu Val Ala Ser Ala Ile Gln
770 775 780

Cys Tyr Met Lys Asp His Pro Lys Ile Ser Glu Glu Ala Leu Gln
785 790 795 800

His Val Tyr Ser Val Met Glu Asn Ala Leu Glu Glu Leu Asn Arg Glu
805 810 815

Phe Val Asn Asn Lys Ile Pro Asp Ile Tyr Lys Arg Leu Val Phe Glu
820 825 830

Thr Ala Arg Ile Met Gln Leu Phe Tyr Met Gln Gly Asp Gly Leu Thr
835 840 845

Leu Ser His Asp Met Glu Ile Lys Glu His Val Lys Asn Cys Leu Phe
850 855 860

Gln Pro Val Ala
865

<210> SEQ ID NO 57

<211> LENGTH: 2089

<212> TYPE: DNA

<213> ORGANISM: Abies grandis

<220> FEATURE:

<221> NAME/KEY: CDS

<222> LOCATION: (73)...(1983)

<223> OTHER INFORMATION: (-)-4S-limonene synthase

<400> SEQUENCE: 57

tgccgttttaa tcggttttaaa gaagctacca tagttcgggt taaagaagct accatagttt 60

aggcaggaat cc atg gct ctc ctt tct atc gta tct ttg cag gtt ccc aaa 111
Met Ala Leu Leu Ser Ile Val Ser Leu Gln Val Pro Lys
1 5 10

tcc tgc ggg ctg aaa tcg ttg atc agt tcc agc aat gtg cag aag gct 159
Ser Cys Gly Leu Lys Ser Leu Ile Ser Ser Ser Asn Val Gln Lys Ala
15 20 25

ctc tgt atc tct aca gca gtc cca aca ctc aga atg cgt agg cga cag 207
Leu Cys Ile Ser Thr Ala Val Pro Thr Leu Arg Met Arg Arg Arg Gln
30 35 40 45

aaa gct ctg gtc atc aac atg aaa ttg acc act gta tcc cat cgt gat 255
Lys Ala Leu Val Ile Asn Met Lys Leu Thr Thr Val Ser His Arg Asp
50 55 60

gat aat ggt ggt ggt gta ctg caa aga cgc ata gcc gat cat cat ccc 303
Asp Asn Gly Gly Gly Val Leu Gln Arg Arg Ile Ala Asp His His Pro
65 70 75

aac ctg tgg gaa gat gat ttc ata caa tca ttg tcc tca cct tat ggg 351
Asn Leu Trp Glu Asp Asp Phe Ile Gln Ser Leu Ser Ser Pro Tyr Gly

-continued

Asn	Glu	Leu	Gln	Leu	Phe	Thr	Asp	Ala	Ile	Lys	Arg	Trp	Asp	Leu	Ser		
	400						405					410					
acg	aca	agg	tgg	ctt	cca	gaa	tat	atg	aaa	gga	gtg	tac	atg	gac	ttg	1359	
Thr	Thr	Arg	Trp	Leu	Pro	Glu	Tyr	Met	Lys	Gly	Val	Tyr	Met	Asp	Leu		
	415				420					425							
tat	caa	tgc	att	aat	gaa	atg	gtg	gaa	gag	gct	gag	aag	act	caa	ggc	1407	
Tyr	Gln	Cys	Ile	Asn	Glu	Met	Val	Glu	Glu	Ala	Glu	Lys	Thr	Gln	Gly		
	430				435				440					445			
cga	gat	atg	ctc	aac	tat	att	caa	aat	gct	tgg	gaa	gcc	cta	ttt	gat	1455	
Arg	Asp	Met	Leu	Asn	Tyr	Ile	Gln	Asn	Ala	Trp	Glu	Ala	Leu	Phe	Asp		
			450						455					460			
acc	ttt	atg	caa	gaa	gca	aag	tgg	atc	tcc	agc	agt	tat	ctc	cca	acg	1503	
Thr	Phe	Met	Gln	Glu	Ala	Lys	Trp	Ile	Ser	Ser	Ser	Tyr	Leu	Pro	Thr		
			465					470					475				
ttt	gag	gag	tac	ttg	aag	aat	gca	aaa	gtt	agt	tct	ggt	tct	cgc	ata	1551	
Phe	Glu	Glu	Tyr	Leu	Lys	Asn	Ala	Lys	Val	Ser	Ser	Gly	Ser	Arg	Ile		
	480					485						490					
gcc	aca	tta	caa	ccc	att	ctc	act	ttg	gat	gta	cca	ctt	cct	gat	tac	1599	
Ala	Thr	Leu	Gln	Pro	Ile	Leu	Thr	Leu	Asp	Val	Pro	Leu	Pro	Asp	Tyr		
	495				500					505							
ata	ctg	caa	gaa	att	gat	tat	cca	tcc	aga	ttc	aat	gag	tta	gct	tcg	1647	
Ile	Leu	Gln	Glu	Ile	Asp	Tyr	Pro	Ser	Arg	Phe	Asn	Glu	Leu	Ala	Ser		
	510				515				520					525			
tcc	atc	ctt	cga	cta	cga	ggt	gac	acg	cgc	tgc	tac	aag	gcg	gat	agg	1695	
Ser	Ile	Leu	Arg	Leu	Arg	Gly	Asp	Thr	Arg	Cys	Tyr	Lys	Ala	Asp	Arg		
			530						535					540			
gcc	cgt	gga	gaa	gaa	gct	tca	gct	ata	tcg	tgt	tat	atg	aaa	gac	cat	1743	
Ala	Arg	Gly	Glu	Glu	Ala	Ser	Ala	Ile	Ser	Cys	Tyr	Met	Lys	Asp	His		
		545						550						555			
cct	gga	tca	ata	gag	gaa	gat	gct	ctc	aat	cat	atc	aac	gcc	atg	atc	1791	
Pro	Gly	Ser	Ile	Glu	Glu	Asp	Ala	Leu	Asn	His	Ile	Asn	Ala	Met	Ile		
		560						565						570			
agt	gat	gca	atc	aga	gaa	tta	aat	tgg	gag	ctt	ctc	aga	ccg	gat	agc	1839	
Ser	Asp	Ala	Ile	Arg	Glu	Leu	Asn	Trp	Glu	Leu	Leu	Arg	Pro	Asp	Ser		
		575				580						585					
aaa	agt	ccc	atc	tct	tcc	aag	aaa	cat	gct	ttt	gac	atc	acc	aga	gct	1887	
Lys	Ser	Pro	Ile	Ser	Ser	Lys	Lys	His	Ala	Phe	Asp	Ile	Thr	Arg	Ala		
		590				595			600					605			
ttc	cat	cat	gtc	tac	aaa	tat	cga	gat	ggt	tac	act	ggt	tcc	aac	aac	1935	
Phe	His	His	Val	Tyr	Lys	Tyr	Arg	Asp	Gly	Tyr	Thr	Val	Ser	Asn	Asn		
			610						615					620			
gaa	aca	aag	aat	ttg	gtg	atg	aaa	acc	gtt	ctt	gaa	cct	ctc	gct	ttg	1983	
Glu	Thr	Lys	Asn	Leu	Val	Met	Lys	Thr	Val	Leu	Glu	Pro	Leu	Ala	Leu		
			625					630						635			
taa	aaacatatag	aatgcattaa	aatgtgggaa	gtctataatc	tagactattc											2036	
tctatctttc	ataatgtaga	tctggatgtg	tattgaaactc	taaaaaaaa	aaa											2089	

<210> SEQ ID NO 58

<211> LENGTH: 637

<212> TYPE: PRT

<213> ORGANISM: Abies grandis

<400> SEQUENCE: 58

Met Ala Leu Leu Ser Ile Val Ser Leu Gln Val Pro Lys Ser Cys Gly
1 5 10 15Leu Lys Ser Leu Ile Ser Ser Ser Asn Val Gln Lys Ala Leu Cys Ile
20 25 30Ser Thr Ala Val Pro Thr Leu Arg Met Arg Arg Arg Gln Lys Ala Leu
35 40 45

-continued

Val Ile Asn Met Lys Leu Thr Thr Val Ser His Arg Asp Asp Asn Gly
 50 55 60

Gly Gly Val Leu Gln Arg Arg Ile Ala Asp His His Pro Asn Leu Trp
 65 70 75 80

Glu Asp Asp Phe Ile Gln Ser Leu Ser Ser Pro Tyr Gly Gly Ser Ser
 85 90 95

Tyr Ser Glu Arg Ala Glu Thr Val Val Glu Glu Val Lys Glu Met Phe
 100 105 110

Asn Ser Ile Pro Asn Asn Arg Glu Leu Phe Gly Ser Gln Asn Asp Leu
 115 120 125

Leu Thr Arg Leu Trp Met Val Asp Ser Ile Glu Arg Leu Gly Ile Asp
 130 135 140

Arg His Phe Gln Asn Glu Ile Arg Val Ala Leu Asp Tyr Val Tyr Ser
 145 150 155 160

Tyr Trp Lys Glu Lys Glu Gly Ile Gly Cys Gly Arg Asp Ser Thr Phe
 165 170 175

Pro Asp Leu Asn Ser Thr Ala Leu Ala Leu Arg Thr Leu Arg Leu His
 180 185 190

Gly Tyr Asn Val Ser Ser Asp Val Leu Glu Tyr Phe Lys Asp Glu Lys
 195 200 205

Gly His Phe Ala Cys Pro Ala Ile Leu Thr Glu Gly Gln Ile Thr Arg
 210 215 220

Ser Val Leu Asn Leu Tyr Arg Ala Ser Leu Val Ala Phe Pro Gly Glu
 225 230 235 240

Lys Val Met Glu Glu Ala Glu Ile Phe Ser Ala Ser Tyr Leu Lys Lys
 245 250 255

Val Leu Gln Lys Ile Pro Val Ser Asn Leu Ser Gly Glu Ile Glu Tyr
 260 265 270

Val Leu Glu Tyr Gly Trp His Thr Asn Leu Pro Arg Leu Glu Ala Arg
 275 280 285

Asn Tyr Ile Glu Val Tyr Glu Gln Ser Gly Tyr Glu Ser Leu Asn Glu
 290 295 300

Met Pro Tyr Met Asn Met Lys Lys Leu Leu Gln Leu Ala Lys Leu Glu
 305 310 315 320

Phe Asn Ile Phe His Ser Leu Gln Leu Arg Glu Leu Gln Ser Ile Ser
 325 330 335

Arg Trp Trp Lys Glu Ser Gly Ser Ser Gln Leu Thr Phe Thr Arg His
 340 345 350

Arg His Val Glu Tyr Tyr Thr Met Ala Ser Cys Ile Ser Met Leu Pro
 355 360 365

Lys His Ser Ala Phe Arg Met Glu Phe Val Lys Val Cys His Leu Val
 370 375 380

Thr Val Leu Asp Asp Ile Tyr Asp Thr Phe Gly Thr Met Asn Glu Leu
 385 390 395 400

Gln Leu Phe Thr Asp Ala Ile Lys Arg Trp Asp Leu Ser Thr Thr Arg
 405 410 415

Trp Leu Pro Glu Tyr Met Lys Gly Val Tyr Met Asp Leu Tyr Gln Cys
 420 425 430

Ile Asn Glu Met Val Glu Glu Ala Glu Lys Thr Gln Gly Arg Asp Met
 435 440 445

Leu Asn Tyr Ile Gln Asn Ala Trp Glu Ala Leu Phe Asp Thr Phe Met
 450 455 460

-continued

Gln	Glu	Ala	Lys	Trp	Ile	Ser	Ser	Ser	Tyr	Leu	Pro	Thr	Phe	Glu	Glu
465					470					475				480	
Tyr	Leu	Lys	Asn	Ala	Lys	Val	Ser	Ser	Gly	Ser	Arg	Ile	Ala	Thr	Leu
			485						490					495	
Gln	Pro	Ile	Leu	Thr	Leu	Asp	Val	Pro	Leu	Pro	Asp	Tyr	Ile	Leu	Gln
			500					505					510		
Glu	Ile	Asp	Tyr	Pro	Ser	Arg	Phe	Asn	Glu	Leu	Ala	Ser	Ser	Ile	Leu
		515					520					525			
Arg	Leu	Arg	Gly	Asp	Thr	Arg	Cys	Tyr	Lys	Ala	Asp	Arg	Ala	Arg	Gly
	530					535					540				
Glu	Glu	Ala	Ser	Ala	Ile	Ser	Cys	Tyr	Met	Lys	Asp	His	Pro	Gly	Ser
545					550					555					560
Ile	Glu	Glu	Asp	Ala	Leu	Asn	His	Ile	Asn	Ala	Met	Ile	Ser	Asp	Ala
				565					570					575	
Ile	Arg	Glu	Leu	Asn	Trp	Glu	Leu	Leu	Arg	Pro	Asp	Ser	Lys	Ser	Pro
			580					585					590		
Ile	Ser	Ser	Lys	Lys	His	Ala	Phe	Asp	Ile	Thr	Arg	Ala	Phe	His	His
		595					600					605			
Val	Tyr	Lys	Tyr	Arg	Asp	Gly	Tyr	Thr	Val	Ser	Asn	Asn	Glu	Thr	Lys
	610					615					620				
Asn	Leu	Val	Met	Lys	Thr	Val	Leu	Glu	Pro	Leu	Ala	Leu			
625					630					635					

We claim the following:

1. An isolated terpene synthase having a region with 20% or greater sequence identity to residues 265 to 535 of SEQ ID NO: 2, said synthase comprising nine α -carbons having interatomic distances in Angstroms between said α -carbons that are ± 2.3 Angstroms of the following interatomic distances:

α -Carbon	1	2	3	4	5	6	7	8	9
1	0.0	8.4	13.7	12.7	11.9	10.2	13.1	9.4	12.8
2	8.4	0.0	11.3	8.7	10.2	7.2	14.8	15.1	17.4
3	13.7	11.3	0.0	3.8	5.4	9.3	6.6	13.9	13.7
4	12.7	8.7	3.8	0.0	3.8	6.0	9.2	15.4	16.1
5	11.9	10.2	5.4	3.8	0.0	5.0	7.8	14.6	15.5
6	10.2	7.2	9.3	6.0	5.0	0.0	12.0	16.1	18.0
7	13.1	14.8	6.6	9.2	7.8	12.0	0.0	10.2	9.5
8	9.4	15.1	13.9	15.4	14.6	16.1	10.2	0.0	3.8
9	12.8	17.4	13.7	16.1	15.5	18.0	9.5	3.8	0.0

the center point of each said α -carbon positioned within a sphere having a radius of 2.3 Angstroms, the center points of each said sphere having the following structural coordinates:

α -Carbon Number	X Position	Y Position	Z Position
1	120.203	38.695	43.506
2	114.058	43.884	41.015
3	106.807	36.336	45.151
4	107.629	38.010	41.804
5	109.375	34.842	40.617
6	111.944	37.854	37.602
7	110.233	31.098	47.361

-continued

α -Carbon Number	X Position	Y Position	Z Position
8	118.846	34.443	51.796
9	116.461	32.848	54.290

each said α -carbon having an associated R-group, said synthase having an ordered arrangement of said R-groups other than the following ordered arrangements of R-groups:

TABLE 9

Ordered Arrangements of α -Carbons 1-9									
	1	2	3	4	5	6	7	8	9
A	W	I	T	T	Y	L	C	T	Y
B	W	I	S	T	Y	L	C	T	Y
C	W	I	C	G	Y	L	C	L	Y
D	W	I	S	G	Y	L	C	L	Y
E	W	L	A	G	Y	I	A	L	Y
F	W	L	T	V	H	L	G	V	Y
G	W	L	A	G	Y	I	A	L	Y
H	W	I	V	G	N	L	F	L	Y
I	W	I	T	A	G	L	S	C	Y
J	W	V	S	C	I	M	G	S	Y
K	F	F	I	T	A	T	G	T	Y
L	W	N	I	S	G	M	L	M	Y
M	W	V	S	S	Y	L	G	L	Y
N	F	F	T	L	A	L	G	S	Y
O	W	N	S	G	P	L	L	M	Y
P	W	N	G	G	I	L	L	I	Y
Q	Y	L	V	T	M	T	G	T	Y
R	W	I	I	S	A	I	L	I	Y
S	W	F	S	S	V	I	L	I	Y
T	W	I	V	A	S	I	L	I	Y
U	W	N	I	S	S	I	F	M	Y
V	L	A	I	G	Q	L	S	I	F
W	S	S	I	A	L	V	G	F	Y
X	L	C	C	G	H	S	L	G	Y

TABLE 9-continued

Ordered Arrangements of α -Carbons 1-9									
	1	2	3	4	5	6	7	8	9
Y	S	F	S	S	V	I	L	V	Y
Z	W	A	S	G	M	L	G	I	Y
AA	A	N	L	T	S	T	C	L	Y
BB	L	C	S	A	Y	V	L	L	Y
CC	W	A	T	G	M	L	S	M	Y
DD	M	C	S	S	G	I	L	V	Y
EE	S	G	V	G	L	C	W	F	Y
FF	S	G	A	L	G	V	G	F	Y
GG	S	G	F	A	L	I	G	F	Y
HH	A	G	F	A	L	I	G	F	Y
II	W	V	T	G	L	V	I	S	Y
JJ	W	A	S	G	M	L	G	I	Y
KK	W	I	S	T	Y	L	C	T	Y
LL	W	I	T	T	Y	L	C	T	Y
MM	W	N	I	S	G	M	L	M	Y
NN	A	A	I	G	Q	L	S	I	F
OO	A	I	V	A	S	I	L	I	Y

2. The synthase of claim 1, wherein said synthase has 25% or greater sequence identity to residues 265 to 535 of SEQ ID 2.

3. The synthase of claim 2, wherein said synthase has 35% or greater sequence identity to residues 265 to 535 of SEQ ID 2.

4. The synthase of claim 1, wherein said synthase catalyses the formation of a terpenoid product from a monoterpene substrate.

5. The synthase of claim 1, wherein said synthase catalyses the formation of a terpenoid product from a sesquiterpene substrate.

6. The synthase of claim 1, wherein said synthase catalyses the formation of a terpenoid product from a diterpene substrate.

7. The synthase of claim 4, wherein said product is a cyclic terpenoid hydrocarbon.

8. The synthase of claim 4, wherein said product is an acyclic terpenoid hydrocarbon.

9. The synthase of claim 4, wherein said product is a cyclic hydroxylated terpenoid hydrocarbon.

10. The synthase of claim 4, wherein said product is an acyclic hydroxylated terpenoid hydrocarbon.

11. The synthase of claim 1, wherein said R-group associated with said α -carbon 1 is selected from the group consisting of Cys, Ser, and Thr.

12. The synthase of claim 1, wherein said R-group associated with said α -carbon 1 is selected from the group consisting of Phe, Tyr and Trp.

13. The synthase of claim 1, wherein said R-group associated with said α -carbon 1 is selected from the group consisting of Pro, Gly, and Ala.

14. The synthase of claim 1, wherein said R-group associated with said α -carbon 1 is selected from the group consisting of Glu and Asp.

15. The synthase of claim 1, wherein said R-group associated with said α -carbon 1 is selected from the group consisting of Met, Ile, Val and Leu.

16. The synthase of claim 1, wherein said R-group associated with said α -carbon 1 is selected from the group consisting of Arg and Lys.

17. The synthase of claim 1, wherein said R-group associated with said α -carbon 1 is selected from the group consisting of Gln, Asn and His.

18. The synthase of claim 1, wherein said R-group associated with said α -carbon 2 is selected from the group consisting of Cys, Ser and Thr.

19. The synthase of claim 1, wherein said R-group associated with said α -carbon 2 is selected from the group consisting of Phe, Tyr and Trp.

20. The synthase of claim 1, wherein said R-group associated with said α -carbon 2 is selected from the group consisting of Pro, Gly, and Ala.

21. The synthase of claim 1, wherein said R-group associated with said α -carbon 2 is selected from the group consisting of Glu and Asp.

22. The synthase of claim 1, wherein said R-group associated with said α -carbon 2 is selected from the group consisting of Met, Ile, Val and Leu.

23. The synthase of claim 1, wherein said R-group associated with said α -carbon 2 is selected from the group consisting of Arg and Lys.

24. The synthase of claim 1, wherein said R-group associated with said α -carbon 2 is selected from the group consisting of Gln, Asn and His.

25. The synthase of claim 1, wherein said R-group associated with said α -carbon 3 is selected from the group consisting of Cys, Ser and Thr.

26. The synthase of claim 1, wherein said R-group associated with said α -carbon 3 is selected from the group consisting of Phe, Tyr and Trp.

27. The synthase of claim 1, wherein said R-group associated with said α -carbon 3 is selected from the group consisting of Pro, Gly, and Ala.

28. The synthase of claim 1, wherein said R-group associated with said α -carbon 3 is selected from the group consisting of Glu and Asp.

29. The synthase of claim 1, wherein said R-group associated with said α -carbon 3 is selected from the group consisting of Met, Ile, Val and Leu.

30. The synthase of claim 1, wherein said R-group associated with said α -carbon 3 is selected from the group consisting of Arg and Lys.

31. The synthase of claim 1, wherein said R-group associated with said α -carbon 3 is selected from the group consisting of Gln, Asn and His.

32. The synthase of claim 1, wherein said R-group associated with said α -carbon 4 is selected from the group consisting of Cys, Ser and Thr.

33. The synthase of claim 1, wherein said R-group associated with said α -carbon 4 is selected from the group consisting of Phe, Tyr and Trp.

34. The synthase of claim 1, wherein said R-group associated with said α -carbon 4 is selected from the group consisting of Pro, Gly, and Ala.

35. The synthase of claim 1, wherein said R-group associated with said α -carbon 4 is selected from the group consisting of Glu and Asp.

36. The synthase of claim 1, wherein said R-group associated with said α -carbon 4 is selected from the group consisting of Met, Ile, Val and Leu.

37. The synthase of claim 1, wherein said R-group associated with said α -carbon 4 is selected from the group consisting of Arg and Lys.

38. The synthase of claim 1, wherein said R-group associated with said α -carbon 4 is selected from the group consisting of Gln, Asn and His.

39. The synthase of claim 1, wherein said R-group associated with said α -carbon 5 is selected from the group consisting of Cys, Ser and Thr.

40. The synthase of claim 1, wherein said R-group associated with said α -carbon 5 is selected from the group consisting of Phe, Tyr and Trp.

41. The synthase of claim 1, wherein said R-group associated with said α -carbon 5 is selected from the group consisting of Pro, Gly, and Ala.

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 6,495,354 B2
APPLICATION NO. : 09/887586
DATED : December 17, 2002
INVENTOR(S) : Chappell et al.

Page 1 of 2

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

IN THE CLAIMS: should read

Column 363, line 22 to line 24

2. The synthase of claim 1, wherein said synthase has 25% or greater sequence identity to residues 265 to 535 of SEQ ID 2.

Column 363, line 25 to line 27

3. The synthase of claim 2, wherein said synthase has 35% or greater sequence identity to residues 265 to 535 of SEQ ID 2.

Column 363, line 53 to line 55

14. The synthase of claim 1, wherein said R-group associated with said α -carbon 1 is selected from the group consisting of Glu and Asp.

Column 366, line 33 to line 36

75. The synthase of claim 1, wherein said ordered arrangement of R-groups in said synthase associated with said α -carbons 1 to 9 is Ser, Ile, Thr, Thr, Tyr, Leu, Cys, Thr and Tyr, respectively.

Column 366, line 58 to line 59

83. The synthase of claim 6, wherein said product is a cyclic terpenoid hydrocarbon.

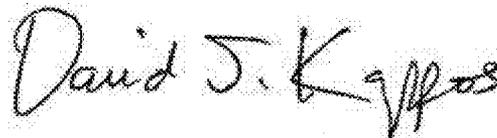
Column 366, line 60 to line 61

84. The synthase of claim 6, wherein said product is an acyclic terpenoid hydrocarbon.

Column 366, line 62 to line 63

85. The synthase of claim 6, wherein said product is a cyclic hydroxylated terpenoid hydrocarbon.

Signed and Sealed this
Fifth Day of April, 2011



David J. Kappos
Director of the United States Patent and Trademark Office

CERTIFICATE OF CORRECTION (continued)

Page 2 of 2

U.S. Pat. No. 6,495,354 B2

Column 366, line 64 to line 65

86. The synthase of claim 6, wherein said product is an acyclic hydroxylated terpenoid hydrocarbon.