(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 January 2013 (31.01.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/016591 A1

(51) International Patent Classification:
C12P 5/02 (2006.01) *C12N 15/00* (2006.01)
C07H 21/00 (2006.01) *C12N 15/09* (2006.01)
C07H 21/04 (2006.01)

(21) International Application Number:
PCT/US2012/048422

(22) International Filing Date:
26 July 2012 (26.07.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/511,954 26 July 2011 (26.07.2011) US

(71) Applicants *(for all designated States except US)*: **ZUVACHEM, INC.** [US/US]; 10751 Falls Road, Suite 300, Lutherville, MD 21093 (US). **BOARD OF TRUSTEES OF MICHIGAN STATE UNIVERSITY** [US/US]; 410 Biochemistry Building, East Lansing, MI 48824 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SHARKEY, Thomas, D.** [US/US]; C/o Michigan State University, 410 Biochemistry Building, East Lansing, MI 48824 (US). **ASPLAND, Simon, E.** [GB/US]; 2507 Fait Avenue, Baltimore, MD 21224 (US).

(74) Agents: **ELLIS, Gregory, C.** et al.; Morgan, Lewis & Bockus, LLP, One Market, Spear Street Tower, San Francisco, CA 94105 (US).

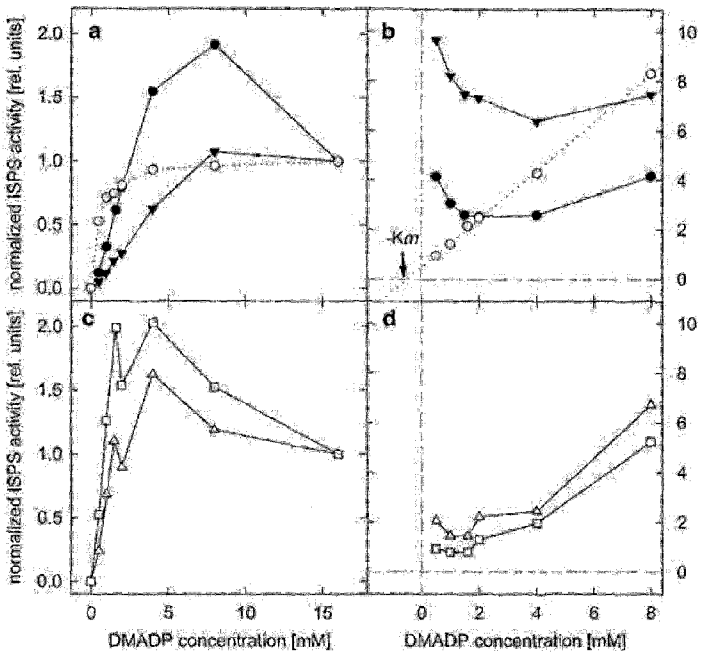
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ,

[Continued on next page]

(54) Title: HIGH EFFICIENCY ISOPRENE SYNTHASES PRODUCED BY PROTEIN ENGINEERING

Figure 1



(57) Abstract: The present invention provides new energy and chemical industry' solutions that are sustainable both environmentally and economically. The invention relates to the use of variant isoprene synthases designed to increase the rate of isoprene production in genetically engineered microorganisms or "biocatalysts" serving as a viable alternative to petroleum-dependent chemicals and energy.

WO 2013/016591 A1

TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

HIGH EFFICIENCY ISOPRENE SYNTHASES PRODUCED BY PROTEIN ENGINEERING

FIELD OF THE INVENTION

[0001] The invention relates, in part, to high efficiency enzymes needed for the economic viability of enzyme driven processes for the production of renewable isoprene.

BACKGROUND OF INVENTION

[0002] Isoprene is a versatile feedstock utilized in production of synthetic rubber (poly-isoprene), elastomers, lubricants, and specialty chemicals. The chemical value of isoprene, CH_2=C(CH_3)CH=CH_2, is provided by two double bonds and a branched methyl group, which enable it to react to form a wide range of valuable products, including poly-isoprene. Furthermore, isoprene has a boiling point of 34°C, a property that offers a number of production and purification advantages. Isoprene is currently produced by the petrochemical industry as a by-product of the thermal cracking of crude oil. The yield of isoprene from this process is small and requires the collection of 5-carbon molecule (C5) streams from several refineries and the separation of individual products from these streams, including isoprene. Approximately 800,000 metric tons of isoprene is produced by this route every year. Efficient polymerization of this product into poly-isoprene is handicapped by impurities that are difficult to separate from the C5 stream of thermally cracked oil.

[0003] The global market for poly-isoprene (rubber) is greater than $25 billion per year. Global demand for rubber (natural and synthetic) by the transportation industry is growing by more than 5% (~300,000 metric tons) per year. Demand is highest in China, which has overtaken the United States in car manufacturing. The global supply of natural rubber is constrained by the productivity of rubber trees, competing uses for irrigated acreage, and the spread of a fungal pathogen which devastated plantations in rubber's native South America. The supply of synthetic isoprene to make synthetic poly-isoprene (also known as Isoprene Rubber) is constrained by refinery capacity and the price of oil. As a result, prices for natural rubber and synthetic isoprene have risen

steadily, doubling over the past 5 years to a present level of approximately $3 per kilogram, and are anticipated to increase at the same rate over the next decade.

[0004] In addition to the poly-isoprene content of rubber, many plants, primarily trees, produce and emit large quantities of monomeric isoprene, over 500,000 metric tons per year. Isoprene is the simplest member of a large (~40,000 members) family of plant isoprenoids including natural rubber. A number of reasons have been suggested for why plants produce isoprene, including its use as: a flowering signal (Terry, et al., (1995) *J. Exp. Bot.* **46**: 1629-1631), a way of releasing phosphate inadvertently converted to dimethyl allyl diphosphate (safety valve) (Rosenstiel, et al., (2004) *Plant Biol.* **6**: 12-21) and a role in defense against being eaten by insects, mediated through the attraction of natural enemies (Laothawornkitkul, et al., (2008) *Plant Cell & Environment.* **31**: 1410-1415). However, the most likely benefits to plants of producing and emitting isoprene are protection from oxidative damage by quenching reactive oxygen species (ROS) and thermo-protection against transient changes in temperature in the leaf canopy by strengthening hydrophobic interactions within membranes.

[0005] Plants produce isoprene from dimethylallyl pyrophosphate, also known as dimethylallyl diphosphate (DMADP), using the enzyme isoprene synthase. To date, confirmed isoprene synthases have been cloned and expressed from several species of Populus and from kudzu: (Miller, et al., (2001) *Planta* **213**: 483-487; Sasaki, et al., (2005) *FEBS Letters* **579**: 2514-2518; Sharkey, et al., (2005) *Plant Physiology* **137**: 700-712; Behnke, et al., (2007) *The Plant Journal* **51**: 485-499; Laothawornkitkul, et al., (2008) *Plant Cell & Environment.* **31**: 1410-1415). Expression of these isoprene synthase genes in microbial species that produce DMADP to make isoprenoids other than isoprene, but that lack an endogenous isoprene synthase, could provide a novel approach to isoprene production.

[0006] Combining the addition of isoprene synthase and genes to increase the production of DMADP and the removal or reduction of endogenous microbial genes that are not vital but divert carbon away from the production of isoprene, results in the production of microorganisms genetically engineered to produce high levels of isoprene, (Whited et al., (2010) *Industrial Biotechnology* **6**: 152-163).

[0007] A key factor determining economic feasibility of producing isoprene by enzyme driven routes, fermentative or otherwise biochemical, is the ability of the isoprene synthase used to efficiently produce a large quantity of isoprene. More specifically, the enzyme synthesizing isoprene preferably produces isoprene at a high rate, e.g., in a fermentative setting greater than 2.5g of isoprene per litre per hour of microbial culture, and, even more preferably, will be able to produce many times its own weight in isoprene. These goals can be realized by engineering an isoprene synthase with appropriate K_m and k_{cat}.

[0008] Currently, all native isoprene synthases that have been studied have a high K_m so require high concentrations of their substrate DMADP in order to reach maximal rates of isoprene production. For the plants that naturally express isoprene synthases, it may be evolutionarily advantageous for their isoprene synthases to have high K_ms to ensure that carbon is first directed to essential functions for the plant and only secondarily to isoprene production through the activity of isoprene synthases. Not wishing to be bound by theory, it could be hypothesized that a very low K_m isoprene synthase may be disadvantageous to plants and therefore less likely to occur in nature. In addition to their high K_m, known isoprene synthases have a very low turnover constant, k_{cat}, or overall catalytic rate. Furthermore, plant isoprene synthases did not evolve to function in an extracellular or non-plant cellular environment such as that found inside microbes grown in a fermenter. Consequently, their performance in such settings will be sub-optimal. Therefore, what is needed to meet the increasing worldwide demand for isoprene and products derived from isoprene is an isoprene synthase with optimized K_m and K_{cat} values, which efficiently synthesizes isoprene in a fermentative mileau. Further, methods for engineering one or a combination of K_m and k_{cat} values of an isoprene synthase would represent a significant advancement over conventional methods of the art.

[0009] This invention provides isoprene synthases having properties necessary for efficient fermentation of feedstocks to produce isoprene, and protein engineering approaches for producing enzymes with an enhanced ability, relative to wild type isoprene synthases, to catalyse the formation of isoprene. Broadly speaking, methods include modifying existing isoprene producing enzymes or converting related enzymes

that do not naturally produce isoprene into enzymes capable of catalyzing isoprene formation. In various embodiments, such enzymes are sufficiently efficient in the catalysis of isoprene formation to be useful in an industrial setting.

SUMMARY OF THE INVENTION

[0010] The present invention provides isoprene synthases able to catalyze the synthesis of isoprene with an activity greater than that found in natually occurring isoprene synthases. Among the factors contributing to the present invention are novel insights into the structure and activity of known isoprene synthases, and other closely related hemiterpenoid and terpenoid synthases. In various embodiments, the superior isoprene synthases of the invention are produced by one or more of the following routes: (a) identification and, preferably, isolation, of a novel native isoprene synthase, (b) conversion of a methyl butenol synthase into an isoprene synthase by the addition, removal or substitution of one or more amino acids, (c) replacement of one or more unstructured arm of a characterized isoprene synthase with the arm of a snapdragon monoterpene synthase, e.g., one that makes the acyclic monoterpenes ocimene and myrcene, (d) conversion of an acyclic monoterpene synthase that removes diphosphate from an allylic isoprenoid precursor into an isoprene synthase by substitution of one or amino acids in their active sites, e.g., with phenylalanine, tyrosine or tryptophan, (e) opening up the active site of a characterized isoprene synthase by replacing one phenylalanine in the active site with tyrosine or tryptophan and replacing another phenylalanine in the active site with an amino acid smaller than phenylalanine, (f) conversion of an enzyme that uses dimethylallyl diphosphate (DMADP) and a cation intermediate into an isoprene synthase by reducing the size of its binding pocket by replacing amino acids that line the active site with phenylalanines, (g) conversion of pentalenene synthase into an isoprene producing enzyme through replacement of amino acids in its active site, (h) conversion of phellandene synthase into an isoprene producing enzyme through replacement of amino acids in its active site, and (g) other mutations of conserved amino acids in the active site through addition, deletion or substitution of amino acids.

[0011] In various embodiments, one or more of the approaches listed above are combined to produce a single novel isoprene synthase to further increase its efficiency and/or activity relative to a wild type isoprene synthase in the synthesis of isoprene. In exemplary embodiments, the active site of the novel isoprene synthases may be modified by opening up their active sites through replacement of the phenylalanines smaller amino acids such as leucine. In various embodiments, replacement of the unstructured arm of an isoprene synthase (as modeled in (Sharkey et al., (2005) *Plant Physiology* **137**: 700-712) and shown in (Köksal et al., (2010) *Journal of Molecular Biology* **402:** 363-373) with the arm of snapdragon monoterpene synthase that lacks a canonical double arginine is combined with a novel isoprene synthase or an isoprene synthase whose active site has been opened up through replacement of the phenylalanines

[0012] In various embodiments of the invention, the thermostability of the novel isoprene synthase will be increased relative to the corresponding wild type. The native isoprene synthases have evolved to function in terrestrial plants at temperatures that vary from the culture conditions ideal for microbes in an industrial bioreactor or fermenter. The present invention provides novel isoprene synthases having optimal activity at the optimal temperature for microbes producing isoprene in a bioreactor or fermenter.

[0013] In other embodiments of the invention, it is provided variant isoprene synthases that are more resistant than the corresponding wild type enzymes to oxidative stress.

[0014] In various embodiments of the invention, the novel isoprene synthases will be modified relative to the corresponding wild type enzyme to facilitate greater activity in their new setting in either microbial hosts or combinations of isolated enzymes. These modifications include changing codon usage from that preferred by terrestrial plants to that preferred by the chosen microbial host. These modifications optionally further include removal of cell localization sequences that direct isoprene to compartments within a terrestrial plant cell, for example, the chloroplast that microbes do not have. These modifications optionally further include changes to increase the stability and solubility of the enzymes producing isoprene.

[0015] In another aspect of the invention, it is provided an isolated nucleic acid sequence having a sequence encoding an isoprene synthase variant of a parent eucalyptus

isoprene synthase, said sequence operably linked to a promoter, wherein the isoprene synthase variant is capable of catalyzing a reaction that synthesizes isoprene from DMADP (dimethylallyl disphosphate) and said isoprene synthase variant is truncated at the N-terminus as compared to the parent eucalyptus isoprene synthase. In some embodiments, the parent eucalyptus isoprene synthase is *E. globulus* of SEQ ID NO: 6. The isoprene synthase variant may have a sequence that has at least five, at least ten, at least fifteen, at least twenty, at least twenty five or at least thirty amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase.

[0016] Alternatively, the isoprene synthase variant has a sequence that has five to thirty five, ten to thirty five, fifteen to thirty five, twenty to thirty five, twenty five to thirty five, or thirty to thirty five amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase. In some embodiments, the isoprene synthase variant has a sequence that has thirty five amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase. In still other embodiments, the isoprene synthase variant has a sequence that has less than thirty five amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase.

[0017] In still another aspect of the invention, it is provided an isolated nucleic acid sequence having a sequence encoding a variant of a parent solanum phellandrene synthase, said sequence operably linked to a promoter, wherein the variant is capable of catalyzing a reaction that synthesizes isoprene from DMADP (dimethylallyl disphosphate) and said variant is truncated at the N-terminus as compared to the parent solanum phellandrene synthase. In some embodiments, the parent solanum phellandrene synthase is *S. lycopersicum* of SEQ ID NO: 23. The isoprene synthase variant may have a sequence that has at least five, at least ten, at least fifteen, at least twenty, at least twenty five, at least thirty, or at least thirty five amino acids truncated from the N-terminus as compared to the parent solanum phellandrene synthas.

[0018] Alternatively, the variant has a sequence that has five to thirty six, ten to thirty six, fifteen to thirty six, twenty to thirty six, twenty five to thirty six, or thirty to thirty six amino acids truncated from the N-terminus as compared to the parent solanum phellandrene synthas. In some embodiments, the variant has a sequence that has thirty six amino acids truncated from the N-terminus as compared to the parent solanum

phellandrene synthas. In still other embodiments, the variant has a sequence that has less than thirty six amino acids truncated from the N-terminus as compared to the parent solanum phellandrene synthas.

[0019] The promoter is, in some embodiments, a prokaryotic promoter, *e.g.* pTrc promoter. Those of skill in the art would be able to select other suitable prokaryotic promoter for use in the present invention. In preferred embodiments, the promoter is not a strong promoter.

[0020] In another aspect of the invention, it is provided an expression vector ocomprising the nucleic acid sequence described above. In still another aspect of the invention, it is provided an isolated host cell comprising the heterologous nucleic acid sequence or expression vector described above. The host cell may in some embodiments be a bacterial cell, *e.g. Escherichia coli*. In some embodiments, to avoid rate-limiting steps in the isoprene production process, the host cell may further comprise one or more recombinant nucleic acid sequence of a MEP pathway gene, such as one or more selected from *dxs*, *ispD*, *ispF*, and *idi*.

[0021] In other embodiments of the invention, it is provided an isolated isoprene synthase variant of a parent eucalyptus isoprene synthase, wherein said variant comprises a truncation in the N-terminal portion of isoprene synthase as compared to the parent eucalyptus isoprene synthase and wherein said variant is capable of catalyzing a reaction that synthesizes isoprene from DMADP (dimethylallyl disphosphate). Also provided is a variant of a parent solanum phellandrene synthase, wherein said variant comprises a truncation in the N-terminal portion of isoprene synthase as compared to the parent eucalyptus isoprene synthase and wherein said variant is capable of catalyzing a reaction that synthesizes isoprene from DMADP (dimethylallyl disphosphate). Also provided are methods of producing isoprene comprising: (a) providing a host cell comprising an expression vector including the nucleic acid sequence described above; and (b) culturing the host cell under conditions suitable for producing isoprene, *e.g.* optionally further comprising (c) recovering the isoprene, *e.g.* still optionally further comprising (d) polymerizing the isoprene.

BRIEF DESCRIPTION OF DRAWINGS

[0022] The accompanying drawings are not intended to be drawn to scale. In the drawings, each identical or nearly identical component that is illustrated in various figures is represented by a like numeral. For purposes of clarity, not every component is labeled in every drawing. In the drawings:

[0023] **FIGURE 1**: Kinetic behavior of isoprene synthase proteins from poplar (Schnitzler et al., (2005) *Planta* **222**: 777-786). Top panels show normal and Hanes-Woolf plots of native (circle) and un-tagged recombinant (triangle) isoprene synthase. Bottom plots show N-terminal (open triangle) and C-terminal (open square) isoprene synthase. These results suggest isoprene synthase may exhibit substrate inhibition and cooperative kinetics.

[0024] **FIGURE 2**: Mechanism of hemiterpene synthesis from dimethylallyl diphosphate (DMADP).

[0025] **FIGURE 3**: Reaction mechanisms leading to hemiterpenes and monoterpenes. Cyclic monoterpenes require a rotation around the 2-3 bond, which is facilitated by reattaching the diphosphate to carbon 3 of the linalyl cation. This step is not required for making isoprene or acyclic monoterpenes.

[0026] **FIGURE 4**: Isoprene is made from dimethylallyl diphosphate by elimination of the diphosphate to yield a carbocation intermediate. Abstraction of any one of six protons of the two methyl groups leads to the formation of isoprene. Similar chemistry beginning with geranyl diphosphate leads to acyclic monoterpenes. Proton abstraction of any of the methyl hydrogens leads to beta myrcene while either E (trans) or Z (cis) beta ocimene is made depending upon which proton of carbon 4 is abstracted.

[0027] **FIGURE 5**: Structure of bornyl diphosphate synthase (PDB 1N1Z). The black line starting at the upper right is the RRX_8W unstructured arm. Gray ribbons are the β-subunit helices with no direct role in catalysis. The orange ribbon is helix A of the α subunit. The short sand colored stretch is the C-terminus. The unstructured black N-terminal amino acids, the orange A-helix amino acids, and the C-terminal amino acids must fold together in the region that joins the two subunits. This may be the cause of frequent misfolding found in isoprene synthases expressed in bacteria. Blue = reaction

product, green = Mg, orange spheres = diphosphate, red = DDXXD sequence. Visualized with MacPyMOL.

[0028] FIGURE 6: N and C termini of bornyl diphosphate synthase and modeled ocimene synthase of snapdragon. Black = N terminus of BPPS, Gray = modeled N terminus of ocimene synthase. Purple = identically located leucine residues that will serve as the cut over location for a chimeric enzyme. Green is BPPS structure while blue is ocimene synthase modeled structure. The sand-colored line is the C-terminus of BPPS that does not occur in ocimene synthase and will be cleaved in the chimeric enzyme. Modeled with SWISS-MODEL (Arnold et al., (2006) *Bioinformatics* **22**: 195-201; Guex and Peitsch, (1997) *Electrophoresis* **18**: 2714-2723; Schwede et al., (2003) *Nucleic Acids Research* **31**: 3381-3385).

[0029] FIGURE 7: View of the binding pocket taken from 3I4X (PDB crystal structure, drawn in MacPyMol). The purple color is surface provided by tyrosines, green are carbons of the DMADP analog DMASP. Exemplary mutations to the active site include, either alone or in combination, converting the lower two tyrosines to phenylalanine and converting one or more amino acids on the facing wall to phenylalanine.

[0030] FIGURE 8: The *in vivo* isoprene production by *E. coli* BL21 strains expressing either *E. globulus* or *P. alba* isoprene synthases, with or without heterologous expression of *dxs*, *ispD*, *ispF*, and *idi*, was measured in a FIS as described in Example 3.3.

[0031] FIGURE 9: Specific activity was obtained at various DMADP concentrations for *E. globulus* isoprene synthase and indicated by the light grey diamonds. Data were fitted to Equation 1 using the constants shown in Table 1 and the fitted equation is shown by the dark grey line as described in Example 3.5.

[0032] FIGURE 10: A comparison of the relative expression and solubility of *E. globulus*, *M. alterniforia*, and *R. pseudoacacia* isoprene synthases is shown. Total, soluble, and insoluble protein fractions are shown after induction of the indicated plasmids described in Example 4.4 and the *E. globulus* isoprene synthase expressed at high levels in soluble form.

[0033] **FIGURE 11:** Specific activity was obtained according to Example 5.2 at various DMADP concentrations for *S. lycopersicum* phellandrene synthase and indicated by the plotted line.

DETAILED DESCRIPTION OF THE INVENTION

Defintions

[0034] This invention is not limited in its application to the details of construction and the arrangement of components set forth in the following description or illustrated in the drawings. The invention is capable of other embodiments and of being practiced or of being carried out in various ways. The phraseology and terminology used herein is for the purpose of description and should not be regarded as limiting.

[0035] "Sustainable energy," as used herein, refers broadly to energy other than fossil fuels. Exemplary sources of sustainable energy include, but are not limited to, solar energy, water power, wind power, geothermal energy, wave energy, and energy produced from other sources, such as wastes and renewables.

[0036] As used herein, the term "hydrocarbon compounds" includes hydrocarbons and hydrocarbon derivatives, e.g., alcohol, halide, thiol, ether, aldehyde, ketone, carboxylic acid, ester, amine, and amide, etc.

[0037] As used herein, the term "carbonaceous chemical," refers to any carbon-containing chemical that can be produced by a biocatalyst. In various embodiments, the carbonaceous chemical is a hydrocarbon, while in other embodiments, the chemical includes one or more heteroatoms, e.g., O, S, N, P and the like. The heteroatoms can be joined to one or more carbon atoms or, when there is more than one heteroatom, they are optionally joined to each other, e.g., SO_3H. The carbonaceous chemical can include residues that are alkyl, heteroalkyl, aryl or heteroaryl residues.

[0038] In various embodiments, the method and system of the invention is of use to produce a carbonaceous chemical in an "essentially pure state." As used herein, the term "essentially pure state," refers to a purity of at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% at least 99.5%, at least 99.9% or at least 99.95%.

[0039] The term "alkyl," by itself or as part of substituent, means, unless otherwise stated, a straight or branched chain, or cyclic hydrocarbon radical, or combination thereof, which may be fully saturated, mono- or polyunsaturated and can include mono-, di- and multivalent radicals, having the number of carbon atoms designated (i.e. C_1-C_{10} means one to ten carbons). In some embodiments, the term "alkyl" means a straight or branched chain, or combinations thereof, which may be fully saturated, mono- or polyunsaturated and can include di- and multivalent radicals. Examples of saturated hydrocarbon radicals include, but are not limited to, groups such as methyl, ethyl, n-propyl, isopropyl, n-butyl, t-butyl, isobutyl, sec-butyl, cyclohexyl, (cyclohexyl)methyl, cyclopropylmethyl, homologs and isomers of, for example, n-pentyl, n-hexyl, n-heptyl, n-octyl, and the like. An unsaturated alkyl group is one having one or more double bonds or triple bonds. Examples of unsaturated alkyl groups include, but are not limited to, vinyl, 2-propenyl, crotyl, 2-isopentenyl, 2-(butadienyl), 2,4-pentadienyl, 3-(1,4-pentadienyl), ethynyl, 1- and 3-propynyl, 3-butynyl, and the higher homologs and isomers. The term "alkyl," unless otherwise noted, is also meant to include those derivatives of alkyl defined in more detail below, such as "heteroalkyl" with the difference that the heteroalkyl group, in order to qualify as an alkyl group, is linked to the remainder of the molecule through a carbon atom. Alkyl groups that are limited to hydrocarbon groups are termed "homoalkyl".

[0040] The term "alkenyl" by itself or as part of another substituent is used in its conventional sense, and refers to a radical derived from an alkene, as exemplified, but not limited, by substituted or unsubstituted vinyl and substituted or unsubstituted propenyl. Typically, an alkenyl group will have from 1 to 24 carbon atoms, with those groups having from 1 to 10 carbon atoms being useful examplars.

[0041] The term "alkylene" by itself or as part of another substituent means a divalent radical derived from an alkane, as exemplified, but not limited, by $-CH_2CH_2CH_2CH_2-$, and further includes those groups described below as "heteroalkylene." Typically, an alkyl (or alkylene) group will have from 1 to 24 carbon atoms, with those groups having 10 or fewer carbon atoms being useful exemplars in the present invention. A "lower alkyl" or "lower alkylene" is a shorter chain alkyl or alkylene group, generally having eight or fewer carbon atoms.

[0042] The terms "alkoxy," "alkylamino" and "alkylthio" (or thioalkoxy) are used in their conventional sense, and refer to those alkyl groups attached to the remainder of the molecule via an oxygen atom, an amino group, or a sulfur atom, respectively.

[0043] The term "heteroalkyl," by itself or in combination with another term, means, unless otherwise stated, a stable straight or branched chain, or cyclic hydrocarbon radical, or combinations thereof, consisting of the stated number of carbon atoms and at least one heteroatom selected from the group consisting of O, N, Si, S, B and P and wherein the nitrogen and sulfur atoms may optionally be oxidized and the nitrogen heteroatom may optionally be quaternized. In some embodiments, the term "heteroalkyl," by itself or in combination with another term, means a stable straight or branched chain, or combinations thereof, consisting of the stated number of carbon atoms and at least one heteroatom. The heteroatom(s) may be placed at any interior position of the heteroalkyl group or at the position at which the alkyl group is attached to the remainder of the molecule. Examples include, but are not limited to, $-CH_2-CH_2-O-CH_3$, $-CH_2-CH_2-NH-CH_3$, $-CH_2-CH_2-N(CH_3)-CH_3$, $-CH_2-S-CH_2-CH_3$, $-CH_2-CH_2,-S(O)-CH_3$, $-CH_2-CH_2-S(O)_2-CH_3$, $-CH=CH-O-CH_3$, $-Si(CH_3)_3$, $-CH_2-CH=N-OCH_3$, and $-CH=CH-N(CH_3)-CH_3$. Up to two heteroatoms may be consecutive, such as, for example, $-CH_2-NH-OCH_3$ and $-CH_2-O-Si(CH_3)_3$. Similarly, the term "heteroalkylene" by itself or as part of another substituent means a divalent radical derived from heteroalkyl, as exemplified, but not limited by, $-CH_2-CH_2-S-CH_2-CH_2-$ and $-CH_2-S-CH_2-CH_2-NH-CH_2-$. For heteroalkylene groups, heteroatoms can also occupy either or both of the chain termini (e.g., alkyleneoxy, alkylenedioxy, alkyleneamino, alkylenediamino, and the like). Still further, for alkylene and heteroalkylene linking groups, no orientation of the linking group is implied by the direction in which the formula of the linking group is written. For example, the formula $-CO_2R'-$ represents both $-C(O)OR'$ and $-OC(O)R'$.

[0044] The terms "cycloalkyl" and "heterocycloalkyl", by themselves or in combination with other terms, represent, unless otherwise stated, cyclic versions of "alkyl" and "heteroalkyl", respectively. Additionally, for heterocycloalkyl, a heteroatom can occupy the position at which the heterocycle is attached to the remainder of the molecule. A "cycloalkyl" or "heterocycloalkyl" substituent may be attached to the remainder of the molecule directly or through a linker, wherein the linker is preferably

alkylene. Examples of cycloalkyl include, but are not limited to, cyclopentyl, cyclohexyl, 1-cyclohexenyl, 3-cyclohexenyl, cycloheptyl, and the like. Examples of heterocycloalkyl include, but are not limited to, 1-(1,2,5,6-tetrahydropyridyl), 1-piperidinyl, 2-piperidinyl, 3-piperidinyl, 4-morpholinyl, 3-morpholinyl, tetrahydrofuran-2-yl, tetrahydrofuran-3-yl, tetrahydrothien-2-yl, tetrahydrothien-3-yl, 1-piperazinyl, 2-piperazinyl, and the like.

[0045] The terms "halo" or "halogen," by themselves or as part of another substituent, mean, unless otherwise stated, a fluorine, chlorine, bromine, or iodine atom. Additionally, terms such as "haloalkyl," are meant to include monohaloalkyl and polyhaloalkyl. For example, the term "halo(C_1-C_4)alkyl" is mean to include, but not be limited to, trifluoromethyl, 2,2,2-trifluoroethyl, 4-chlorobutyl, 3-bromopropyl, and the like.

[0046] The term "aryl" means, unless otherwise stated, a polyunsaturated, aromatic, substituent that can be a single ring or multiple rings (preferably from 1 to 3 rings), which are fused together or linked covalently. The term "heteroaryl" refers to aryl groups (or rings) that contain from one to four heteroatoms selected from N, O, S, Si and B, wherein the nitrogen and sulfur atoms are optionally oxidized, and the nitrogen atom(s) are optionally quaternized. A heteroaryl group can be attached to the remainder of the molecule through a heteroatom. Non-limiting examples of aryl and heteroaryl groups include phenyl, 1-naphthyl, 2-naphthyl, 4-biphenyl, 1-pyrrolyl, 2-pyrrolyl, 3-pyrrolyl, 3-pyrazolyl, 2-imidazolyl, 4-imidazolyl, pyrazinyl, 2-oxazolyl, 4-oxazolyl, 2-phenyl-4-oxazolyl, 5-oxazolyl, 3-isoxazolyl, 4-isoxazolyl, 5-isoxazolyl, 2-thiazolyl, 4-thiazolyl, 5-thiazolyl, 2-furyl, 3-furyl, 2-thienyl, 3-thienyl, 2-pyridyl, 3-pyridyl, 4-pyridyl, 2-pyrimidyl, 4-pyrimidyl, 5-benzothiazolyl, purinyl, 2-benzimidazolyl, 5-indolyl, 1-isoquinolyl, 5-isoquinolyl, 2-quinoxalinyl, 5-quinoxalinyl, 3-quinolyl, and 6-quinolyl. Substituents for each of the above noted aryl and heteroaryl ring systems are selected from the group of acceptable substituents described below.

[0047] For brevity, the term "aryl" when used in combination with other terms (e.g., aryloxy, arylthioxy, arylalkyl) optionally includes both aryl and heteroaryl rings as defined above. Thus, the term "arylalkyl" is meant to include those radicals in which an aryl group is attached to an alkyl group (e.g., benzyl, phenethyl, pyridylmethyl and the

like) including those alkyl groups in which a carbon atom (e.g., a methylene group) has been replaced by, for example, an oxygen atom (e.g., phenoxymethyl, 2-pyridyloxymethyl, 3-(1-naphthyloxy)propyl, and the like).

[0048] Each of the above terms (e.g., "alkyl," "heteroalkyl," "aryl" and "heteroaryl") are meant to include both substituted and unsubstituted forms of the indicated radical. Exemplary substituents for each type of radical are provided below.

[0049] Substituents for the alkyl and heteroalkyl radicals (including those groups often referred to as alkylene, alkenyl, heteroalkylene, heteroalkenyl, alkynyl, cycloalkyl, heterocycloalkyl, cycloalkenyl, and heterocycloalkenyl) are generically referred to as "alkyl group substituents," and they can be one or more of a variety of groups selected from, but not limited to: substituted or unsubstituted alkyl, substituted or unsubstituted heteroalkyl, substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, substituted or unsubstituted cycloalkyl, substituted or unsubstituted heterocycloalkyl, -OR', =O, =NR', =N-OR', -NR'R", -SR', -halogen, -SiR'R"R"', -OC(O)R', -C(O)R', -CO_2R', -CONR'R", -OC(O)NR'R", -NR"C(O)R', -NR'-C(O)NR"R"', -NR"$C(O)_2$R', -NR-C(NR'R"R"')=NR"", -NR-C(NR'R")=NR"', -S(O)R', -$S(O)_2$R', -$S(O)_2$NR'R", -$NRSO_2$R', -CN and $-NO_2$ in a number ranging from zero to (2m'+1), where m' is the total number of carbon atoms in such radical. R', R", R"' and R"" each preferably independently refer to hydrogen, substituted or unsubstituted heteroalkyl, substituted or unsubstituted aryl, e.g., aryl substituted with 1-3 halogens, substituted or unsubstituted alkyl, alkoxy or thioalkoxy groups, or arylalkyl groups. When a compound of the present inventions includes more than one R group, for example, each of the R groups is independently selected as are each R', R", R"' and R"" groups when more than one of these groups is present. When R' and R" are attached to the same nitrogen atom, they can be combined with the nitrogen atom to form a 5-, 6-, or 7-membered ring. For example, -NR'R" is meant to include, but not be limited to, 1-pyrrolidinyl and 4-morpholinyl. From the above discussion of substituents, one of skill in the art will understand that the term "alkyl" is meant to include groups including carbon atoms bound to groups other than hydrogen groups, such as haloalkyl (e.g., -CF_3 and $-CH_2CF_3$) and acyl (e.g., -$C(O)CH_3$, -$C(O)CF_3$, -$C(O)CH_2OCH_3$, and the like).

[0050] Similar to the substituents described for the alkyl radical, substituents for the aryl and heteroaryl groups are generically referred to as "aryl group substituents." The substituents are selected from, for example: substituted or unsubstituted alkyl, substituted or unsubstituted heteroalkyl, substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, substituted or unsubstituted cycloalkyl, substituted or unsubstituted heterocycloalkyl, -OR', =O, =NR', =N-OR', -NR'R", -SR', -halogen, -SiR'R"R"', -OC(O)R', -C(O)R', -CO_2R', -CONR'R", -OC(O)NR'R", -NR"C(O)R', -NR'-C(O)NR"R"', -NR"$C(O)_2$R', -NR-C(NR'R"R'")=NR"", -NR-C(NR'R")=NR'", -S(O)R', -$S(O)_2$R', -$S(O)_2$NR'R", -$NRSO_2$R', -CN and –NO_2, -R', -N_3, -$CH(Ph)_2$, fluoro(C_1-C_4)alkoxy, and fluoro(C_1-C_4)alkyl, in a number ranging from zero to the total number of open valences on the aromatic ring system; and where R', R'', R''' and R'''' are preferably independently selected from hydrogen, substituted or unsubstituted alkyl, substituted or unsubstituted heteroalkyl, substituted or unsubstituted aryl and substituted or unsubstituted heteroaryl. When a compound of the present invention includes more than one R group, for example, each of the R groups is independently selected as are each R', R'', R''' and R'''' groups when more than one of these groups is present.

[0051] Two of the substituents on adjacent atoms of the aryl or heteroaryl ring may optionally be replaced with a substituent of the formula –T-C(O)-$(CRR')_q$-U-, wherein T and U are independently –NR-, -O-, -CRR'- or a single bond, and q is an integer of from 0 to 3. Alternatively, two of the substituents on adjacent atoms of the aryl or heteroaryl ring may optionally be replaced with a substituent of the formula –A-$(CH_2)_r$-B-, wherein A and B are independently –CRR'-, -O-, -NR-, -S-, -S(O)-, -$S(O)_2$-, -$S(O)_2$NR'- or a single bond, and r is an integer of from 1 to 4. One of the single bonds of the new ring so formed may optionally be replaced with a double bond. Alternatively, two of the substituents on adjacent atoms of the aryl or heteroaryl ring may optionally be replaced with a substituent of the formula –$(CRR')_s$-X-$(CR"R'")_d$-, where s and d are independently integers of from 0 to 3, and X is –O-, -NR'-, -S-, -S(O)-, -$S(O)_2$-, or –$S(O)_2$NR'-. The substituents R, R', R" and R'" are preferably independently selected from hydrogen or substituted or unsubstituted (C_1-C_6)alkyl.

[0052] As used herein, the term "acyl" describes a substituent containing a carbonyl residue, C(O)R. Exemplary species for R include H, halogen, substituted or

unsubstituted alkyl, substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, and substituted or unsubstituted heterocycloalkyl.

[0053] As used herein, the term "fused ring system" means at least two rings, wherein each ring has at least 2 atoms in common with another ring. "Fused ring systems" may include aromatic as well as non-aromatic rings. Examples of "fused ring systems" are naphthalenes, indoles, quinolines, chromenes and the like.

[0054] As used herein, the term "heteroatom" includes oxygen (O), nitrogen (N), sulfur (S), silicon (Si) and boron (B).

[0055] The symbol "R" is a general abbreviation that represents a substituent group. Exemplary substituent groups include substituted or unsubstituted alkyl, substituted or unsubstituted heteroalkyl, substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, and substituted or unsubstituted heterocycloalkyl groups.

[0056] By "amino acid" and "amino acid identity" as used herein is meant one of the 20 naturally occurring amino acids or any non-natural analogues that may be present at a specific, defined position. By "protein" herein is meant at least two covalently attached amino acids, which includes proteins, polypeptides, oligopeptides and peptides. The protein may be made up of naturally occurring amino acids and peptide bonds, or synthetic peptidomimetic structures, i.e. "analogs", such as peptoids (see, Simon et al., PNAS USA 89(20):9367 (1992)) particularly when LC peptides are to be administered to a patient. Thus "amino acid", or "peptide residue", as used herein means both naturally occurring and synthetic amino acids. For example, homophenylalanine, citrulline and norleucine are considered amino acids for the purposes of the invention.

[0057] "Amino acid" also includes imino acid residues such as proline and hydroxyproline. The side chain may be in either the (R) or the (S) configuration. In the preferred embodiment, the amino acids are in the (S) or L-configuration. If non-naturally occurring side chains are used, non-amino acid substituents may be used, for example to prevent or retard in vivo degradation.

[0058] As used herein, the terms "starting gene" and "parent gene" refer to a nucleic acid which is a gene of interest that encodes a protein of interest that is to be improved and/or changed using the present invention. Likewise, the terms "starting protein" and

"parent protein" refer to a protein of interest that is to be improved and/or changed using the present invention.

[0059] In various embodiments, the nucleic acid is a recombinant nucleic acid. For instance, in some embodiments, an isoprene synthase nucleic acid is operably linked to another nucleic acid encoding all or a portion of another polypeptide such that the recombinant nucleic acid encodes a fusion polypeptide that includes an isoprene synthase and all or part of another polypeptide (e.g., a peptide that facilitates purification or detection of the fusion polypeptide, such as a His-tag). In some embodiments, part or all of a recombinant nucleic acid is chemically synthesized. In some aspects, the nucleic acid is a heterologous nucleic acid. By "heterologous nucleic acid" is meant a nucleic acid whose nucleic acid sequence is not identical to that of another nucleic acid naturally found in the same host cell.

[0060] In particular embodiments the nucleic acid includes a segment of or the entire nucleic acid sequence of any naturally-occurring isoprene synthase nucleic acid. In some embodiments, the nucleic acid includes at least or about 50, 100, 150, 200, 300, 400, 500, 600, 700, 800, or more contiguous nucleotides from a naturally-occurring isoprene synthase nucleic acid. In some aspects, the nucleic acid has one or more mutations compared to the sequence of a wild-type (i.e., a sequence occurring in nature) isoprene synthase nucleic acid. In some embodiments, the nucleic acid has one or more mutations (e.g., a silent mutation) that increase the transcription or translation of isoprene synthase nucleic acid. In some embodiments, the nucleic acid is a degenerate variant of any nucleic acid encoding an isoprene synthase polypeptide.

[0061] An isoprene synthase nucleic acid can be incorporated into a vector, such as an expression vector, using standard techniques (Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor, 2001, which is hereby incorporated by reference in its entirety, particularly with respect to the screening of appropriate DNA sequences and the construction of vectors). Methods used to ligate the DNA construct comprising a nucleic acid of interest such as isoprene synthase, a promoter, a terminator, and other sequences and to insert them into a suitable vector are well known in the art. Additionally, vectors can be constructed using known recombination techniques (e.g., Invitrogen Life Technologies, Gateway Technology).

[0062] As used herein, "homologous genes" refers to a pair of genes from different, but usually related species, which correspond to each other and which are identical or very similar to each other. The term encompasses genes that are separated by speciation (i.e., the development of new species) (e.g., orthologous genes), as well as genes that have been separated by genetic duplication (e.g., paralogous genes).

[0063] As used herein, "ortholog" and "orthologous genes" refer to genes in different species that have evolved from a common ancestral gene (i.e., a homologous gene) by speciation. Typically, orthologs retain the same function during the course of evolution. Identification of orthologs finds use in the reliable prediction of gene function in newly sequenced genomes.

[0064] As used herein, "paralog" and "paralogous genes" refer to genes that are related by duplication within a genome. While orthologs retain the same function through the course of evolution, paralogs evolve new functions, even though some functions are often related to the original one.

[0065] As used herein, "homology" refers to sequence similarity or identity, with identity being preferred. This homology is determined using standard techniques known in the art (See e.g., Smith and Waterman, Adv Appl Math, 2:482, 1981; Needleman and Wunsch, J Mol Biol, 48:443, 1970; Pearson and Lipman, Proc Natl Acad Sci USA, 85:2444, 1988; programs such as GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, Madison, Wis.; and Devereux et al., Nucl Acid Res, 12:387-395, 1984).

[0066] As used herein, an "analogous sequence" of an isoprene synthase is one wherein the function of the gene is essentially the same as the gene based on the kudzu isoprene synthase. Additionally, analogous genes include at least 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99% or 100% sequence identity with the sequence of the kudzu isoprene synthase. In additional embodiments more than one of the above properties applies to the sequence. Analogous sequences are determined by known methods of sequence alignment. A commonly used alignment method is BLAST, although as indicated above and below, there are other methods that also find use in aligning sequences.

[0067] "Percent sequence identity," "percent amino acid sequence identity," "percent gene sequence identity," and/or "percent nucleic acid/polynucleotide sequence identity," with respect to two amino acids, polynucleotide and/or gene sequences (as appropriate), refer to the percentage of residues that are identical in the two sequences when the sequences are optimally aligned. Thus, 80% amino acid sequence identity means that 80% of the amino acids in two optimally aligned polypeptide sequences are identical.

[0068] The phrase "substantially identical" in the context of two nucleic acids or polypeptides thus refers to a polynucleotide or polypeptide that comprising at least 70% sequence identity, preferably at least 75%, preferably at least 80%, preferably at least 85%, preferably at least 90%, preferably at least 95%, preferably at least 97%, preferably at least 98% and preferably at least 99% sequence identity as compared to a reference sequence using the programs or algorithms (e.g., BLAST, ALIGN, CLUSTAL) using standard parameters. One indication that two polypeptides are substantially identical is that the first polypeptide is immunologically cross-reactive with the second polypeptide. Typically, polypeptides that differ by conservative amino acid substitutions are immunologically cross-reactive. Thus, a polypeptide is substantially identical to a second polypeptide, for example, where the two peptides differ only by a conservative substitution. Another indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions (e.g., within a range of medium to high stringency).

[0069] Various isoprene synthase polypeptides and nucleic acids can be used in the compositions and methods of the invention. As used herein, "polypeptides" includes polypeptides, proteins, peptides, fragments of polypeptides, and fusion polypeptides that include part or all of a first polypeptide (e.g., an isoprene synthase) and part or all of a second polypeptide (e.g., a peptide that facilitates purification or detection of the fusion polypeptide, such as a His-tag). In various embodiments, a polypeptide has at least or about 50, 100, 150, 175, 200, 250, 300, 350, 400, or more amino acids. In some embodiments, the polypeptide fragment contains at least or about 25, 50, 75, 100, 150, 200, 300, or more contiguous amino acids from a full-length polypeptide and has at least or about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 100% or greater than 100% of an activity of a corresponding full-length polypeptide. In particular

embodiments, the polypeptide includes a segment of or the entire amino acid sequence of any naturally-occurring isoprene synthase. In some embodiments, the polypeptide has one or more mutations compared to the sequence of a wild-type (i.e., a sequence occurring in nature) isoprene synthase.

[0070] In some embodiments, the polypeptide is a heterologous polypeptide. By "heterologous polypeptide" is meant a polypeptide whose amino acid sequence is not identical to that of another polypeptide naturally expressed in the same host cell.

[0071] By "variant polypeptide" as used herein is meant a polypeptide sequence that differs from that of a parent polypeptide sequence by addition, deletion or substitution of at least one amino acid modification. Variant polypeptide may refer to the polypeptide itself, a composition comprising the polypeptide, or the nucleic acid sequence that encodes it. Preferably, the variant polypeptide has at least one amino acid modification compared to the parent polypeptide, e.g. from about one to about ten amino acid modifications, and preferably from about one to about five amino acid modifications compared to the parent. The variant polypeptide sequence herein will preferably possess at least about 80% identity with a parent polypeptide sequence, and most preferably at least about 90% identity, more preferably at least about 95% identity. Accordingly, by "isoprene synthase variant" as used herein is meant an isoprene synthase sequence that differs from that of a parent isoprene synthase sequence by virtue of at least one amino acid modification.

[0072] As used herein, the terms "active site," "binding site" or "binding pocket" refer to a region of a polypeptide or a molecular complex comprising the polypeptide that, as a result of the primary amino acid sequence of the polypeptide and/or its three-dimensional shape, favorably associates with another chemical entity or compound including ligands or inhibitors. Thus, an active site may include or consist of features such as interfaces between domains. Chemical entities or compounds that may associate with an active site include, but are not limited to, compounds, ligands, cofactors, substrates, inhibitors, agonists, antagonists, etc.

[0073] "Structural reaction residues" and "site-construction residues" refer to a three-dimensional collection of amino acids involved in an enzymatic reaction. For example, these would include those forming the active site, those coordinating metal ions and

those forming the substrate bind region. In particular, those forming the flexible loops and N-terminus and the adjacent residues that stabilize the flexible segments when substrate is bound. In an exemplary embodiment, the invention provides an enzyme with isoprene synthase activity that is a variant polypeptide modified by addition, deletion or substitution of at least one amino acid that is a structural reaction residue.

[0074] As used herein, "equivalent or homologous residues" refers to amino acid residues that are shared by certain proteins. Equivalent residues may be identified by determining homology at the level of tertiary structure for a terpene synthase (e.g., isoprene synthase) whose tertiary structure has been determined by x-ray crystallography. Equivalent residues are defined as those for which the atomic coordinates of two (2) or more of the main chain atoms of a particular amino acid residue of the terpene synthase having putative equivalent residues and the substrate of interest (e.g., N on N, CA on CA, C on C and O on O) are within 0.2 nm and preferably 0.15 nm after alignment. Alignment is achieved after the best model has been oriented and positioned to give the maximum overlap of atomic coordinates of non-hydrogen protein atoms of the terpene synthases and substrates analyzed. The preferred model is the crystallographic model giving the lowest R factor for experimental diffraction data at the highest resolution available, determined using methods known to those skilled in the art of crystallography and protein characterization/analysis. For example, equivalent residues which are functionally analogous to a specific residue of isoprene synthase are defined as those amino acids at a structurally homologous synthase which may adopt a conformation such that they either alter, modify, or contribute to protein structure, substrate binding or catalysis in a manner defined or attributed to a specific residue of isoprene synthase.

"Biocatalyst", as used herein refers to a microbe that is converts a carbon source into a product, in this case isoprene, that is expelled from the microbe that produced it.

THE EMBODIMENTS

[0075] Embodiments of the present invention provide a key component of new energy and chemical industry solutions that are not dependent on fossil fuels and are consequently sustainable both environmentally and economically. The present invention provides compositions for producing isoprene. Specifically, the composition includes

novel enzymes capable of producing isoprene, some of which are isoprene synthases, that have favorable characteristics in isolation or when expressed in a biocatalyst. In either case they are useful for the efficient and economical production of isoprene from a variety of carbonaceous feedstocks.

Isoprene Synthase Enzyme Kinetics

[0076] Isoprene synthase genes have been cloned and characterized from two plant genera: Populus (poplars and aspen) (Miller, et al., (2001) *Planta* **213**: 483-487; Sharkey, et al., (2005) *Plant Physiology* **137**: 700-712; Sasaki, et al., (2005) *FEBS Letters* **579**: 2514-2518 and Behnke, et al., (2007) *The Plant Journal* **51**: 485-499) and *Pueraria montana* (kudzu) (Sharkey, et al., (2005) *Plant Physiology* **137**: 700-712).

[0077] The enzyme catalyzes the divalent cation-dependent, irreversible, and stoichiometric conversion of dimethylallyl diphosphate (DMADP) to isoprene and pyrophosphate. The wild type protein, first purified from aspen leaves, was determined to be functional as a heterodimer (Silver and Fall, (1995) *J. Biological Chemistry* **270**: 13010-13016). The associated monomers were found to be similar, but not identical (62 and 58 kDa). However, monomers and/or homodimers have activity since single gene products can make isoprene. The native enzyme is sensitive to denaturation due to dilution as well as oxidation, and therefore the samples were kept concentrated and in a high concentration of DTT. This suggests that a free thiol may be required for catalytic activity. The enzyme required a divalent cation for activity, either Mg^{2+} or Mn^{2+}. Kinetic analysis of the native enzyme demonstrates maximal activity at pH 8.0. The apparent k_{cat} and K_m for the enzyme were estimated to be 1.7 s^{-1} and 8 mM respectively, but the enzyme also exhibited significant substrate inhibition (Eq. 1). The highest activity of the native protein was estimated to be 900 nmol isoprene min^{-1} mg-1 at 10 mM DMADP.

Equation 1. Enzymatic rate equation for enzyme with substrate inhibition.

$$Rate = \frac{E_t k_{cat} S}{K_M + S + \frac{S^2}{K_{IS}}}$$

[0078] Characterization of the recombinant hybrid poplar protein expressed in *E. coli* yielded additional insights into this enzyme. Activity was highest from protein obtained when a leader signal peptide was deleted from the gene (Miller, et al., (2001) *Planta* **213**: 483-487). Recombinant proteins with N- and C-terminal polyhistidine tags were compared to the unmodified recombinant protein as well as native protein isolated from hybrid poplars (Schnitzler et al., (2005) *Planta* **222**: 777-786). Unlike the protein from aspen, the hybrid poplar protein and its recombinant versions were all monomeric, although the purification conditions were different from the previous work with the aspen enzyme. This suggests that the heterodimer observed in the aspen leaves may be comprised of alternatively modified versions of the same gene product. The addition of the polyhistidine tags to the protein significantly affected the activity of the recombinant isoprene synthase protein. All of the proteins exhibited a broad pH optimum of 8.0 except for the protein with N-terminal tag, which exhibited a pH optimum of 9.0. All of the enzymes had temperature optima of 40 °C except for the protein with the C-terminal tag, which was more thermostable (optimal activity at 45-50 °C). All of the proteins also displayed interesting kinetic behavior (Figure 1). When plotted in a standard Michaelis-Menten format, it appears that these enzymes also suffer from substrate inhibition above 10 mM DMADP. However, by examining Hanes-Woolf plots, it appears that the enzymes also exhibit a sigmoidal substrate dependency, indicative of positive cooperativity (Eq. 2, Figure 1), and this is most pronounced in the absence of the polyhistidine tag. Positive co-operativity is unusual but not unheard of in monomeric enzymes. Apparent K_m values obtained from the data were in the range of 2-3 mM for all of the poplar enzymes.

Equation 2. Enzymatic rate equation for positive co-operativity.

$$Rate = \frac{E_t k_{cat} S^n}{K_M + S^n}$$

[0079] Recently, a modified Michaelis-Menten equation allowing for both cooperativity and substrate inhibition has been shown to fit methyl butenol and isoprene production of methyl butenol synthase another hemiterpene sythnase (Eq. 3) (Gray et al, (2011) http://www.jbc.org/content/early/2011/04/19/jbc.M111.237438.short). In this equation H is the Hill coefficient describing cooperativity, K_{is} is the binding affinity of

the substrate to the enzyme-substrate complex and x is the number of substrate molecules that could bind to and inactivate the enzyme substrate complex. It is likely that this equation will most closely model the activity of isoprene synthases.

Equation 3. a modified Michaelis-Menten equation allowing for both cooperativity and substrate inhibition

$$v = \frac{V_{max}}{1 + \frac{K_m^H}{S^H} + \frac{S^x}{K_{is}^x}}$$

[0080] The recombinant enzyme from *Populus alba* was not expressed with a polyhistidine tag, and was found to be very similar to that of the hybrid poplar (Sasaki, et al., (2005) *FEBS Letters* **579**: 2514-2518). This protein was also monomeric, had a pH optima of 8.0, a temperature optima of 40 °C, and an apparent K_m of 8.7 mM. No k_{cat} was determined.

[0081] The recombinant enzyme from kudzu was expressed without the leader sequence and with an N-terminal polyhistidine tag (Sharkey, et al., (2005) *Plant Physiology* **137**: 700-712). The kinetic behavior of the enzyme was measured, and a clear sigmoidal dependence on the substrate concentration was observed (Eq. 2), again indicative of cooperative behavior. This results in an apparent K_m of 7.7 mM, Hill coefficient (n) of 4.1, and a k_{cat} of 0.088 mol mol^{-1} s^{-1}. The enzyme was not characterized further, nor was the aspen isoprene synthase protein, which was also cloned in this work.

Isoprene Synthase Mechanism of Action

[0082] The mechanism of action of isoprene synthase can be hypothesized based on the reaction mechanisms of other terpene synthases (Figure 2). The most commonly used reaction model is the well-accepted mechanism for limonene synthase (Rajaonarivony et al., (1992) *Arch. Biochem. Biophys.* **296**: 49-57). A ten-carbon precursor (geranyl diphosphate, GDP) binds to the limonene synthase enzyme. Three magnesium atoms are coordinated around the diphosphate and a signature sequence of DDXXD interacts with some of these magnesiums. A long arm of the protein that has a highly conserved double arginine (RR) moves to close off the active site. This RR has a

conserved glutamate (E) a few base pairs upstream and this constitutes the beginning (N-terminus) of the functional enzyme. The E-RRX_8W beginning of the enzyme binds and closes the active site. The RR is required for the first steps of the reaction. The diphosphate is released from the end of GDP making a linalyl cation (Figure 3). The diphosphate is reattached at carbon 3. This reattachment ensures that the bond between carbons 2 and 3 is a single bond, and this allows rotation around that bond required to convert the linalyl cation to a neryl cation. The neryl cation is the immediate precursor of cyclic monoterpenes (e.g. limonene). Proton abstraction leads to quenching of the cation and typically bond formation between carbons 1 and 6 making a six-membered ring.

[0083] The complicated detachment and reattachment of the diphosphate is almost universal among monoterpene synthases but two notable exceptions exist. First, it was reported that snapdragon flowers have a different class of monoterpene synthases that do not have the canonical RRX_8W. These enzymes are able to make acyclic monoterpenes (ocimene and myrcene) but not cyclic monoterpenes (Dudareva et al., (2003) *The Plant Cell* **15**:1227-1241). It was speculated that the lack of the double R prevents this enzyme from making the linalyl to neryl conversion and so cannot form cyclic monoterpenes. However, acyclic monoterpenes can be formed from the linalyl cation. Second, it was reported that trichomes of tomato have an enzyme (a phellandrene synthase) that uses neryl diphosphate (NDP), the cis isomer of GDP (Schilmiller et al., (2009) *PNAS* **106**: 10865-10870). This molecule forms the neryl cation as soon as the diphosphate is removed and so does not require the rotation around the 2-3 bond required when GDP is the starting material. This simplifies the reaction mechanism. This enzyme can use GDP at a low rate. Like many terpene synthases, phellandrene synthase makes various products. When NDP is the substrate only cyclic products are made but when GDP is the substrate some acyclic monoterpenes are made. This indicates that the acyclic monoterpenes are made from the linalyl cation and when NDP is supplied the neryl cation is not converted to the linalyl cation.

[0084] Additional insight into terpene synthases come from a recent report on kaurene synthase. The bifunctional copalyl diphosphate synthase/kaurene synthase of *Physcomatrella patens* has a similar mechanism in which a cation is quenched by proton

abstraction. Most kaurene synthases make only kaurene but the *P. patens* enzyme will sometimes quench the cation with water resulting in kaurane (Kawaide et al., (2011) *FEBS Journal* **278**: 123-133). A similar reaction in a hemiterpene synthase would produce methyl butenol. Unpublished sequences of methyl butenol synthase show a similar amino acid substitution as the *P. patens* kaurane synthase and may explain why this enzyme makes methyl butenol instead of isoprene.

[0085] Finally, there are now additional insights into what makes an enzyme an isoprene synthase. Modeling (Sharkey et al., (2005) *Plant Physiology* **137**: 700-712) and the crystal structure (Köksal et al., (2010) *J Mol. Biol.* **402**: 363-373) show that two phenylalanines reduce the relative size of the active site in isoprene synthase explaining why these enzymes may not function as monoterpene synthases. However, if this were the only explanation for what makes an isoprene synthase then all monoterpene synthases should act as isoprene synthases. At the very least all acyclic monoterpene synthases should show isoprene synthase activity. This is because the acyclic monoterpene synthase reaction involves proton abstraction from either the methyl carbon to make myrcene of from carbon 4 to make the ocimenes (which ocimene is made depends upon which proton is abstracted) (Figure 4). While it has not been extensively tested, there are no reports of significant isoprene synthase activity in an ocimene or myrcene synthase. Extensive analyses of isoprene synthase active sites have shown that not only are there two phenylalanines making the site smaller, several other phenylalanines line the active site. This makes the active site rich in pi electrons (from the double bonds in phenylalanine). A strong but often overlooked molecular interaction is cation-pi interactions (Dougherty, (1996) *Science* **271**: 163-169; Gallivan and Dougherty, (1999) *PNAS* **96**: 9459-9464). The cation made upon elimination of the diphosphate from DMADP could be stabilized by cation-pi interactions with some of the phenylalanines of the active site. This could give time for the proton abstraction and proton abstraction of any of the six methyl protons would result in the formation of isoprene. Cation quenching by proton abstraction may occur by a number of different mechanisms and may account for the large variety of terpenes that are made by terpene synthases.

[0086] Current published isoprene synthases (from the Poplar family and kudzu) have a low catalytic activity (k_{cat}) and high K_m. The present invention provides an isoprene synthase in which at least one of these parameters are improved. In an exemplary embodiment, one or both of these parameters is improved by at least about 1, by at least about two or by at least about three orders of magnitude. In various embodiments, the improved k_{cat} provides higher yields, e.g., at maximum substrate availability. In various embodiments, the lower K_m provides advantages in the synthesis of isoprene, e.g., the alleviation of probable feedback on substrate production. In an exemplary embodiment, the improved K_m enhances the ease of steering carbon into isoprene production. In various embodiments the invention provides enzymes with isoprene synthase activity with modified stability and/or solubility (e.g., in the fermentation medium or intracellular mileau), thereby providing an increased level of isoprene synthase that can be expressed in a cell or *in vitro* system. In an exemplary embodiment the alterations to stability and/or solubility increase isoprene synthase activity in a cell or *in vitro* systems by at least about an order of magnitude, at least about two orders of magnitude or at least about three orders of magnitude. It is likely that a total increase in isoprene synthase activity in a cell or *in vitro* system of approximately three orders of magnitude will be necessary to make bio-catalyst driven production of isoprene economically viable. Accordingly, the present invention provides enzymes with an isoprene synthase activity that is at least about one order of magnitude greater, at least about two orders of magnitude greater and at least about three orders of magnitude greater in a cell or in vitro environment than the corresponding activity in the parent polypeptide or provide a platform for directed evolution or structure-based prediction of additional changes that will achieve such favourable kinetics..

[0087] The following sections set out how these large gains in isoprene synthase activity are attained.

Identifying novel isoprene synthases with improved native characteristics

[0088] In an exemplary embodiment, the invention provides a novel polypeptide with isoprene synthase activity.

[0089] All efforts to date to improve isoprene synthase have relied on just two platforms, the group of highly similar Populus sequences (SEQ ID NOs: 1-4) and the

kudzu sequence (SEQ ID NO: 5). These have been chosen on the basis of availability, not suitability. With many new sequences now available to us three more platforms are added, the Eucalyptus (*Eucalyptus globulus*) (SEQ ID NO: 6), the Tea Tree (*Melaleuca alternifolia*) (SEQ ID NO: 7) and the Black Locust (*Robinia pseudoacacia*) (SEQ ID NO: 8) isoprene synthases which are still somewhat related to the current genes, have a very different evolutionary history and some unique properties. Isoprene synthase is similar enough to the angiosperm terpene synthases that structures can be reasonably predicted but different enough to provide significant novelty from which to select high value constructs.

Gymnosperm Hemiterpene Based Approach: Converting methyl butenol synthases into isoprene synthases

[0090] In an exemplary embodiment, the invention provides a variant polypeptide with isoprene synthase activity based on a parent methyl butenol synthase polypeptide or other hemiterpene synthase from gymnosperm species converted by addition, deletion or substitution of one or more amino acid into a polypeptide with isoprene synthase activity.

[0091] Currently a highly active methyl butenol synthase has been cloned and expressed from *Pinus sabiniana* (SEQ ID NO: 9) and an inactive isoprene synthase has been cloned from *Picea pungens* (SEQ ID NO: 10). Because of new insight into the control of cation quenching by water (Kawaide et al., (2011) *FEBS Journal* **278**:123-133) the active methyl butenol synthase, which currently makes 3% of its product as isoprene, is converted to produce all isoprene with a single amino acid change (S440F or S440Y) (SEQ ID NOs: 11 and 12).

Acyclic-Monoterpene-Synthase-Based Approach One: Replacement of an unstructured arm of characterized isoprene synthases with the structured arm of myrcene synthase

[0092] In various embodiments, the invention provides a variant polypeptide with isoprene synthase activity. In these embodiments, the unstructured arm of isoprene synthase is replaced with the structured arm of myrcene synthase in a variant polypeptide.

[0093] The *Tps-g* group of monoterpene synthases (Dudareva et al., (2003) *The Plant Cell* **15**: 1227-1241) are the seventh terpenoid synthase subfamily identified. The

subfamilies identified earlier are designated *Tps-a* through *Tps-f* (Bohlmann et al., (1998) *PNAS* **95**: 4126-4133). The *Tps-g* subfamily makes acyclic monoterpenes, presumably because they have a very different initial domain, in that they lack the RRx_8W motif, which is a characteristic feature of the other families of monoterpene synthases: the *Tsp-b* family of angiosperm monoterpene synthases and the *Tsp-d* family of conifer monoterpene synthases. This initial domain closes off the active site and likely moves during catalysis. It is often unresolved in crystal structures but is resolved in PDB 1N1Z (Figure 5). Lack of the RR in *Tps-g* enzymes likely prevents the linalyl to neryl cation conversion required for cyclic monoterpenes but this conversion is not required for acyclic synthases or hemiterpene synthases. This arm is likely to be a significant component of the difficulty in expressing isoprene synthases because of its flexibility and lack of secondary structure. Chimeric proteins based on the native isoprene synthases are generated with the long initial arm replaced by amino acids found in *Tps-g* genes. These genes have a very different initial sequence with amino acids that may help form secondary structure. Homology modeling is used to identify the exact location for switching from the isoprene synthase gene sequence to the *Tps-g* sequence. Specifically, fusion should occur at a conserved leucine at the base of the long arm of all isoprene synthases. The arm substitution is designed to help solubility and has potential to improve the activity in other ways including but not limited to: not allowing the linalyl to neryl rotation that might slow down the enzyme.

[0094] By switching from one enzyme to another at exactly this leucine there is the best chance of making an enzyme that retains the desired activity. The following indicate the position of this leucine in the isoprene synthase identified to date. It is predicted that all future native isoprene synthases will also contain this conserved leucine. The position of this conserved leucine is L148 in the *Populus* sequence (SEQ ID NOs: 1, 2, 3 and 4), L153 in the kudzu (SEQ NO: 5), L137 in the *Eucalyptus* sequence (SEQ ID NO: 6), L137 in the *Melaleuca* sequence (SEQ ID NO: 8) and L91 of the *Robinia* sequence (SEQ ID NO: 9). All amino acids N-terminal of this leucine are replaced with amino acids D46 through L100 of the Snapdragon gene AY195608 (SEQ ID NO: 13) (Figure 6) to produce the hybrids (SEQ ID NOs: 14, 15, 16, 17, and 18). In addition, any changes are made that might be needed to accommodate the new arm sequence are made.

Acyclic Monoterpene Synthase Based Approach Two: Conversion of acyclic monoterpene synthases into isoprene synthases by replacing amino acids in their active sites

[0095] In an exemplary embodiment, the invention provides a variant polypeptide with isoprene synthase activity. Enzymes according to this embodiment, are produced by converting a parent acyclic monoterpene synthase polypeptide into an isoprene synthase by adding, deleting or substituting one or more amino acid residues.

[0096] Given the hypothesis that cation-pi interactions may play a significant role, enzymes that start with diphosphate removal from an allylic isoprenoids precursor can be converted into a variant polypeptide with isoprene synthase activity. Suitable families of enzymes belong to the following named groups further identified by their Enzyme Commission (EC) Numbers. Enzymes that remove diphosphate from geranyl diphosphate including: myrcene synthases (EC 4.2.3.15), S-linalool synthases (EC 4.2.3.25) and R-linalool synthases (EC 4.2.3.26). Enzymes that remove diphosphate from farnesyl diphosphate including: α-farnesene synthase (EC 4.2.3.46) and β-farnesene synthase (EC 4.2.3.47), (3S, 6E) nerolidol synthase (EC 4.2.3.48) and (3R, 6E) nerolidol synthase (EC 4.2.3.49). Enzymes that remove diphosphate from neryl diphosphate including: β-phellandrene synthase (EC 4.2.3.52).

[0097] In order to convert the parent polypeptide to the variant polypeptide one or more amino acids are substituted that line the active site, at the bottom of the pocket that would accommodate DMADP, with phenylalanine, tyrosine, or tryptophan to increase the surface area of the active site that is accounted for by pi electrons. Homology modeling may further refine the predictions of which amino acids should be converted into which other amino acids. One specific example is provided to illustrate the broader point: conversion of Snapdragon myrcene synthase (SEQ ID NO: 13) into an isoprene synthase by conversion of V331 and/or V444 to phenylalanine (F shown), tyrosine (Y) or tryptophan (W) (not shown) (SEQ ID NOs: 20, 21 and 22).

Conversion of phellandrene synthase into an isoprene producing enzyme

[0098] In various embodiments, the invention provides an enzyme with isoprene synthase activity by providing a variant polypeptide derived from parent phellandrene

synthase polypeptide. The phellandrene synthase is converted, by variation (addition, deletion or substitution) of one or more amino acid residues in the protein. In various embodiments, the amino acid is in the active site. Many monoterpenes are produced by monoterpene synthases from the prenyl diphosphate precursor geranyl diphosphate (GPP) in which two isoprene units are joined in the *trans* (*E*) configuration. The type VI glandular trichomes, small hair-like structures, on the surface of M82 variety of cultivated tomato (*Solanum lycopersicum*) produce an unusual monoterpene synthase that uses neryl diphosphate (NPP) the cis-isomer of GPP as a substrate. This monoterpene synthase is a β-phellandrene synthase (EC 4.2.3.52) called PHS1 (Schilmiller et al., (2009) *PNAS* **106**: 10865-10870). PHS1 has a high k_{cat} producing more than 4 molecules of phellandrene per second using NPP as a substrate. Phellandrene synthase is different from typical cyclic monoterpene synthases in that it does not have the RR at the beginning of the active site. The reason for this is that it does not have to do the linalyl to neryl cation chemical conversion that most cyclic monoterpene synthases have to do. Based on homology to the structure of isoprene synthases, it is converted into an isoprene producing enzyme by either or both of the mutations V524F and M672F (SEQ ID NO: 23).

Pi-Electron Based Approaches

[0099] In an exemplary embodiment, the invention provides a variant polypeptide in which the pi-electron structure of the active site of the relevant parent polypeptide is varied to lower K_m, increase k_{cat} or both, or to convert an enzyme without (or with minimal) isoprene synthase activity into an enzyme with isoprene synthase activity.

[0100] It is postulated that if pi electrons are important to stabilizing the cation intermediate, this could slow the reaction mechanism by making the reaction intermediate too stable. This might be evolutionarily advantageous by keeping the K_m high (with an active site slightly smaller than optimal). This would ensure that isoprene synthase did not interfere with other metabolism needed by plants. However, in an industrial setting a low K_m form of the enzyme has significant advantages. One or the

other of the two site-constructing phenylalanines, shown in bold and underlined in *Populus tremuloides* isoprene synthase (IspS) F338 and F485 (SEQ ID NO: 24), *Populus alba* IspS F338 and F485 (SEQ ID NO: 25), *Populus nigra* IspS F338 and F485 (SEQ ID NO: 26), *Populus trichocarpa* IspS F301 and F448 (SEQ ID NO: 27), *Pueraria Montana* IspS F343 and F493 (SEQ ID NO: 28), *Eucalyptus globulus* IspS F326 and F473 (SEQ ID NO: 29), *Melaleuca alternifolia* F326 and F473 (SEQ ID NO: 30) and *Robinia pseudoaccacia* IspS F281 and F428 (SEQ ID NO: 31), with a tyrosine or tryptophan and the other phenylalanine are replaced with a smaller amino acid, for example leucine, to provide sufficient pi-electron character to the active site without unduly restricting the binding of DMADP. The active site is examined to find other potential locations for a phenylalanine or other pi-electron contributing amino acids in ways that do not constrict the active site. Variation in the location of phenylalanines has been found in unpublished sequences of isoprene synthases from spruce F337 and F541 (*Picea pungens*) (SEQ ID NO: 10) and hops F354 and F530 (*Humulus lupulus*) (SEQ ID NO: 32) and indicating that it should be possible to vary the location of the pi-electron contributing amino acids without destroying enzyme activity.

Conversion of dimethylallyl tryptophan synthase into an isoprene producing enzyme

[0101] In an exemplary embodiment, the present invention provides a variant polypeptide of a parent dimethylallyl tryptophan synthase with isoprene synthase activity. Exemplary enzymes according to this embodiment have one or more amino acid removed, added or substituted relative to the amino acid sequence of the parent.

[0102] A variety of enzymes that use dimethylally diphosphate (DMADP) and a cation intermediate are all good targets for conversion into isoprene synthases. One example is a gene unrelated to known isoprene synthases, dimethylallyl tryptophan synthase. This protein uses DMADP and has a high (poor) K_m (8 μM) and higher k_{cat} (0.37 per second). The structure of this enzyme is known (3I4X, PDB) (Figure 7) and it is known to use a carbocation intermediate of DMADP similar to that in the isoprene synthase reaction (Luk and Tanner, (2009) *J Am Chem Soc* **131**: 13932-13933). Based on knowledge of isoprene synthase evolution it is predicted that adding phenylalanine residues to locations likely to reduce the size of the binding pocket or converting some tyrosine residues that line the pocket with phenylalanine residues have potential for

converting this enzyme into a novel isoprene synthase. The binding pocket residues on dimethylallyl tryptophan synthase are L81, T82, R83, Y191, Y345 and Y398 (SEQ ID NO: 33). One, two, three, four, five or all six of these residues are converted to phenylalanine.

[0103] The families of enzymes that use dimethylally diphosphate (DMADP) and a cation intermediate and can be converted into isoprene synthases to form a variant polypeptide include the following identified by name and enzyme commission (EC) number include but are not limited to: dimethylallyltranstransferases (EC 2.5.1.1), geranyltranstransferases (EC 2.5.1.10), trans-octaprenyltranstransferases (EC 2.5.1.11), adenylate dimethylallytranstransferases (EC 2.5.1.27), dimethylallylcistransferases (EC 2.5.1.28), farnesyltranstransferase (EC 2.5.1.29), trihydroxypterocarpan dimethylallyltransferases (EC 2.5.1.33), dimethylallytryptophan synthase (EC 2.5.1.34), chyrsanthemyl diphosphate synthase (EC 2.5.1.67) and lavandulyl diphosphate synthase (EC 2.5.1.11).

Conversion of pentalenene synthase into an isoprene producing enzyme

[0104] In an exemplary embodiment, the invention provides a variant polypeptide of a parent pentalenene synthase that produces isoprene. The variant is prepared by the addition, deletion or substitution of one or more amino acid relative to the parent polypeptide.

[0105] Pentalenene is a complex tricyclic sesquiterpene (a terpenoid containing 15 carbon atoms), that is produced from farnesyl diphosphate by the action of pentalenene synthase (EC. 4.3.2.7). Pentalenene, is the hydrocarbon precursor of the pentalenolactone family of antibiotics. The crystal structure of pentalene synthase has been resolved (Lesburg et al., (1997) *Science* **277**:1820-1824), revealing the fact that it is unusually a single domain terpenoid synthase. For this reason it may behave much better in bacteria and in other ways be more stable. Pentalenene synthase of *Streptomyces* sp. UC5319 (SEQ ID NO: 34) is converted into an isoprene synthase by converting V179 and/or N219 into a F, W or Y (SEQ ID NOs: 35, 36 and 37).

EXAMPLES

Example 1: Isoprene synthase expression and purification

[0106] Utilizing theoretical amino acid protein sequences in this disclosure, bacterial expression constructs, N-terminal or C-terminal flag tagged, capable of regulated expression of recombinant proteins described in this disclosure are generated. The backbone vector used is the pET vector system (Novagen) or similar. The gene sequences are codon optimized for bacterial expression and synthetically generated to match the amino acid sequences in this disclosure with a Flag tag. The synthetically created gene sequences is then sub-cloned using unique, engineered restriction sites into a commercially-available expression construct (e.g. Novagen) for expression of the protein. Upon transformation into DH5α, six clones for each construct are tested for presence of the insert by restriction digest analysis.

[0107] The expression constructs generated are used to transform *E. coli* BL21(DE3) to generate clones for expression screening. Following the selection of the best expressing clones, induction experiments (3hrs and overnight, 37°C and 18°C) are performed on the selected clones. SDS-PAGE or immunoblot is run to compare protein expression (total and soluble) and the optimal parameters for soluble expression are chosen for production at the 1L scale. The selected clones are grown in LB medium and the cultures are induced at an OD between 0.8 and 1.0. After induction, cell paste is harvested by centrifugation and stored at -20°C until purification. SDS-PAGE or Western blot is performed on an analytical sample to confirm the protein expression.

[0108] Recombinant protein is purified from cell paste using Flag affinity chromatography. Briefly, cells will be lysed using a microfluidics homogenizer and clarified by centrifugation. Clarified lysate is loaded onto a chromatography column containing M2 anti-Flag resin. The resin is washed to remove contaminating proteins and bound protein is competitively eluted with flag peptide according to standard methods. Fractions containing the protein of interest are pooled, dialyzed into PBS + 5% glycerol, concentrated, aliquotted, and stored frozen at -80°C pending studies. All steps of the purification are monitored by SDS-PAGE-Coomassie.

[0109] For the un-tagged recombinant proteins, an alternative separation protocol is used based on ion exchange chromatography (Silver and Fall (1995) *J Biol. Chem.* **270**: 13010-13016). The pI of the recombinant protein is 5.5, and this will be is exploited to

separate the proteins on a Sepharose Q column. Active fractions from the ion exchange column are also concentrated and purified by gel filtration and handled the same as the tagged proteins.

Example 2: Isoprene synthase assays.

[0110] One of the products of the isoprene synthase reaction is pyrophosphate, and a continuous enzyme-coupled assay has been developed for assaying phosphate (Webb (1992) *PNAS* **89**: 4884-4887), and this has been extended to monitor pyrophosphate for enzymatic kinetic parameter estimation (Miller et al. (2007) *J. Bacteriology* **189**: 8196-8205). There are a number of commercially available kits for measuring enzyme kinetics by monitoring pyrophosphate production. These include but are not limited to a PPiLight (Lonza), PiPer Pyrophosphate Assay kit (Invitrogen), as well as EnzCheck (Molecular Probes). These assays are modified for the continuous measurement of isoprene synthase mutant kinetics.

[0111] In order to determine kinetic parameters for the mutant and wild type enzymes, enzymatic reaction rates are needed at varying substrate concentrations. Since this enzyme may display substrate inhibition and/or substrate cooperativeness (Eq. 1, Eq. 2 and together Eq. 3), it will be important to collect appropriate data in order to distinguish between, and quantify the impacts of, these effects. Initial rates of the enzymes are determined with varying initial DMADP concentrations. In order to ensure that only initial rates are estimated, the rates are calculated only during the consumption of the first 10% of the initial DMADP substrate. At least 4 values above and below the K_m of the enzyme are used so that an accurate estimation of the kinetic parameters can be made. The experimental data is fit to the appropriate enzymatic rate equations (Eq. 1, Eq. 2 and together Eq. 3) using a least squares method. If the data matches predictions that both substrate inhibition and positive cooperativeness are present, then Equation 3 will be used.

[0112] The performance of the mutant isoprene synthase proteins is assessed using kinetic analysis described above. In addition, the activity of the isoprene synthase proteins in a more oxidative environment (with a decreased concentration of DTT) is

explored. Once beneficial mutations are identified from the analysis of the mutants, the mutations can be combined to determine if the beneficial impacts are additive. In addition, the mutants are examined with and without a tag, as this has previously been shown to affect the performance of the recombinant isoprene synthase enzymes. It is anticipated that this approach may be able to significantly improve the k_{cat}, K_m, substrate inhibition, and coopertivity issues with the isoprene synthase enzyme. In addition, it may be possible to impact the susceptibility of the enzyme to oxidative stress.

[0113] Further, the activity of the isoprene synthase proteins at a range of different temperatures is explored. Isoprene synthases produce bursts of isoprene in response to transient heat stress. Repeating the assay described above at a range of temperatures enables examination of this phenomenon on isolated proteins *in vitro*.

Confirming the production of isoprene and analysis of other products

[0114] Hemiterpene and other terpenoid synthases may produce more than one product. In addition to the pyrophosphate assay, samples of the product of these enzymes are adsorbed onto a Solid Phase Micro Extraction (SPME) fiber and then the fiber is heated in a GC MS.

Example 3

Example 3.1: Plasmids and Strains

pTrcHis2B-P1IspS

[0115] The *P. alba* isoprene synthase (except the N terminal plastid targeting sequence, amino acids 1 to 36) was codon optimized for *E. coli* expression and synthesized by GenScript. A 1.6 kb BsaI to XhoI fragment was cloned into the NcoI and XhoI sites of pTrcHis2B to yield pTrcHis2B-P1IspS (SEQ ID NO:38).

pTrcHis2B-E2IspS

[0116] The *E. globulus* isoprene synthase (except the N terminal plastid targeting sequence, amino acids 1 to 35) was codon optimized for *E. coli* expression and synthesized by GenScript. A 1.6 kb NcoI to XhoI fragment was cloned into the NcoI and XhoI sites of pTrcHis2B to yield pTrcHis2B-E2IspS (SEQ ID NO:39).

pJexpress401-E4IspS-His

[0117] The *E. globulus* isoprene synthase (except the N terminal plastid targeting sequence, amino acids 1 to 35) with a C terminal linker and 6 histidine motif was codon optimized for *E. coli* expression, synthesized, and cloned into pJexpress401 by DNA2.0 (SEQ ID NO:40, pJExpress401-E4IspS-His).

pCL-SDFi

[0118] Four *E. coli* MEP pathway genes (*dxs*, *ispD*, *ispF*, and *idi*) were placed under control of the pTrc promoter and cloned into the SmaI site of pCL1920 (SEQ ID NO:41, pCL-SDFi).

Example 3.2: Isoprene quantification

[0119] In vivo isoprene production was measured by injecting flask headspace samples into a Tianmei GC9700 GC equipped with a photoionization detector. The column was a PLOT-Q (Tianmei), the column temperature was 140 C, the injector temperature was 150 C, and the detector temperature was 150 C. The injection volume was 1 mL gas. Isoprene standards were used to quantify the signal.

[0120] Samples to measure in vitro production were assayed on a fast isoprene sensor (Hills Scientific) per the manufacturer instructions. Isoprene standards were used to quantify the signal.

Example 3.3: In vivo isoprene production

[0121] To make pre-seed cultures, single colonies from each strain to be tested were used to inoculate 4 mL LB medium plus appropriate antibiotics and grown overnight at 37 C with shaking. To make seed cultures, for each strain, 1 mL of pre-seed was inoculated into 75 mL mineral medium (37.9 mM $(NH_4)_2SO_4$, 8.3 mM $MgSO_4$, 36.8 mM KH_2PO_4, 85.1 mM Na_2PO_4, 1.4 mM $CaCl_2$, 0.007 mM thiamine, 0.476 mM citric acid monohydrate, 0.107 mM $MnSO_4$, 0.018 mM $FeSO_4$, 0.007 mM $ZnSO_4$, 0.002 mM $CuSO_4$, 0.003 mM $CoCl_2$, 0.003 mM NaMoO) plus 10 g/L glucose, 5 g/L yeast extract, and appropriate antibiotics and cultured at 37 C with shaking for 10 hours. For each strain, the seed culture was diluted to OD600 of 0.1 in 30 mL mineral medium plus 5 g/L glucose, 1 g/L yeast extract, appropriate antibiotics, and 0.4 mM IPTG and incubated at 30 C with shaking. Isoprene was measured at each time point then the headspace of each

flask was flushed with a mix of 50% oxygen and 50% air. At 24 and 34 hours, additional glucose was added to each flask to a final concentration of 2 g/L.

Results

[0122] The *E. globulus* isoprene synthase, when expressed in *E. coli* cells, produces more isoprene than the *P. alba* isoprene synthase (Figure 8). *E. coli* BL21 strains expressing either *E. globulus* or *P. alba* isoprene synthases, plus or minus heterologous expression of *dxs*, *ispD*, *ispF*, and *idi*, were grown in minimal medium under aerobic conditions and isoprene production measured. Cumulative culture productivity of isoprene is shown for *E. coli* BL21 expressing heterologous MEP genes and *E. globulus* isoprene synthase (inverted triangles), heterologous MEP genes and *P. alba* isoprene synthase (circles), *E. globulus* isoprene synthase (triangles), and *P. alba* isoprene synthase (squares).

Example 3.4: *E. globulus* isoprene synthase protein purification

[0123] Overnight seed culture of FM5 cells containing the pJexpress401-E4IspS-His plasmid were diluted 1:100 into 500 mL LB cultures and grown to OD600 ~0.9 at 37 C. IPTG was added to a final concentration of 1 mM and induction occurred for ~4 hours at room temperature. Cultures were centrifuged and cell pellets frozen.

[0124] Frozen pellets were resuspended in lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0 with Roche Mini EDTA-free tablets per the manufacturer's instructions) and cells disrupted by sonication. Cell lysates were treated with RNAse A and DNAse I (Qiagen) and cleared by centrifugation. Cleared lysates were incubated with nickel NTA resin for 1 hour at 4 C then loaded into a disposable chromatography column. The column was washed with 5 times the bed volume with wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) then eluted with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). Glycerol was added to 15% and the samples flash frozen in liquid nitrogen and stored at -80 C.

Example 3.5: In vitro protein kinetics

[0125] Purified isoprene synthase was assayed for kinetic properties using in vitro conversion of DMADP to isoprene. For each of six different concentrations of DMADP

(0, 0.25, 0.5, 1, 2, and 4 mM), 75 uL of appropriately diluted DMADP solution was mixed with 73 uL reaction buffer (100 mM HEPES, 40 mM KCl, 20 mM $MgCl_2$, 10% glycerol, pH 7.8), 1 uL 1 mM DTT, and 2.5 uL of the purified isoprene synthase at 0.3 ug/mL. The reaction was allowed to occur in a sealed tube at 37 C for 12 minutes, after which 1 mL of headspace was injected into a FIS for measurement of the isoprene produced. Controls with no added protein were used as appropriate to subtract non-specific isoprene production. Specific activity measurements were fitted to Equation 1 to determine K_M, k_{cat}, and K_{IS}.

Equation 1: Enzymatic rate equation for enzyme with substrate inhibition.

$$Rate = \frac{E_t k_{cat} S}{K_M + S + \frac{S^2}{K_{IS}}}$$

Results

[0126] The K_M for *E. globulus* isoprene synthase is lower (more preferable for production of isoprene) than is reported for other isoprene synthases. Specific activity was obtained at various DMADP concentrations for *E. globulus* isoprene synthase (Figure 9, light grey diamonds). The data were fitted to Equation 1 using the constants shown in Table 1 and the fitted equation is shown (Figure 9, dark grey line).

Table 1: Kinetic parameters of purified *E. globulus* isoprene synthase.

Kinetic parameter	Value
K_M	0.03 mM
k_{cat}	0.03 s^{-1}
K_{IS}	9.33 mM

[0127] The published K_M values for various non-*Eucalyptus* isoprene synthases range from 0.3 mM to 9 mM (Rasulov, et al., (2009) Plant Physiology **149**: 1609-1618 and references therein and Sasaki, et al., (2005) FEBS Letters **579**: 2514-2518). The K_M value of 0.03 mM for *E. globulus* favorably compares to that of other published isoprene synthases, such as those from *P. alba* and other *Poplar* species, and thus represents a superior isoprene synthase for the commercial production of isoprene.

Example 4

Example 4.1: Plasmids and Strains

pTrc-E4-6His

[0128] The entire coding sequence of the *E. globulus* isoprene synthase (except the N terminal plastid targeting sequence, amino acids 1 to 35) in pJexpress401-E4IspS-His was transferred to the NcoI/XhoI site of pTrcHis2B using PCR to introduce an NcoI site at the N terminus.

E. globulus active site mutant IspS

[0129] Sequences identical to pJexpress401-E4IspS-His except the indicated amino acid mutations were synthesized and cloned into pJexpress404 by DNA2.0.

pTrc-M1-6His

[0130] The *M. alternifolia* isoprene synthase (SEQ ID NO 7, except the N terminal plastid targeting sequence, amino acids 1 to 32) with a C terminal linker and 6 histidine motif (ENLYFQSGSGSGSGHHHHHH) was codon optimized for *E. coli* expression and synthesized by DNA2.0. A BsmBI/XhoI fragment containing the isoprene synthase was then ligated into an NcoI/XhoI digest of pTrcHis2B to yield pTrc-M1-6His.

pTrc-R1-6His

[0131] The *R. pseudoaccacia* isoprene synthase (SEQ ID NO 8) with a C terminal linker and 6 histidine motif (ENLYFQSGSGSGSGHHHHHH) was codon optimized for *E. coli* expression and synthesized by DNA2.0. An NcoI/XhoI fragment containing the isoprene synthase was then ligated into an NcoI/XhoI digest of pTrcHis2B to yield pTrc-R1-6His.

Myrcene synthase

[0132] The *A. majus* myrcene synthase (SEQ ID NO 13, except the N terminal plastid targeting sequence, amino acids 1 to 45) with a C terminal linker and 6 histidine motif (ENLYFQSGSGSGSGHHHHHH) was codon optimized for *E. coli* expression, synthesized, and cloned into pJexpress404 by DNA2.0. Variants were made with the indicated amino acid substitutions.

Dimethylallyl tryptophan synthase

[0133] The *A. fumigatus* dimethylallyl tryptophan synthase (SEQ ID NO 33) with a C terminal linker and 6 histidine motif (ENLYFQSGSGSGSGHHHHHH) was codon optimized for *E. coli* expression, synthesized, and cloned into pJexpress404 by DNA2.0. Variants were made with the indicated amino acid substitutions, with "all" representing L81F, T82F, R83F, Y191F, Y345F, and Y398F.

Phellandrene synthase

[0134] The *S. lycopersicum* phellandrene synthase (SEQ ID NO 23, except the N terminal plastid targeting sequence, amino acids 1 to 36) with a C terminal linker and 6 histidine motif (ENLYFQSGSGSGSGHHHHHH) was codon optimized for *E. coli* expression, synthesized, and cloned into pJexpress404 by DNA2.0. Variants were made with the indicated amino acid substitutions.

MBO synthase

[0135] The *P. sabiniana* methyl butanol synthase (SEQ ID NO 9, except the N terminal plastid targeting sequence, amino acids 1 to 43) with a C terminal linker and 6 histidine motif (ENLYFQSGSGSGSGHHHHHH) was codon optimized for *E. coli* expression, synthesized, and cloned into pJexpress404 by DNA2.0. Variants were made with the indicated amino acid substitutions.

Pentalenene synthase

[0136] The *Streptomyces* sp. UC5319 pentalenene synthase (SEQ ID NO 34) with a C terminal linker and 6 histidine motif (ENLYFQSGSGSGSGHHHHHH) was codon optimized for *E. coli* expression, synthesized, and cloned into pJexpress401 by DNA2.0. Variants were made with the indicated amino acid substitutions.

pA1

[0137] Five *B. subtilis* MEP pathway genes (*dxs*, *dxr*, *ispD*, *ispF*, and *idi*) were placed under the control of strong *E. coli* promoters and cloned into the HindIII/SacI sites of pCL1920.

Example 4.2: Isoprene quantification

[0138] In vivo isoprene production was measured by injecting flask headspace samples into a Tianmei GC9700 GC equipped with a photoionization detector. The column was a PLOT-Q (Tianmei), the column temperature was 140 C, the injector temperature was 150 C, and the detector temperature was 150 C. The injection volume was 1 mL gas. Isoprene standards were used to quantify the signal.

Example 4.3: In vivo isoprene production

[0139] To make pre-seed cultures, single colonies from each strain to be tested were used to inoculate 4 mL LB medium plus appropriate antibiotics and grown overnight at 37 C with shaking. To make seed cultures, for each strain, 300 uL of pre-seed was inoculated into 30 mL MOPS mineral medium (37.9 mM $(NH_4)_2SO_4$, 4.2 mM $MgSO_4$, 2.9 mM KH_2PO_4, 7.1 mM Na_2PO_4, 0.7 mM $CaCl_2$, 100 mM 3-(N-morpholino) propanesulfoic acid, 0.2 mg/L thiamine, 0.476 mM citric acid monohydrate, 0.107 mM $MnSO_4$, 0.018 mM $FeSO_4$, 0.007 mM $ZnSO_4$, 0.002 mM $CuSO_4$, 0.003 mM $CoCl_2$, 0.003 mM NaMoO) plus 10 g/L glucose, 5 g/L yeast extract, and appropriate antibiotics and cultured at 37 C with shaking for 6.5 hours. For each strain, 0.5 mL seed culture was diluted into 9.5 mL fresh MOPS mineral medium plus 5 g/L glucose, 1 g/L yeast extract, appropriate antibiotics, and 1 mM IPTG and incubated at 37 C in an air tight flask with shaking. Isoprene in the headspace was measured after 16 hours.

Results

[0140] The *E. globulus* isoprene synthase, when expressed in *E. coli* cells, produces far more isoprene than the *M. alterniforia* or *R. pseudoacacia* isoprene synthase, or any of the non-isoprene synthases with variants predicted to increase isoprene production. *E. coli* FM5 strains expressing the indicated terpenoid synthase and heterologous expression of *dxs*, *dxr*, *ispD*, *ispF*, and *idi*, were grown in minimal medium under aerobic conditions and isoprene production measured (Table 2).

Table 2: *E. globulus* isoprene synthase is superior to other enzymes.

Gene Name	Average Culture Productivity of Isoprene (mg/L)	Standard Error of the Mean
E. globulus wild type IspS pJexpress-E4-his	1.94	0.109

Gene Name	Average Culture Productivity of Isoprene (mg/L)	Standard Error of the Mean
pTrc-E4-his	1.54	0.060
E. globulus active site mutant IspS		
Euc_F326L_F473W	0.02	0.002
Euc_F326L_F473Y	0.14	0.004
Euc_F326W_F473L	0.08	0.003
Euc_F326Y_F473L	0.03	0.006
Non E. globulus wild type IspS		
pTrc-M1-6His	0.03	0.006
pTrc-R1-6His	0.03	0.013
Myrcene synthase		
McS_WT	0.01	0.003
McS_V331F	0.01	0.001
McS_V444F	0.01	0.002
McS_V331F_V444F	0.01	0.000
Dimethylallyl tryptophan synthase		
fgaPT2_WT	0.01	0.002
fgaPT2_L81F	0.00	0.003
fgaPT2_T82F	0.01	0.005
fgaPT2_R83F	0.01	0.001
fgaPT2_Y191F	0.01	0.002
fgaPT2_Y345F	0.01	0.005
fgaPT2_Y398F	0.01	0.001
fgaPT2_all	0.02	0.010
Phellandrene synthase		
PhS_WT	0.02	0.007
PhS_V524F	0.02	0.007
PhS_M672F	0.01	0.001
PhS_V524F_M672F	0.01	0.003
MBO synthase		
MBO_WT	0.01	0.003
MBO_S440F	0.01	0.001
MBO_S440Y	0.02	0.005
Pentalenene synthase		
PS_WT	0.01	0.003
PS_I177F	0.02	0.003
PS_N219F	0.01	0.005
PS_M73F_F77V_F78T_I177F	0.02	0.011

Example 4.4: In vivo expression and solubility

[0141] Cells were grown and induced as above in MOPS mineral medium then 15 OD600 units of cells were pelleted by centrifugation. Cell extracts were prepared by cell lysis in a French press, and total protein samples saved. The remainder was subjected to centrifugation at 14,000 g for 10 minutes to separate soluble and insoluble proteins. Protein samples were electrophoresed on SDS-PAGE gels with Life Technologies Benchmark His-tagged protein standard and transferred to PVDF membranes. Proteins were visualized with anti-his antibodies and BCIP/NBT per the manufacturer's instructions.

Results

[0142] The expression and/or solubility of the *E. globulus* isoprene synthase is far superior to that of *M. alterniforia* or *R. pseudoacacia* isoprene synthases. Total, soluble, and insoluble protein fractions are shown after induction of the indicated plasmids, and the *E. globulus* isoprene synthase was shown to express at high level in soluble form, unlike the *M. alterniforia* or *R. pseudoacacia* isoprene synthases (Figure 10).

Example 5

Example 5.1: Protein Purification

[0134] Commercially obtained E. coli strain FM5-Alper (also called strain Q) were transformed with the pJexpress404 vector (DNA 2.0) containing sequence for a histidine-tagged phellandrene synthase (PhS) or a modified version (Mod A) of the His-tagged PhS. Both sequences were behind a T5 promoter. A resulting E. coli transformant containing each PhS sequence was grown in a 250 ml flask of LB overnight atroom temperature (OD ≈ 1). Cultures were induced by adding enough IPTG to bring the culture concentration to 400 μM. Cultures were allowed to grow an additional eight hours until they reached an OD ≈ 4. Cultures were then concentrated by centrifugation and cells were broken using a combination of freezing, sonication, and lysozyme. Histidine-tagged protein was purified on Ni-agarose (Qiagen) according to manufacture's directions. Protein elution from the columns was determined by image analysis of Coomassie-stained protein gels and protein concentration used in the assays were determined by a Lowry assay.

Example 5.2: Enzyme Assays

[0134] Dimethylallyl diphosphate (DMADP) from either Echelon Biosciences or Isoprenoids LLC was assayed to determine the exact concentration of DMADP(a necessary step because DMADP is labile). Assays were carried out in a total volume of 200 µL of assay buffer (100 mM Hepes pH 7.8, 40 mM KCl, 20 mM MgCl2 10% glycerol). DMADP stock or an equivalent amount of 2 mM NH4HCO3 (the buffer used to dilute the DMADP) was used to establish different concentrations ofDADP in the assay. There was 20 µg protein in each sample. Three additional vials were assayed that had all of these components except the protein to determine the non-enzymatic conversion of DMADP in the system (1 to 3 %). The nonenzymatic signal was subtracted from the signal for the sample with protein. The vial was 2 mL. Vials were incubated for 1 hr at 37°C. At the end of 1 hr, 1 mL of head-space air was withdrawn by syringe while simultaneously adding 1 mL water to avoid creating a vacuum. The head-space air was injected into a chemiluminescent isoprene detection system (Fast Isoprene Sensor, Hills Scientific) that had a flowing gas stream. Isoprene standards were drawn from a tank of compressed nitrogen with 3.25 ± 0.16 PPM isoprene (Airgas). The isoprene signal from each injection lasted about 15 seconds. All of the FISsignal for 30 seconds centered on the injection peak were summed and 15 seconds of baseline signal before and after the peak were summed and subtracted from the injection signal. In all cases n=3 except in two cases where a large outlier datapoint was dropped (this did not change the value of the kineticparameters). Activities were calculated in terms of moles of isoprene per mole of enzyme per second (specific activity). The Vmax estimated from Michaelis-Menton kinetics is the effective kcat when data are plotted this way. See Figure 11.

[0143] While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications may be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications

are intended to be within the scope of the claims appended hereto. All publications and patents cited herein are incorporated by reference in their entirety for all purposes.

SEQUENCES

Legend: The italicized amino acids represent leader sequences that are optionally completely or partially deleted or substituted. The bold amino acids indicate exemplary sites for modification, singly or in combination. Said modification encompasses substitution, deletion and insertion. Substitutions are indicated using the notation [original residue in one-letter-code][position number][substitution residue in one-letter-code]. For example, the notation S440F indicates that a serine residue at position 440 is being replaced with a phenylalanine residue. Position numbers are counted from the N-terminus of the sequences as represented in the instant specification.

SEQ ID NO 1

Populus tremuloides isoprene synthase
Adapted from Genbank: AY341431.1

MATELLCLHR PISLTHKLFR NPLPKVIQAT PLTLKLRCSV STENVSFSET
ETETRRSANY EPNSWDYDYL LSSDTDESIE VHKDKAKKLE AEVRREINNE
KAEFLTLLEL IDNVQRLGLG YRFESDIRRA LDRFVSSGGF DGVTKTSLHG
TALSFRLLRQ HGFEVSQEAF SGFKDQNGNF LENLKEDIKA ILSLYEASFL

ALEGENILDE AKVFAISHLK ELSEEKIGKE LAEQVSHALE LPLHRRTQRL
EAVWSIEAYR KKEDANQVLL ELAILDYNMI QSVYQRDLRE TSRWWRRVGL
ATKLHFARDR LIESFYWAVG VAFEPQYSDC RNSVAKMFSF VTIIDDIYDV
YGTLDELELF TDAVERWDVN AINDLPDYMK LCFLALYNTI NEIAYDNLKD
KGENILPYLT KAWADLCNAF LQEAKWLYNK STPTFDDYFG NAWKSSSGPL
QLIFAYFAVV QNIKKEEIEN LQKYHDIISR PSHIFRLCND LASASAEIAR
GETANSVSCY MRTKGISEEL ATESVMNLID ETWKKMNKEK LGGSLFAKPF
VETAINLARQ SHCTYHNGDA HTSPDELTRK RVLSVITEPI LPFER

SEQ ID NO 2

Populus alba Isoprene synthase
Adapted from Genbank: AB198180.1

MATELLCLHR PISLTHKLFR NPLPKVIQAT PLTLKLRCSV STENVSFTET
ETEARRSANY EPNSWDYDYL LSSDTDESIE VYKDKAKKLE AEVRREINNE
KAEFLTLLEL IDNVQRLGLG YRFESDIRGA LDRFVSSGGF DAVTKTSLHG
TALSFRLLRQ HGFEVSQEAF SGFKDQNGNF LENLKEDIKA ILSLYEASFL
ALEGENILDE AKVFAISHLK ELSEEKIGKE LAEQVNHALE LPLHRRTQRL
EAVWSIEAYR KKEDANQVLL ELAILDYNMI QSVYQRDLRE TSRWWRRVGL
ATKLHFARDR LIESFYWAVG VAFEPQYSDC RNSVAKMFSF VTIIDDIYDV
YGTLDELELF TDAVERWDVN AINDLPDYMK LCFLALYNTI NEIAYDNLKD
KGENILPYLT KAWADLCNAF LQEAKWLYNK STPTFDDYFG NAWKSSSGPL
QLVFAYFAVV QNIKKEEIEN LQKYHDTISR PSHIFRLCND LASASAEIAR
GETANSVSCY MRTKGISEEL ATESVMNLID ETWKKMNKEK LGGSLFAKPF
VETAINLARQ SHCTYHNGDA HTSPDELTRK RVLSVITEPI LPFER

SEQ ID NO 3

Populus nigra Isoprene synthase
Adapted from Genbank: HQ684728.1

MATELLCLHR PISLTHKLFR NPLPKVIQAT PLTLKLRCSV STENVSFTET
ETETRRSANY EPNSWDYDYL LSSDTDESIE VYKDKAKKLE AEVRREINNE
KAEFLTLPEL IDNVQRLGLG YRFESDIRRA LDRFVSSGGF DAVTKTSLHA
TALSFRLLRQ HGFEVSQEAF SGFKDQNGNF LKNLKEDIKA ILSLYEASFL
ALEGENILDE AKVFAISHLK ELSEEKIGKD LAEQVNHALE LPLHRRTQRL
EAVWSIEAYR KKEDADQVLL ELAILDYNMI QSVYQRDLRE TSRWWRRVGL
ATKLHFARDR LIESFYWAVG VAFEPQYSDC RNSVAKMFSF VTIIDDIYDV
YGTLDELELF TDAVERWDVN AIDDLPDYMK LCFLALYNTI NEIAYDNLKD
KGENILPYLT KAWADLCNAF LQEAKWLYNK STPTFDEYFG NAWKSSSGPL
QLVFAYFAVV QNIKKEEIDN LQKYHDIISR PSHIFRLCND LASASAEIAR
GETANSVSCY MRTKGISEEL ATESVMNLID ETWKKMNKEK LGGSLFAKPF
VETAINLARQ SHCTYHNGDA HTSPDELTRK RVLSVITEPI LPFER

SEQ ID NO 4

Populus trichocarpa Isoprene synthase

Adapted from Genbank: EU693027.1

CSVSTENVSF TETETETRRS ANYEPNSWDY DYLLSSDTDE SIEVYKDKAK
KLEAEVRREI NNEKAEFLTL LELIDNVQRL GLGYRFESDI RRALDRFVSS
GGFDAVTKTS LHATALSFRL LRQHGFEVSQ EAFSGFKDQN GNFLENLKED
IKAILSLYEA SFLALEGENI LDEAKVFAIS HLKELSEEKI GKDLAEQVNH
ALELPLHRRT QRLEAVLSIE AYRKKEDADQ VLLELAILDY NMIQSVYQRD
LRETSRWWRR VGLATKLHFA RDRLIESFYW AVGVAFEPQY SDCRNSVAKM
FSFVTIIDDI YDVYGTLDEL ELFTNAVERW DVN**A**IDDLPD YMKLCFLALY
NTINEIAYDN LKEKGENILP YLTKAWADLC NAFLQEAKWL YNKSTPTFDD
YFGNAWK**SSS** GPLQLVFAYF AVVQNIKKEE IENLQKYHDI ISRPSHI**F**RL
C**N**DLASASAE IARGET**AN**SV SCYMRTKGIS EELATESVMN LIDETWKKMN
KEKLGGSLFA KPFVETAINL AR**QS**H**C**TYHN GDAHTSPDEL TRKRVLSVIT
EPILPFER

SEQ ID NO 5

Pueraria Montana Isoprene synthase
Adapted from Genbank: AY316691.1

MATNLLCLSN KLSSPTPTPS TRFPQSKNFI TQKTSLANPK PWRVICATSS
*QFTQITE*HNS RRSANYQPNL WNFEFLQSLE NDLKVEKLEE KATKLEEEVR
CMINRVDTQP LSLLELIDDV QRLGLTYKFE KDIIKALENI VLLDENKKNK
SDLHATALSF RLLRQHGFEV SQDVFERFKD KEGGFSGELK GDVQGLLSLY
EASYLGFEGE NLLEEARTFS ITHLKNNLKE GINTKVAEQV SHALELPYHQ
RLHRLEARWF LDKYEPKEPH HQLLLELAKL DFNMVQTLHQ KELQDLSRWW
TEMGLASKLD FVRDRLMEVY FWALGMAPDP QFGECRKAVT KM**F**GLVTIID
DVYDVYGTLD ELQLFTDAVE RWDVN**A**INTL PDYMKLCFLA LYNTVNDTSY
SILKEKGHNN LSYLTKSWRE LCKAFLQEAK WSNNKIIPAF SKYLENASV**S**
SSGVALLAPS YFSVCQQQED ISDHALRSLT DFHGLVRSSC VI**F**RLC**N**DLA
TSAAELERGE T**TN**SIISYMH ENDGTSEEQA REELRKLIDA EWKKMNRERV
SDSTLLPKAF MEIAVNMAR**V** **S**H**C**TYQYGDG LGRPDYATEN RIKLLLIDPF
PINQLMYV

SEQ ID NO 6

Eucalyptus globulus "mts-1 mRNA for monoterpene synthase" actually an isoprene synthase
Adapted from Genbank: AB266390.1

*MALRLLFTPH LPVLSSRRAN GRVRCSASTQ ISD*PQEGRRS ANYQPSVWTY
NYLQSIVAGE GRQSRREVEQ QKEKVQILEE EVRGALNDEK AETFTIFATV
DDIQRLGLGD HFEEDISNAL RRCVSKGAVF MSLQKSLHGT ALGFRLLRQH
GYEVSQDVFK IFLDESGSFV KTLGGDVQGV LSLYEASHLA FEEEHILHKA
RSFAIKHLEN LNSDVDKDLQ DQVKHELELP LHRRMPLLEA RRSIEAYSRR
GYTNPQILEL ALTDFNVSQS YLQRDLQEML GWWNNTGLAK RLSFARDRLI
ECFFWAVGIA HEPSLSICRK AVTKA**F**ALIL VLDDVYDVFG TLEELELFTD
AVRRWDLN**A**V EDLPVYMKLC YLALYNSVNE MAYETLKEKG ENVIPYLAKA

WYDLCKAFLQ EAKWSNSRII PGVEEYLNNG WV**SSS**GSVML IHAYFLASPS
IRKEELESLE HYHDLLRLPS LI**F**RLT**N**DIA SSSAELERGE T**TN**SIRCFMQ
EKGISELEAR ECVKEEIDTA WKKMNKYMVD RSTFNQSFVR MTYNLAR**MA**H
CVYQDGDAIG SPDDLSWNRV HSLIIKPISP AA

SEQ ID NO 7

Melaleuca alternifolia "putative monoterpene synthase mRNA", actually an isoprene synthase
Adapted from GenBank: AY279379.1

*MALRLLSTPH LPQLCSRRVS GRVHCSASTQ VS*DAQGGRRS ANYQPSVWTY
NYLQSLVADD IRRSRREVEQ EREKAQILEE DVRGALNDGN AEPMAIFALV
DDIQRLGLGR YFEEDISKAL RRCLSQYAVT GSLQKSLHGT ALSFRVLRQH
GFEVSQDVFK IFMDESGSFM KTLGGDVQGM LSLYEASHLA FEEEDILHKA
KTFAIKHLEN LNHDIDQDLQ DHVNHELELP LHRRMPLLEA RRFIEAYSRR
SNVNPRILEL AVMKFNSSQL TLQRDLQDML GWWNNVGLAK RLSFARDRLM
ECFFWAVGIA REPALSNCRK GVTKA**F**SLIL VLDDVYDVFG TLDELELFTD
AVRRWHED**A**V ENLPGYMKLC FLALYNSVND MAYETLKETG ENVTPYLTKV
WYDLCKAFLQ EAKWSYNKIT PGVEEYLNNG WV**SSS**GQVML THAYFLSSPS
LRKEELESLE HYHDLLRLPS LI**F**RLT**N**DLA TSSAELGRGE T**TN**SILCYMR
EKGFSESEAR KQVIEQIDTA WRQMNKYMVD HSTFNRSFMQ MTYNLAR**MA**H
CVYQDGDAIG APDDQSWNRV HSLIIKPVSL APC

SEQ ID NO 8
Robinia pseudoaccacia isoprene synthase
Unpublished sequence

RRSANYQPNL WNFEFLQSQE YDLMVETLQE RATKLEEEVR RLINRVDIEP
LKLLELVDNV QRLGLTYKFE DDINKALERI VSLDEREKSG LHATALIFRL
LRQHGFEVSQ DVFESTRDKE GRFKAEIKGD VQGLLSLYEA SYLGSEGENL
LDEAREFSMT HLKNLNEGVV TPKLAEQINH ALELPYHRRF QRLEARWFIE
NYEVKEPHDR LLVELAKLDF NMVQSLQKKE VGESSRWWKE IGLTSKLDFV
RDRLVEVYFW ASGMAPDPQL SECRKAVTKM **F**GLVTIIDDV YDVYGTLDEL
ELFTNAVERW DVN**A**VDTLPD YMKLCFFALY NTVNDTAYNL LKEKGDNNLP
YLAKSWSDLC KAFLQEAKWS NNKIIPSFNK YIENASV**SSS** GGALLTPCYF
SIRQDITNQA LDSLTNYHGP VRSSCAI**F**RL **C**NDLATSAAE LERGET**TN**SI
TSCMQDNGIS EEQARDELRN LIDAEWKQMN RERVFDQTFP KAFIETAINM
AR**VS**H**C**TYQY GDGLGRPDNT AENRIKLLLI DPLSN

SEQ ID NO 9

Pinus sabiniana Methyl butonal synthase
Unpublished sequence with S440 shown in bold and highlighted

*MALLSVAPLA PRWCVHKSLV TSTKVKVVRR TISTSIRMCR ITT*ESGEGVQ
RRIANHHSNL WDDNFIQSLS TPYGAISYHE SAQKLIGEVK EMINSISLKD
GELITPSNDL LMRLSIVDSI ERLGIDRHFK SEIKSALDYV YSYWNEKGIG

WGRDSVVADL NSTALGLRTL RLHGYPVSSD VLQHFKEQKG QFACSAIQTE
GEIRSVLNLF RASQIAFPGE KVMEEAEVFS TIYLKEAILK LPVCGLSREI
SYVLEYGWHI NLPRLEARNY IDVFGEDPIY LTPNMKTQKL LELAKLEFNM
FHSLQQQELK LLSRWWKDSG FSQMTFPRHR HVEYYTLASC IDSEPQHSSF
RLGFAKIFHL ATVLDDIYDT FGTMDELELF TAAVKRWHPS ATEWLPEYMK
GVYMVLYETV NEMAGEAEKS QGRDTLNYGR NALEAYIDAS MEEAKWIFSG
FLPTFEEYLD NGKVSFGYGI GTLQPILTLG IPFPHHILQE IDFPSRLNDV
ASSILRLKGD IHTYQAERSR GEKSSCISCY MEENPESTEE DAINHINSMV
DKLLKELNWE YLRPDSNVPI TSKKHAFDIL RAFYHLYKYR DGFSVANYEI
KNLVMTTVIE PVPL

SEQ ID NO 10

Picea pungens isoprene synthase (Spruce)
Unpublished sequence

MALLSVAPLV STWCVDKPLV GSSEAKALLR KIPTLEMCRL TKSVTPSISM
*CLTTTVS*DDG VQRRIANHHS NLWDDNFIQS LSTPYGATAY HERAQKLIGE
VKVIINSILV EDGELITAPN DLLQRLSIVD SIERLGIDRH FKNEIKSALD
YVYSYWNEKG IGCGRDSVVN DLNTTALGLR TLRLHGYPVS SDVLEQFKDQ
NGQFACSAIQ TEGEIKKVLN LFRASLIAFP GEKVMEEAEI FSTIYLKEAL
LKIPVCSFSR EIAYVLEYGW HMNLPRLEAR NYIDVFGQDA IYLTPNMRTQ
KILELAKLEF NIFHSLQQKE LKHLSRWWKD SVFSQLTFPR HRHVEYYTLA
SCIDIDPQHS SFRLGFAKI**F** HLATVLDDIY DTFGMMDELE LFTAAVKRWH
PS**A**AEWLPEY MKGVYMMLYE TVYEMAREAE KSQGRDTLNY ARQALEAYID
SYMKEAKWIS SGFLPTFEEY LDNGKV**SFG**Y RIGTLQPILT LGIPFPHHIL
QEIDFPSRLN DLAGSILRLK GDIHSYQAER SRGEESSGIS CYMKDNPEST
EEDAVTYINA MVNRLLKELN WELLKPHSNV PITSKKHAFD ILRA**F**YHLYK
DRDGFSVTRN EIRNLVMTTV IEPVPL

SEQ ID NO 11

Pinus sabiniana Methyl butonal synthase converted into an isoprene synthase
Route 1: making the change S440F, shown in bold and highlighted

*MALLSVAPLA PRWCVHKSLV TSTKVKVVRR TISTSIRMCR ITTE*SGEGVQ
RRIANHHSNL WDDNFIQSLS TPYGAISYHE SAQKLIGEVK EMINSISLKD
GELITPSNDL LMRLSIVDSI ERLGIDRHFK SEIKSALDYV YSYWNEKGIG
WGRDSVVADL NSTALGLRTL RLHGYPVSSD VLQHFKEQKG QFACSAIQTE
GEIRSVLNLF RASQIAFPGE KVMEEAEVFS TIYLKEAILK LPVCGLSREI
SYVLEYGWHI NLPRLEARNY IDVFGEDPIY LTPNMKTQKL LELAKLEFNM
FHSLQQQELK LLSRWWKDSG FSQMTFPRHR HVEYYTLASC IDSEPQHSSF
RLGFAKIFHL ATVLDDIYDT FGTMDELELF TAAVKRWHPS ATEWLPEYMK
GVYMVLYETV NEMAGEAEKS QGRDTLNYGR NALEAYIDA**F** MEEAKWIFSG
FLPTFEEYLD NGKVSFGYGI GTLQPILTLG IPFPHHILQE IDFPSRLNDV
ASSILRLKGD IHTYQAERSR GEKSSCISCY MEENPESTEE DAINHINSMV
DKLLKELNWE YLRPDSNVPI TSKKHAFDIL RAFYHLYKYR DGFSVANYEI
KNLVMTTVIE PVPL

SEQ ID NO 12

Pinus sabiniana Methyl butonal synthase converted into an isoprene synthase
Route 1: making the change S440Y, shown in bold and highlighted

*MALLSVAPLA PRWCVHKSLV TSTKVKVVRR TISTSIRMCR ITT*ESGEGVQ
RRIANHHSNL WDDNFIQSLS TPYGAISYHE SAQKLIGEVK EMINSISLKD
GELITPSNDL LMRLSIVDSI ERLGIDRHFK SEIKSALDYV YSYWNEKGIG
WGRDSVVADL NSTALGLRTL RLHGYPVSSD VLQHFKEQKG QFACSAIQTE
GEIRSVLNLF RASQIAFPGE KVMEEAEVFS TIYLKEAILK LPVCGLSREI
SYVLEYGWHI NLPRLEARNY IDVFGEDPIY LTPNMKTQKL LELAKLEFNM
FHSLQQQELK LLSRWWKDSG FSQMTFPRHR HVEYYTLASC IDSEPQHSSF
RLGFAKIFHL ATVLDDIYDT FGTMDELELF TAAVKRWHPS ATEWLPEYMK
GVYMVLYETV NEMAGEAEKS QGRDTLNYGR NALEAYIDA**Y** MEEAKWIFSG
FLPTFEEYLD NGKVSFGYGI GTLQPILTLG IPFPHHILQE IDFPSRLNDV
ASSILRLKGD IHTYQAERSR GEKSSCISCY MEENPESTEE DAINHINSMV
DKLLKELNWE YLRPDSNVPI TSKKHAFDIL RAFYHLYKYR DGFSVANYEI
KNLVMTTVIE PVPL

SEQ ID NO 13

Snapdragon (*Antirrhinum majus*) myrcene synthase
Adapted from Genbank: AY195608.1
Amino acids to be transferred bold and highlighted

*MAFCISYVGA LLPCSLSTRT KFAICHNTSK LHRAAYKTSR WNIPG***DVGST**
PPPSKLHQAL CLNEHSLSCM AELPMDYEGK IKETRHLLHL KGENDPIESL
IFVDATLRLG VNHHFQKEIE EILRKSYATM KSPIICEYHT LHEVSLFFRL
MRQHGRYVSA DVFNNFKGES GRFKEELKRD TRGLVELYEA AQLSFEGERI
LDEAENFSRQ ILHGNLAGME DNLRRSVGNK LRYPFHTSIA RFTGRNYDDD
LGGMYEWGKT LRELALMDLQ VERSVYQEEL LQVSKWWNEL GLYKKLNLAR
NRPFEFYTWS MVILADYINL SEQRVELTKS VAFIYLIDDI FDVYGTLDEL
IIFTEAVNKW DYSATDTLPE NMKMCCMTLL DTINGTSQKI YEKHGYNPID
SLKTTWKSLC SAFLVEAKWS ASGSLPSANE YLENEKVSSG VYVVLVHLFC
LMGLGGTSRG SIELNDTQEL MSSIAIIFRL WNDLGSAKNE HQNGKDGSYL
NCYKKEHINL TAAQAHEHAL ELVAIEWKRL NKESFNLNHD SVSSFKQAAL
NLARMVPLMY SYDHNQRGPV LEEYVKFMLS D

SEQ ID NO 14 (the hybrid)

Hybrid between *Populus alba* Isoprene synthase (Genbank: AB198180.1) deleted up to L148 and the structured arm of Snapdragon (*Antirrhinum majus*) myrcene synthase

DVGSTPPPSK LHQALCLNEH SLSCMAELPM DYEGKIKETR HLLHLKGEND
PIESLHGTAL SFRLLRQHGF EVSQEAFSGF KDQNGNFLEN LKEDIKAILS
LYEASFLALE GENILDEAKV FAISHLKELS EEKIGKELAE QVNHALELPL
HRRTQRLEAV WSIEAYRKKE DANQVLLELA ILDYNMIQSV YQRDLRETSR

WWRRVGLATK LHFARDRLIE SFYWAVGVAF EPQYSDCRNS VAKMFSFVTI
IDDIYDVYGT LDELELFTDA VERWDVNAIN DLPDYMKLCF LALYNTINEI
AYDNLKDKGE NILPYLTKAW ADLCNAFLQE AKWLYNKSTP TFDDYFGNAW
KSSSGPLQLV FAYFAVVQNI KKEEIENLQK YHDTISRPSH IFRLCNDLAS
ASAEIARGET ANSVSCYMRT KGISEELATE SVMNLIDETW KKMNKEKLGG
SLFAKPFVET AINLARQSHC TYHNGDAHTS PDELTRKRVL SVITEPILPF
ER

SEQ ID NO 15

Hybrid between *Pueraria Montana* Isoprene synthase (Genbank: AY316691.1) deleted up to L153 and the structured arm of Snapdragon (*Antirrhinum majus*) myrcene synthase

DVGSTPPPSK LHQALCLNEH SLSCMAELPM DYEGKIKETR HLLHLKGEND
PIESLHATAL SFRLLRQHGF EVSQDVFERF KDKEGGFSGE LKGDVQGLLS
LYEASYLGFE GENLLEEART FSITHLKNNL KEGINTKVAE QVSHALELPY
HQRLHRLEAR WFLDKYEPKE PHHQLLLELA KLDFNMVQTL HQKELQDLSR
WWTEMGLASK LDFVRDRLME VYFWALGMAP DPQFGECRKA VTKMFGLVTI
IDDVYDVYGT LDELQLFTDA VERWDVNAIN TLPDYMKLCF LALYNTVNDT
SYSILKEKGH NNLSYLTKSW RELCKAFLQE AKWSNNKIIP AFSKYLENAS
VSSSGVALLA PSYFSVCQQQ EDISDHALRS LTDFHGLVRS SCVIFRLCND
LATSAAELER GETTNSIISY MHENDGTSEE QAREELRKLI DAEWKKMNRE
RVSDSTLLPK AFMEIAVNMA RVSHCTYQYG DGLGRPDYAT ENRIKLLLID
PFPINQLMYV

SEQ ID NO 16

Hybrid between *Eucalyptus globulus* isoprene synthase (AB266390.1) deleted up to L137 and the structured arm of Snapdragon (*Antirrhinum majus*) myrcene synthase

DVGSTPPPSK LHQALCLNEH SLSCMAELPM DYEGKIKETR HLLHLKGEND
PIESLHGTAL GFRLLRQHGY EVSQDVFKIF LDESGSFVKT LGGDVQGVLS
LYEASHLAFE EEHILHKARS FAIKHLENLN SDVDKDLQDQ VKHELELPLH
RRMPLLEARR SIEAYSRRGY TNPQILELAL TDFNVSQSYL QRDLQEMLGW
WNNTGLAKRL SFARDRLIEC FFWAVGIAHE PSLSICRKAV TKAFALILVL
DDVYDVFGTL EELELFTDAV RRWDLNAVED LPVYMKLCYL ALYNSVNEMA
YETLKEKGEN VIPYLAKAWY DLCKAFLQEA KWSNSRIIPG VEEYLNNGWV
SSSGSVMLIH AYFLASPSIR KEELESLEHY HDLLRLPSLI FRLTNDIASS
SAELERGETT NSIRCFMQEK GISELEAREC VKEEIDTAWK KMNKYMVDRS
TFNQSFVRMT YNLARMAHCV YQDGDAIGSP DDLSWNRVHS LIIKPISPAA

SEQ ID NO 17

Hybrid between *Melaleuca alternifolia* isoprene synthase (GenBank: AY279379.1) deleted up to L137 and the structured arm of Snapdragon (*Antirrhinum majus*) myrcene synthase

DVGSTPPPSK LHQALCLNEH SLSCMAELPM DYEGKIKETR HLLHLKGEND
PIESLHGTAL SFRVLRQHGF EVSQDVFKIF MDESGSFMKT LGGDVQGMLS
LYEASHLAFE EEDILHKAKT FAIKHLENLN HDIDQDLQDH VNHELELPLH
RRMPLLEARR FIEAYSRRSN VNPRILELAV MKFNSSQLTL QRDLQDMLGW
WNNVGLAKRL SFARDRLMEC FFWAVGIARE PALSNCRKGV TKAFSLILVL
DDVYDVFGTL DELELFTDAV RRWHEDAVEN LPGYMKLCFL ALYNSVNDMA
YETLKETGEN VTPYLTKVWY DLCKAFLQEA KWSYNKITPG VEEYLNNGWV
SSSGQVMLTH AYFLSSPSLR KEELESLEHY HDLLRLPSLI FRLTNDLATS
SAELGRGETT NSILCYMREK GFSESEARKQ VIEQIDTAWR QMNKYMVDHS
TFNRSFMQMT YNLARMAHCV YQDGDAIGAP DDQSWNRVHS LIIKPVSLAP
C

SEQ ID NO 18

Hybrid between *Robinia pseudoaccacia* isoprene synthase deleted up to L91 and the structured arm of Snapdragon (*Antirrhinum majus*) myrcene synthase

DVGSTPPPSK LHQALCLNEH SLSCMAELPM DYEGKIKETR HLLHLKGEND
PIESLHATAL IFRLLRQHGF EVSQDVFEST RDKEGRFKAE IKGDVQGLLS
LYEASYLGSE GENLLDEARE FSMTHLKNLN EGVVTPKLAE QINHALELPY
HRRFQRLEAR WFIENYEVKE PHDRLLVELA KLDFNMVQSL QKKEVGESSR
WWKEIGLTSK LDFVRDRLVE VYFWASGMAP DPQLSECRKA VTKMFGLVTI
IDDVYDVYGT LDELELFTNA VERWDVNAVD TLPDYMKLCF FALYNTVNDT
AYNLLKEKGD NNLPYLAKSW SDLCKAFLQE AKWSNNKIIP SFNKYIENAS
VSSSGGALLT PCYFSIRQDI TNQALDSLTN YHGPVRSSCA IFRLCNDLAT
SAAELERGET TNSITSCMQD NGISEEQARD ELRNLIDAEW KQMNRERVFD
QTFPKAFIET AINMARVSHC TYQYGDGLGR PDNTAENRIK LLLIDPLSN

SEQ ID NO 19

Physcomitrella patens PpCPS/KS mRNA for ent-kaurene synthase
Adapted from Genbank: AB302933
S709 is shown in bold and highlighted.

*MASSTLIQNR SCGVTSSMSS FQIFRGQPLR FPGTRTPAAV Q*CLKKRRCLR
PTESVLESSP GSGSYRIVTG PSGINPSSNG HLQEGSLTHR LPIPMEKSID
NFQSTLYVSD IWSETLQRTE CLLQVTENVQ MNEWIEEIRM YFRNMTLGEI
SMSPYDTAWV ARVPALDGSH GPQFHRSLQW IIDNQLPDGD WGEPSLFLGY
DRVCNTLACV IALKTWGVGA QNVERGIQFL QSNIYKMEED DANHMPIGFE
IVFPAMMEDA KALGLDLPYD ATILQQISAE REKKMKKIPM AMVYKYPTTL
LHSLEGLHRE VDWNKLLQLQ SENGSFLYSP ASTACALMYT KDVKCFDYLN
QLLIKFDHAC PNVYPVDLFE RLWMVDRLQR LGISRYFERE IRDCLQYVYR
YWKDCGIGWA SNSSVQDVDD TAMAFRLLRT HGFDVKEDCF RQFFKDGEFF
CFAGQSSQAV TGMFNLSRAS QTLFPGESLL KKARTFSRNF LRTKHENNEC
FDKWIITKDL AGEVEYNLTF PWYASLPRLE HRTYLDQYGI DDIWIGKSLY
KMPAVTNEVF LKLAKADFNM CQALHKKELE QVIKWNASCQ FRDLEFARQK
SVECYFAGAA TMFEPEMVQA RLVWARCCVL TTVLDDYFDH GTPVEELRVF
VQAVRTWNPE LINGLPEQAK ILFMGLYKTV NTIAEEAFMA QKRDVHHHLK

HYWDKLITSA LKEAEWAESG YVPTFDEYME VAEISVALEP IVCSTLFFAG
HRLDEDVLDS YDYHLVMHLV NRVGRILNDI QGMKREASQG KISSVQIYME
EHPSVPSEAM AIAHLQELVD NSMQQLTYEV LRFTAVPKSC KRIHLNMAKI
MHAFYKDTDG FSSLTAMTGF VKKVLFEPVPE

SEQ ID NO 20

Snapdragon (*Antirrhinum majus*) myrcene synthase converted into an isoprene synthase by making the change V331F (shown), V331W or V331Y (not shown).
Adapted from Genbank: AY195608.1
V331F is shown bold and highlighted

*MAFCISYVGA LLPCSLSTRT KFAICHNTSK LHRAAYKTSR WNIPG*DVGST
PPPSKLHQAL CLNEHSLSCM AELPMDYEGK IKETRHLLHL KGENDPIESL
IFVDATLRLG VNHHFQKEIE EILRKSYATM KSPIICEYHT LHEVSLFFRL
MRQHGRYVSA DVFNNFKGES GRFKEELKRD TRGLVELYEA AQLSFEGERI
LDEAENFSRQ ILHGNLAGME DNLRRSVGNK LRYPFHTSIA RFTGRNYDDD
LGGMYEWGKT LRELALMDLQ VERSVYQEEL LQVSKWWNEL GLYKKLNLAR
NRPFEFYTWS MVILADYINL SEQRVELTKS **F**AFIYLIDDI FDVYGTLDEL
IIFTEAVNKW DYSATDTLPE NMKMCCMTLL DTINGTSQKI YEKHGYNPID
SLKTTWKSLC SAFLVEAKWS ASGSLPSANE YLENEKVSSG VYVVLVHLFC
LMGLGGTSRG SIELNDTQEL MSSIAIIFRL WNDLGSAKNE HQNGKDGSYL
NCYKKEHINL TAAQAHEHAL ELVAIEWKRL NKESFNLNHD SVSSFKQAAL
NLARMVPLMY SYDHNQRGPV LEEYVKFMLS D

SEQ ID NO 21

Snapdragon (*Antirrhinum majus*) myrcene synthase converted into an isoprene synthase by making the change V444F (shown), V444W or V444Y (not shown).
Adapted from Genbank: AY195608.1
V444F is shown bold and highlighted

*MAFCISYVGA LLPCSLSTRT KFAICHNTSK LHRAAYKTSR WNIPG*DVGST
PPPSKLHQAL CLNEHSLSCM AELPMDYEGK IKETRHLLHL KGENDPIESL
IFVDATLRLG VNHHFQKEIE EILRKSYATM KSPIICEYHT LHEVSLFFRL
MRQHGRYVSA DVFNNFKGES GRFKEELKRD TRGLVELYEA AQLSFEGERI
LDEAENFSRQ ILHGNLAGME DNLRRSVGNK LRYPFHTSIA RFTGRNYDDD
LGGMYEWGKT LRELALMDLQ VERSVYQEEL LQVSKWWNEL GLYKKLNLAR
NRPFEFYTWS MVILADYINL SEQRVELTKS VAFIYLIDDI FDVYGTLDEL
IIFTEAVNKW DYSATDTLPE NMKMCCMTLL DTINGTSQKI YEKHGYNPID
SLKTTWKSLC SAFLVEAKWS ASGSLPSANE YLENEKVSSG VYV**F**LVHLFC
LMGLGGTSRG SIELNDTQEL MSSIAIIFRL WNDLGSAKNE HQNGKDGSYL
NCYKKEHINL TAAQAHEHAL ELVAIEWKRL NKESFNLNHD SVSSFKQAAL
NLARMVPLMY SYDHNQRGPV LEEYVKFMLS D

SEQ ID NO 22

Snapdragon (*Antirrhinum majus*) myrcene synthase converted into an isoprene synthase by making the change V331F and V444F (shown), (V331F and V444W), (V331F and V444Y), (V331W and V444F), (V331W and V444W), (V331W and V444Y), (V331Y and V444F), (V331Y and V444W) or (V331Y and V444Y) (not shown).
Adapted from Genbank: AY195608.1
V331F and V444F are shown bold and highlighted

*MAFCISYVGA LLPCSLSTRT KFAICHNTSK LHRAAYKTSR WNIPG*DVGST
PPPSKLHQAL CLNEHSLSCM AELPMDYEGK IKETRHLLHL KGENDPIESL
IFVDATLRLG VNHHFQKEIE EILRKSYATM KSPIICEYHT LHEVSLFFRL
MRQHGRYVSA DVFNNFKGES GRFKEELKRD TRGLVELYEA AQLSFEGERI
LDEAENFSRQ ILHGNLAGME DNLRRSVGNK LRYPFHTSIA RFTGRNYDDD
LGGMYEWGKT LRELALMDLQ VERSVYQEEL LQVSKWWNEL GLYKKLNLAR
NRPFEFYTWS MVILADYINL SEQRVELTKS **F**AFIYLIDDI FDVYGTLDEL
IIFTEAVNKW DYSATDTLPE NMKMCCMTLL DTINGTSQKI YEKHGYNPID
SLKTTWKSLC SAFLVEAKWS ASGSLPSANE YLENEKVSSG VYV**F**LVHLFC
LMGLGGTSRG SIELNDTQEL MSSIAIIFRL WNDLGSAKNE HQNGKDGSYL
NCYKKEHINL TAAQAHEHAL ELVAIEWKRL NKESFNLNHD SVSSFKQAAL
NLARMVPLMY SYDHNQRGPV LEEYVKFMLS D

SEQ ID NO 23

Solanum lycopersicum cultivar M82 β-phellandrene synthase (PHS1)

Adapted from GenBank FJ797957.1

This β-phellandrene synthase is converted into an isoprene producing enzyme using a combination of one or both of the changes V524F and M672F. V524 and M672 are shown bold, underlined and high-lighted.

*MIVGYRSTII TLSHPKLGNG KTISSNAIFQ RSCRVR*CSHS TTSSMNGFED
ARDRIRESFG KLELSPSSYD TAWVAMVPSR HSLNEPCFPQ CLDWIIENQR
EDGSWGLNPT HPLLLKDSLS STLACLLALT KWRVGDEQIK RGLGFIETYG
WAVDNKDQIS PLGFEVIFSS MIKSAEKLDL NLPLNLHLVN LVKCKRDSTI
KRNVEYMGEG VGELCDWKEM IKLHQRQNGS LFDSPATTAA ALIYHQHDQK
CYQYLNSIFQ QHKNWVPTMY PTKVHSLLCL VDTLQNLGVH RHFKSEIKKA
LDEIYRLWQQ KNEQIFSNVT HCAMAFRLLR MSYYDVSSDE LAEFVDEEHF
FATNGKYKSH VEILELHKAS QLAIDHEKDD ILDKINNWTR AFMEQKLLNN
GFIDRMSKKE VELALRKFYT TSHLAENRRY IKSYEENNFK ILKAAYRSPN
INNKDLLAFS IHDFELCQAQ HREELQQLKR WFEDYRLDQL GLAERYIHAS
YLFGVTVIPE PELSDARLMY AKY**Y**MLLTIV DDHFESFASK DECFNIIELV
ERWDDYASVG YKSEKVKVFF SVFYKSIEEL ATIAEIKQGR SVKNHLINLW
LELMKLMLME RVEWCSGKTI PSIEEYLYVT SITFCAKLIP LSTQYFLGIK
ISKDLLESDE ICGLWNCSGR V**M**RILNDLQD SKREQKEVSI NLVTLLMKSM
SEEEAIMKIK EILEMNRREL LKMVLVQKKG SQLPQLCKDI FWRTSKWAHF
TYSQTDGYRI AEEMKNHIDE VFYKPLNH

SEQ ID NO 24

Populus tremuloides isoprene synthase
Adapted from Genbank: AY341431.1
The active site is opened up by making the mutation F338L and F485W (shown bold and underlined), F338L and F485Y, F338W and F485L, F338Y and F485L (not shown)

MATELLCLHR PISLTHKLFR NPLPKVIQAT PLTLKLRCSV STENVSFSET
ETETRRSANY EPNSWDYDYL LSSDTDESIE VHKDKAKKLE AEVRREINNE
KAEFLTLLEL IDNVQRLGLG YRFESDIRRA LDRFVSSGGF DGVTKTSLHG
TALSFRLLRQ HGFEVSQEAF SGFKDQNGNF LENLKEDIKA ILSLYEASFL
ALEGENILDE AKVFAISHLK ELSEEKIGKE LAEQVSHALE LPLHRRTQRL
EAVWSIEAYR KKEDANQVLL ELAILDYNMI QSVYQRDLRE TSRWWRRVGL
ATKLHFARDR LIESFYWAVG VAFEPQYSDC RNSVAKM**L**SF VTIIDDIYDV
YGTLDELELF TDAVERWDVN AINDLPDYMK LCFLALYNTI NEIAYDNLKD
KGENILPYLT KAWADLCNAF LQEAKWLYNK STPTFDDYFG NAWKSSSGPL
QLIFAYFAVV QNIKKEEIEN LQKYHDIISR PSHI**W**RLCND LASASAEIAR
GETANSVSCY MRTKGISEEL ATESVMNLID ETWKKMNKEK LGGSLFAKPF
VETAINLARQ SHCTYHNGDA HTSPDELTRK RVLSVITEPI LPFER

SEQ ID NO 25

Populus alba Isoprene synthase
Adapted from Genbank: AB198180.1
The active site is opened up by making the mutation F338L and F485W (shown bold and underlined), F338L and F485Y, F338W and F485L, F338Y and F485L (not shown)

MATELLCLHR PISLTHKLFR NPLPKVIQAT PLTLKLRCSV STENVSFTET
ETEARRSANY EPNSWDYDYL LSSDTDESIE VYKDKAKKLE AEVRREINNE
KAEFLTLLEL IDNVQRLGLG YRFESDIRGA LDRFVSSGGF DAVTKTSLHG
TALSFRLLRQ HGFEVSQEAF SGFKDQNGNF LENLKEDIKA ILSLYEASFL
ALEGENILDE AKVFAISHLK ELSEEKIGKE LAEQVNHALE LPLHRRTQRL
EAVWSIEAYR KKEDANQVLL ELAILDYNMI QSVYQRDLRE TSRWWRRVGL
ATKLHFARDR LIESFYWAVG VAFEPQYSDC RNSVAKM**L**SF VTIIDDIYDV
YGTLDELELF TDAVERWDVN AINDLPDYMK LCFLALYNTI NEIAYDNLKD
KGENILPYLT KAWADLCNAF LQEAKWLYNK STPTFDDYFG NAWKSSSGPL
QLVFAYFAVV QNIKKEEIEN LQKYHDTISR PSHI**W**RLCND LASASAEIAR
GETANSVSCY MRTKGISEEL ATESVMNLID ETWKKMNKEK LGGSLFAKPF
VETAINLARQ SHCTYHNGDA HTSPDELTRK RVLSVITEPI LPFER

SEQ ID NO 26

Populus nigra Isoprene synthase
Adapted from Genbank: HQ684728.1
The active site is opened up by making the mutation F338L and F485W (shown bold and underlined), F338L and F485Y, F338W and F485L, F338Y and F485L (not shown)

MATELLCLHR PISLTHKLFR NPLPKVIQAT PLTLKLRCSV STENVSFTET
ETETRRSANY EPNSWDYDYL LSSDTDESIE VYKDKAKKLE AEVRREINNE

KAEFLTLPEL IDNVQRLGLG YRFESDIRRA LDRFVSSGGF DAVTKTSLHA
TALSFRLLRQ HGFEVSQEAF SGFKDQNGNF LKNLKEDIKA ILSLYEASFL
ALEGENILDE AKVFAISHLK ELSEEKIGKD LAEQVNHALE LPLHRRTQRL
EAVWSIEAYR KKEDADQVLL ELAILDYNMI QSVYQRDLRE TSRWWRRVGL
ATKLHFARDR LIESFYWAVG VAFEPQYSDC RNSVAKM**L**SF VTIIDDIYDV
YGTLDELELF TDAVERWDVN AIDDLPDYMK LCFLALYNTI NEIAYDNLKD
KGENILPYLT KAWADLCNAF LQEAKWLYNK STPTFDEYFG NAWKSSSGPL
QLVFAYFAVV QNIKKEEIDN LQKYHDIISR PSHI**W**RLCND LASASAEIAR
GETANSVSCY MRTKGISEEL ATESVMNLID ETWKKMNKEK LGGSLFAKPF
VETAINLARQ SHCTYHNGDA HTSPDELTRK RVLSVITEPI LPFER

SEQ ID NO 27

Populus trichocarpa Isoprene synthase
Adapted from Genbank: EU693027.1
The active site is opened up by making the mutation F301L and F448W (shown bold and underlined), F301L and F448Y, F301W and F448L, F301Y and F448L (not shown)

CSVSTENVSF TETETETRRS ANYEPNSWDY DYLLSSDTDE SIEVYKDKAK
KLEAEVRREI NNEKAEFLTL LELIDNVQRL GLGYRFESDI RRALDRFVSS
GGFDAVTKTS LHATALSFRL LRQHGFEVSQ EAFSGFKDQN GNFLENLKED
IKAILSLYEA SFLALEGENI LDEAKVFAIS HLKELSEEKI GKDLAEQVNH
ALELPLHRRT QRLEAVLSIE AYRKKEDADQ VLLELAILDY NMIQSVYQRD
LRETSRWWRR VGLATKLHFA RDRLIESFYW AVGVAFEPQY SDCRNSVAKM
LSFVTIIDDI YDVYGTLDEL ELFTNAVERW DVNAIDDLPD YMKLCFLALY
NTINEIAYDN LKEKGENILP YLTKAWADLC NAFLQEAKWL YNKSTPTFDD
YFGNAWKSSS GPLQLVFAYF AVVQNIKKEE IENLQKYHDI ISRPSHI**W**RL
CNDLASASAE IARGETANSV SCYMRTKGIS EELATESVMN LIDETWKKMN
KEKLGGSLFA KPFVETAINL ARQSHCTYHN GDAHTSPDEL TRKRVLSVIT
EPILPFER

SEQ ID NO 28

Pueraria Montana Isoprene synthase
Adapted from Genbank: AY316691.1
The active site is opened up by making the mutation F343L and F493W (shown bold and underlined), F343L and F493Y, F343W and F493L, F343Y and F493L (not shown)

MATNLLCLSN KLSSPTPTPS TRFPQSKNFI TQKTSLANPK PWRVICATSS
*QFTQITE*HNS RRSANYQPNL WNFEFLQSLE NDLKVEKLEE KATKLEEEVR
CMINRVDTQP LSLLELIDDV QRLGLTYKFE KDIIKALENI VLLDENKKNK
SDLHATALSF RLLRQHGFEV SQDVFERFKD KEGGFSGELK GDVQGLLSLY
EASYLGFEGE NLLEEARTFS ITHLKNNLKE GINTKVAEQV SHALELPYHQ
RLHRLEARWF LDKYEPKEPH HQLLLELAKL DFNMVQTLHQ KELQDLSRWW
TEMGLASKLD FVRDRLMEVY FWALGMAPDP QFGECRKAVT KM**L**GLVTIID
DVYDVYGTLD ELQLFTDAVE RWDVNAINTL PDYMKLCFLA LYNTVNDTSY
SILKEKGHNN LSYLTKSWRE LCKAFLQEAK WSNNKIIPAF SKYLENASVS
SSGVALLAPS YFSVCQQQED ISDHALRSLT DFHGLVRSSC VI**W**RLCNDLA

TSAAELERGE TTNSIISYMH ENDGTSEEQA REELRKLIDA EWKKMNRERV
SDSTLLPKAF MEIAVNMARV SHCTYQYGDG LGRPDYATEN RIKLLLIDPF
PINQLMYV

SEQ ID NO 29

Eucalyptus globulus "mts-1 mRNA for monoterpene synthase" actually an isoprene synthase
Adapted from Genbank: AB266390.1
The active site is opened up by making the mutation F326L and F473W (shown bold and underlined), F326L and F473Y, F326W and F473L, F326Y and F473L (not shown)

*MALRLLFTPH LPVLSSRRAN GRVRCSASTQ ISD*PQEGRRS ANYQPSVWTY
NYLQSIVAGE GRQSRREVEQ QKEKVQILEE EVRGALNDEK AETFTIFATV
DDIQRLGLGD HFEEDISNAL RRCVSKGAVF MSLQKSLHGT ALGFRLLRQH
GYEVSQDVFK IFLDESGSFV KTLGGDVQGV LSLYEASHLA FEEEHILHKA
RSFAIKHLEN LNSDVDKDLQ DQVKHELELP LHRRMPLLEA RRSIEAYSRR
GYTNPQILEL ALTDFNVSQS YLQRDLQEML GWWNNTGLAK RLSFARDRLI
ECFFWAVGIA HEPSLSICRK AVTKA**L**ALIL VLDDVYDVFG TLEELELFTD
AVRRWDLNAV EDLPVYMKLC YLALYNSVNE MAYETLKEKG ENVIPYLAKA
WYDLCKAFLQ EAKWSNSRII PGVEEYLNNG WVSSSGSVML IHAYFLASPS
IRKEELESLE HYHDLLRLPS LI**W**RLTNDIA SSSAELERGE TTNSIRCFMQ
EKGISELEAR ECVKEEIDTA WKKMNKYMVD RSTFNQSFVR MTYNLARMAH
CVYQDGDAIG SPDDLSWNRV HSLIIKPISP AA

SEQ ID NO 30

Melaleuca alternifolia "putative monoterpene synthase mRNA", actually an isoprene synthase
Adapted from GenBank: AY279379.1
The active site is opened up by making the mutation F326L and F473W (shown bold and underlined), F326L and F473Y, F326W and F473L, F326Y and F473L (not shown)

*MALRLLSTPH LPQLCSRRVS GRVHCSASTQ VSD*AQGGRRS ANYQPSVWTY
NYLQSLVADD IRRSRREVEQ EREKAQILEE DVRGALNDGN AEPMAIFALV
DDIQRLGLGR YFEEDISKAL RRCLSQYAVT GSLQKSLHGT ALSFRVLRQH
GFEVSQDVFK IFMDESGSFM KTLGGDVQGM LSLYEASHLA FEEEDILHKA
KTFAIKHLEN LNHDIDQDLQ DHVNHELELP LHRRMPLLEA RRFIEAYSRR
SNVNPRILEL AVMKFNSSQL TLQRDLQDML GWWNNVGLAK RLSFARDRLM
ECFFWAVGIA REPALSNCRK GVTKA**L**SLIL VLDDVYDVFG TLDELELFTD
AVRRWHEDAV ENLPGYMKLC FLALYNSVND MAYETLKETG ENVTPYLTKV
WYDLCKAFLQ EAKWSYNKIT PGVEEYLNNG WVSSSGQVML THAYFLSSPS
LRKEELESLE HYHDLLRLPS LI**W**RLTNDLA TSSAELGRGE TTNSILCYMR
EKGFSESEAR KQVIEQIDTA WRQMNKYMVD HSTFNRSFMQ MTYNLARMAH
CVYQDGDAIG APDDQSWNRV HSLIIKPVSL APC

SEQ ID NO 31
Robinia pseudoaccacia isoprene synthase

Unpublished sequence
The active site is opened up by making the mutation F281L and F428W (shown bold and underlined), F281L and F428Y, F281W and F428L, F281Y and F428L (not shown)

RRSANYQPNL WNFEFLQSQE YDLMVETLQE RATKLEEEVR RLINRVDIEP
LKLLELVDNV QRLGLTYKFE DDINKALERI VSLDEREKSG LHATALIFRL
LRQHGFEVSQ DVFESTRDKE GRFKAEIKGD VQGLLSLYEA SYLGSEGENL
LDEAREFSMT HLKNLNEGVV TPKLAEQINH ALELPYHRRF QRLEARWFIE
NYEVKEPHDR LLVELAKLDF NMVQSLQKKE VGESSRWWKE IGLTSKLDFV
RDRLVEVYFW ASGMAPDPQL SECRKAVTKM **F**GLVTIIDDV YDVYGTLDEL
ELFTNAVERW DVN**A**VDTLPD YMKLCFFALY NTVNDTAYNL LKEKGDNNLP
YLAKSWSDLC KAFLQEAKWS NNKIIPSFNK YIENASV**SSS** GGALLTPCYF
SIRQDITNQA LDSLTNYHGP VRSSCAI**F**RL **C**NDLATSAAE LERGET**TN**SI
TSCMQDNGIS EEQARDELRN LIDAEWKQMN RERVFDQTFP KAFIETAINM
AR**VS**H**C**TYQY GDGLGRPDNT AENRIKLLLI DPLSN

SEQ ID NO 32

Humulus lupulus (Hops) isoprene synthase
Unpublished sequence
F354 and F530 are indicated in bold, underlined and high-lighted

MQCMAVHQFA PLLSLLNCSR ISSDFGRLFT PKTSTKSRSS TCHPIQCTVV
*NNT*DRRSANY EPSIWSFDYI QSLTSQYKGK SYSSRLNELK KEVKMMEDGT
KECLAQLDLI DTLQRLGISY HFEDEINTIL KRKYINIQNN INHNYNLYST
ALQFRLLRQH GYLVTQEVFN AFKDETGKFK TYLSDDIMGV LSLYEASFYA
MKHENVLEEA RVFSTECLKE YMMKMEQNKV LLDHDLDHND NFNVNHHVLI
INHALELPLH WRITRSEARW FIDVYEKKQD MDSTLLEFAK LDFNMVQSTH
QEDLKHLSRW WRHSKLGEKL NFARDRLMEA FLWEVGLKFE PEFSYFKRIS
ARL**F**VLITII DDIYDVYGTL EELELFTKAV ERWDVNAINE LPEYMKMPFL
VLHNTINEMA FDVLGDQNFL NIEYLKKSLV DLCKCYLQEA KWYYSGYQPT
LQEYIEMAWL SIGGPVILVH AYFCFTNPIT KESMKFFTEG YPNIIQQSCL
IVRLADDFGT FSDELNRGDV PKSIQCYMYD TGASEDEARE HIKFLICETW
KDMNKNDEDN SCFSETFVEV CKNLARTAL**F** MYQYGDGHAS QNCLSKERIF
ALIINPINFH ERK

SEQ ID NO 33

Aspergillus fumigatus tryptophan dimethylallyltransferase
Adapted from Genbank AY775787
Resides that line the active site and can be converted into phenylalanines (F) are shown bold and highlighted.

MKAANASSAE AYRVLSRAFR FDNEDQKLWW HSTAPMFAKM LETANYTTPC
QYQYLITYKE CVIPSLGCYP TNSAPRWLSI **LTR**YGTPFEL SLNCSNSIVR
YTFEPINQHT GTDKDPFNTH AIWESLQHLL PLEKSIDLEW FRHFKHDLTL
NSEESAFLAH NDRLVGGTIR TQNKLALDLK DGRFALKTYI **Y**PALKAVVTG
KTIHELVFGS VRRLAVREPR ILPPLNMLEE YIRSRGSKST ASPRLVSCDL

TSPAKSRIKI YLLEQMVSLE AMEDLWTLGG RRRDASTLEG LSLVRELWDL
IQLSPGLKSY PAPYLPLGVI PDERLPLMAN FTLHQNDPVP EPQV**Y**FTTFG
MNDMAVADAL TTFFERRGWS EMARTYETTL KSYYPHADHD KLNYLHA**Y**IS
FSYRDRTPYL SVYLQSFETG DWAVANLSES KVKCQDAACQ PTALPPDLSK
TGVYYSGLH

SEQ ID NO 34

Streptomyces sp. UC5319 pentalenene synthase gene
Adapted from Genbank: U05213.1

MPQDVDFHIP LPGRQSPDHA RAEAEQLAWP RSLGLIRSDA AAERHLRGGY
ADLASRFYPH ATGADLDLGV DLMSWFFLFD DLFDGPRGEN PEDTKQLTDQ
VAAALDGPLP DTAPPIAHGF ADIWRRTCEG MTPAWCARSA RHWRNYFDGY
VDEAESRFWN APCDSAAQYL AMRRHTIGVQ PTVDLAERAG RFEVPHRVFD
SAVMSAMLQI AVDVNLLLND IASLEKEEAR GEQNNMVMIL RREHGWSKSR
SVSHMQNEVR ARLEQYLLLE SCLPKVGEIY QLDTAEREAL ERYRTDAVRT
VIRGSYDWHR SSGRYDAEFA LAAGAQGYLE ELGSSAH

SEQ ID NO 35

Streptomyces sp. UC5319 pentalenene synthase gene converted into an isoprene synthase by making the change V179F (shown), V179W or V179Y (not shown)
Adapted from Genbank: U05213.1
V179F is shown in bold and highlighted.

MPQDVDFHIP LPGRQSPDHA RAEAEQLAWP RSLGLIRSDA AAERHLRGGY
ADLASRFYPH ATGADLDLGV DLMSWFFLFD DLFDGPRGEN PEDTKQLTDQ
VAAALDGPLP DTAPPIAHGF ADIWRRTCEG MTPAWCARSA RHWRNYFDGY
VDEAESRFWN APCDSAAQYL AMRRHTIG**F**Q PTVDLAERAG RFEVPHRVFD
SAVMSAMLQI AVDVNLLLND IASLEKEEAR GEQNNMVMIL RREHGWSKSR
SVSHMQNEVR ARLEQYLLLE SCLPKVGEIY QLDTAEREAL ERYRTDAVRT
VIRGSYDWHR SSGRYDAEFA LAAGAQGYLE ELGSSAH

SEQ ID NO 36

Streptomyces sp. UC5319 pentalenene synthase gene converted into an isoprene synthase by making the change N219F (shown), N219W or N219Y (not shown)
Adapted from Genbank: U05213.1
N219F is shown in bold and highlighted.

MPQDVDFHIP LPGRQSPDHA RAEAEQLAWP RSLGLIRSDA AAERHLRGGY
ADLASRFYPH ATGADLDLGV DLMSWFFLFD DLFDGPRGEN PEDTKQLTDQ
VAAALDGPLP DTAPPIAHGF ADIWRRTCEG MTPAWCARSA RHWRNYFDGY
VDEAESRFWN APCDSAAQYL AMRRHTIGVQ PTVDLAERAG RFEVPHRVFD
SAVMSAMLQI AVDVNLLL**F**D IASLEKEEAR GEQNNMVMIL RREHGWSKSR
SVSHMQNEVR ARLEQYLLLE SCLPKVGEIY QLDTAEREAL ERYRTDAVRT
VIRGSYDWHR SSGRYDAEFA LAAGAQGYLE ELGSSAH

SEQ ID NO 37

Streptomyces sp. UC5319 pentalenene synthase gene converted into an isoprene synthase by making the change V179F and N219F (shown), (V179F and N219W), (V179F and N219Y), (V179W and N219F), (V179W and N219W), (V179W and N219Y), (V179Y and N219F), (V179Y and N219W) or (V179Y and N219Y) (not shown)
Adapted from Genbank: U05213.1
V179F and N219F are shown in bold and highlighted.

MPQDVDFHIP LPGRQSPDHA RAEAEQLAWP RSLGLIRSDA AAERHLRGGY
ADLASRFYPH ATGADLDLGV DLMSWFFLFD DLFDGPRGEN PEDTKQLTDQ
VAAALDGPLP DTAPPIAHGF ADIWRRTCEG MTPAWCARSA RHWRNYFDGY
VDEAESRFWN APCDSAAQYL AMRRHTIG**F**Q PTVDLAERAG RFEVPHRVFD
SAVMSAMLQI AVDVNLLL**F**D IASLEKEEAR GEQNNMVMIL RREHGWSKSR
SVSHMQNEVR ARLEQYLLLE SCLPKVGEIY QLDTAEREAL ERYRTDAVRT
VIRGSYDWHR SSGRYDAEFA LAAGAQGYLE ELGSSAH

SEQ ID NO:38

pTrcHis2B-P1IspS

gtttgacagcttatcatcgactgcacggtgcaccaatgcttctggcgtcaggca
gccatcggaagctgtggtatggctgtgcaggtcgtaaatcactgcataattcgtgtcgc
tcaaggcgcactcccgttctggataatgtttttgcgccgacatcataacggttctggc
aaatattctgaaatgagctgttgacaattaatcatccggctcgtataatgtgtggaatt
gtgagcggataacaatttcacacaggaaacagcgccgctgagaaaaagcgaagcggcac
tgctctttaacaatttatcagacaatctgtgtgggcactcgaccggaattatcgattaa
ctttattattaaaaattaaagaggtatatattaatgtatcgattaaataaggaggaata
aaccatgagatgtagcgtgtccaccgaaaatgtgtctttcaccgaaactgaaaccgaag
ctcgtcgttctgcgaactacgaacctaacagctgggactatgattacctgctgtcctcc
gacacggacgagtccatcgaagtatacaaagacaaagcgaaaaagctggaagccgaagt
tcgtcgcgagattaataacgaaaaagcagaatttctgaccctgctggaactgattgaca
acgtccagcgcctgggcctgggttaccgtttcgagtctgatatccgtggtgcgctggat
cgcttcgtttcctccggcggcttcgatgcggtaaccaagacttccctgcacggtacggc
actgtctttccgtctgctgcgtcaacacggttttgaggtttctcaggaagcgttcagcg
gcttcaaagaccaaaacggcaacttcctggagaacctgaaggaagatatcaaagctatc
ctgagcctgtacgaggccagcttcctggctctggaaggcgaaaacatcctggacgaggc
gaaggttttcgcaatctctcatctgaaagaactgtctgaagaaaagatcggtaaagagc
tggcagaacaggtgaaccatgcactggaactgccactgcatcgccgtactcagcgtctg
gaagcagtatggtctatcgaggcctaccgtaaaaaggaggacgcgaatcaggttctgct
ggagctggcaattctggattacaacatgatccagtctgtataccagcgtgatctgcgtg
aaacgtcccgttggtggcgtcgtgtgggtctggcgaccaaactgcactttgctcgtgac
cgcctgattgagagcttctactgggccgtgggtgtagcattcgaaccgcaatactccga
ctgccgtaactccgtcgcaaaaatgttttctttcgtaaccattatcgacgatatctacg
atgtatacggcaccctggacgaactggagctgtttactgatgcagttgagcgttgggac
gtaaacgccatcaacgacctgccggattacatgaaactgtgctttctggctctgtataa
cactattaacgaaatcgcctacgacaacctgaaagataaaggtgagaacatcctgccgt

atctgaccaaagcctgggctgacctgtgcaacgctttcctgcaagaagccaagtggctg
tacaacaaatctactccgacctttgacgactacttcggcaacgcatggaaatcctcttc
tggcccgctgcaactggtgttcgcttacttcgctgtcgtgcagaacattaaaaaggaag
agatcgaaaacctgcaaaaataccatgacaccatctctcgtccttcccatatcttccgt
ctgtgcaatgacctggctagcgcgtctgcggaaattgcgcgtggtgaaaccgcaaatag
cgtttcttgttacatgcgcactaaaggtatctccgaagaactggctaccgaaagcgtga
tgaatctgatcgatgaaacctggaaaaagatgaacaaggaaaaactgggtggtagcctg
ttcgcgaaaccgttcgtggaaaccgcgatcaacctggcacgtcaatctcactgcactta
tcataacggcgacgcgcatacctctccggatgagctgacccgcaaacgcgttctgtctg
taatcactgaaccgattctgccgtttgaacgctaactcgagatctgcagctggtaccat
atgggaattcgaagctttctagaacaaaaactcatctcagaagaggatctgaatagcgc
cgtcgaccatcatcatcatcatcattgagtttaaacggtctccagcttggctgttttgg
cggatgagagaagattttcagcctgatacagattaaatcagaacgcagaagcggtctga
taaaacagaatttgcctggcggcagtagcgcggtggtcccacctgaccccatgccgaac
tcagaagtgaaacgccgtagcgccgatggtagtgtggggtctccccatgcgagagtagg
gaactgccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgtttt
atctgttgtttgtcggtgaacgctctcctgagtaggacaaatccgccgggagcggattt
gaacgttgcgaagcaacggcccggagggtggcgggcaggacgcccgccataaactgcca
ggcatcaaattaagcagaaggccatcctgacggatggccttttgcgtttctacaaact
ctttttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccc
tgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgt
cgcccttattccctttttgcggcattttgccttcctgtttttgctcacccagaaacgc
tggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatcgaactg
gatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgat
gagcacttttaaagttctgctatgtggcgcggtattatcccgtgttgacgccgggcaag
agcaactcggtcgccgcatacactattctcagaatgacttggttgagtactcaccagtc
acagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataac
catgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagc
taaccgcttttttgcacaacatgggggatcatgtaactcgccttgatcgttgggaaccg
gagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgtagcaatggc
aacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaat
taatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccg
gctggctggtttattgctgataaatctggagccggtgagcgtgggtctcgcggtatcat
tgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacgggga
gtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgatt
aagcattggtaactgtcagaccaagtttactcatatatactttagattgatttaaaact
tcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaaaa
tcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaagga
tcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccacc
gctaccagcggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaa
ctggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggc
caccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttacc
agtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagt
taccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttg
gagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccac
gcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggag
agcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggttt
cgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatg

gaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctc
acatgttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgag
tgagctgataccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgagga
agcggaagagcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacacc
gcatatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagtataca
ctccgctatcgctacgtgactgggtcatggctgcgccccgacacccgccaacacccgct
gacgcgccctgacgggcttgtctgctcccggcatccgcttacagacaagctgtgaccgt
ctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggcagc
agatcaattcgcgcgcgaaggcgaagcggcatgcatttacgttgacaccatcgaatggt
gcaaaacctttcgcggtatggcatgatagcgcccggaagagagtcaattcagggtggtg
aatgtgaaaccagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagac
cgtttcccgcgtggtgaaccaggccagccacgtttctgcgaaaacgcgggaaaaagtgg
aagcggcgatggcggagctgaattacattcccaaccgcgtggcacaacaactggcgggc
aaacagtcgttgctgattggcgttgccacctccagtctggccctgcacgcgccgtcgca
aattgtcgcggcgattaaatctcgcgccgatcaactgggtgccagcgtggtggtgtcga
tggtagaacgaagcggcgtcgaagcctgtaaagcggcggtgcacaatcttctcgcgcaa
cgcgtcagtgggctgatcattaactatccgctggatgaccaggatgccattgctgtgga
agctgcctgcactaatgttccggcgttatttcttgatgtctctgaccagacacccatca
acagtattattttctcccatgaagacggtacgcgactgggcgtggagcatctggtcgca
ttgggtcaccagcaaatcgcgctgttagcgggcccattaagttctgtctcggcgcgtct
gcgtctggctggctggcataaatatctcactcgcaatcaaattcagccgatagcggaac
gggaaggcgactggagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgag
ggcatcgttcccactgcgatgctggttgccaacgatcagatggcgctgggcgcaatgcg
cgccattaccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatacgacg
ataccgaagacagctcatgttatatcccgccgtcaaccaccatcaaacaggattttcgc
ctgctggggcaaaccagcgtggaccgcttgctgcaactctctcagggccaggcggtgaa
gggcaatcagctgttgcccgtctcactggtgaaaagaaaaaccaccctggcgcccaata
cgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtt
tcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagcgcgaattga
tctg

SEQ ID NO:39

pTrcHis2B-E2IspS

gtttgacagcttatcatcgactgcacggtgcaccaatgcttctggcgtcaggca
gccatcggaagctgtggtatggctgtgcaggtcgtaaatcactgcataattcgtgtcgc
tcaaggcgcactcccgttctggataatgtttttgcgccgacatcataacggttctggc
aaatattctgaaatgagctgttgacaattaatcatccggctcgtataatgtgtggaatt
gtgagcggataacaatttcacacaggaaacagcgccgctgagaaaaagcgaagcggcac
tgctctttaacaatttatcagacaatctgtgtgggcactcgaccggaattatcgattaa
ctttattattaaaaattaaagaggtatatattaatgtatcgattaaataaggaggaata
aaccatggagggtcgccgttccgcgaactaccagccgtctgtgtggacctacaactacc
tgcaatccatcgtcgccggcgaaggccgtcaatcccgccgtgaagttgaacagcagaaa
gaaaaagtacagatcctggaagaagaggtgcgtggcgcactgaacgatgaaaaagcgga
gacgttcaccatcttcgctactgtggatgacattcagcgtctgggcctgggcgatcatt
tcgaggaagacatcagcaacgcgctgcgtcgttgcgtgagcaaaggtgcagtcttcatg
tctctgcaaaaaagcctgcacggcaccgcgctgggttttcgtctgctgcgccagcatgg

ctatgaagtatctcaggacgtatttaagatctttctggatgaaagcggcagctttgtca
aaaccctgggtggcgacgtgcagggcgtgctgtcctgtacgaggcttctcatctggcg
tttgaagaagagcacattctgcacaaagcgcgttcttttgcgatcaaacacctggaaaa
cctgaactctgacgtagataaagatctgcaagaccaggtgaaacacgaactggagctgc
cgctgcaccgtcgcatgccgctgctggaagctcgtcgtagcattgaggcttactctcgc
cgtggctacactaacccgcagattctggagctggctctgactgactttaacgtctctca
gagctacctgcaacgtgatctgcaagaaatgctgggttggtggaacaacaccggtctgg
ctaagcgcctgagctttgcgcgcgaccgcctgatcgaatgcttcttctgggccgtcggc
attgctcatgaaccgagcctgagcatctgccgcaaggcagtgaccaaagcttttgccct
gatcctggtactggacgatgtctacgatgttttcggtacgctggaagagctggagctgt
tcaccgacgcggttcgccgttgggatctgaacgcggttgaagatctgccggtgtatatg
aaactgtgctatctggctctgtacaactctgtaaatgagatggcttacgaaaccctgaa
agaaaaaggcgaaaatgtcattccgtacctggcgaaagcgtggtacgacctgtgtaaag
cattcctgcaagaagccaaatggagcaactctcgtattatcccgggtgtggaagagtac
ctgaacaacggttgggtttcttcttccggctccgtaatgctgattcacgcatatttcct
ggcttccccgagcattcgtaaagaggagctggaatctctggagcactatcacgatctgc
tgcgcctgccgtccctgatcttccgcctgaccaacgacattgcatcttccagcgctgaa
ctggaacgcggcgaaaccaccaacagcatccgttgtttcatgcaggaaaaaggtatctc
cgaactggaagcacgcgagtgtgtcaaagaagagatcgacactgcatggaaaaagatga
acaagtacatggtagatcgctccactttcaaccagtcctttgtccgcatgacttacaac
ctggcacgtatggcgcactgcgtataccaggacggtgacgcgattggttccccggacga
tctgtcctggaaccgtgtacactctctgatcattaaaccgatttctccggcggcttaac
tcgagatctgcagctggtaccatatgggaattcgaagctttctagaacaaaaactcatc
tcagaagaggatctgaatagcgccgtcgaccatcatcatcatcatcattgagtttaaac
ggtctccagcttggctgttttggcggatgagagaagattttcagcctgatacagattaa
atcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagtagcgcggtgg
tcccacctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtagtgtg
gggtctccccatgcgagagtagggaactgccaggcatcaaataaaacgaaaggctcagt
cgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctcctgagtagg
acaaatccgccgggagcggatttgaacgttgcgaagcaacggcccggagggtggcgggc
aggacgcccgccataaactgccaggcatcaaattaagcagaaggccatcctgacggatg
gccttttgcgtttctacaaactcttttgtttatttttctaaatacattcaaatatgt
atccgctcatgagacaataaccctgataaatgcttcaataatattgaaaaaggaagagt
atgagtattcaacatttccgtgtcgcccttattcccttttttgcggcattttgccttcc
tgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtg
cacgagtgggttacatcgaactggatctcaacagcggtaagatccttgagagttttcgc
cccgaagaacgttttccaatgatgagcacttttaaagttctgctatgtggcgcggtatt
atcccgtgttgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatg
acttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaaga
gaattatgcagtgctgccataaccatgagtgataacactgcggccaacttacttctgac
aacgatcggaggaccgaaggagctaaccgcttttttgcacaacatggggatcatgtaa
ctcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgac
accacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactact
tactctagcttcccggcaacaattaatagactggatggaggcggataaagttgcaggac
cacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggt
gagcgtgggtctcgcggtatcattgcagcactggggccagatggtaagccctcccgtat
cgtagttatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcg
ctgagataggtgcctcactgattaagcattggtaactgtcagaccaagtttactcatat

atactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatcct
ttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcag
accccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatctgc
tgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagct
accaactcttttccgaaggtaactggcttcagcagagcgcagataccaaatactgtcc
ttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatac
ctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttac
cgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggg
gttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacag
cgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggt
aagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggt
atctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgc
tcgtcaggggggcggagcctatggaaaaacgccagcaacgcggcctttttacggttcct
ggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctgtgg
ataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgaccgag
cgcagcgagtcagtgagcgaggaagcggaagagcgcctgatgcggtattttctccttac
gcatctgtgcggtatttcacaccgcatatggtgcactctcagtacaatctgctctgatg
ccgcatagttaagccagtatacactccgctatcgctacgtgactgggtcatggctgcgc
cccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatcc
gcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtc
atcaccgaaacgcgcgaggcagcagatcaattcgcgcgcgaaggcgaagcggcatgcat
ttacgttgacaccatcgaatggtgcaaaacctttcgcggtatggcatgatagcgcccgg
aagagagtcaattcagggtggtgaatgtgaaaccagtaacgttatacgatgtcgcagag
tatgccggtgtctcttatcagaccgtttcccgcgtggtgaaccaggccagccacgtttc
tgcgaaaacgcgggaaaaagtggaagcggcgatggcggagctgaattacattcccaacc
gcgtggcacaacaactggcgggcaaacagtcgttgctgattggcgttgccacctccagt
ctggccctgcacgcgccgtcgcaaattgtcgcggcgattaaatctcgcgccgatcaact
gggtgccagcgtggtggtgtcgatggtagaacgaagcggcgtcgaagcctgtaaagcgg
cggtgcacaatcttctcgcgcaacgcgtcagtgggctgatcattaactatccgctggat
gaccaggatgccattgctgtggaagctgcctgcactaatgttccggcgttatttcttga
tgtctctgaccagacacccatcaacagtattattttctcccatgaagacggtacgcgac
tgggcgtggagcatctggtcgcattgggtcaccagcaaatcgcgctgttagcgggccca
ttaagttctgtctcggcgcgtctgcgtctggctggctggcataaatatctcactcgcaa
tcaaattcagccgatagcggaacgggaaggcgactggagtgccatgtccggttttcaac
aaaccatgcaaatgctgaatgagggcatcgttcccactgcgatgctggttgccaacgat
cagatggcgctgggcgcaatgcgcgccattaccgagtccgggctgcgcgttggtgcgga
tatctcggtagtgggatacgacgataccgaagacagctcatgttatatcccgccgtcaa
ccaccatcaaacaggattttcgcctgctggggcaaaccagcgtggaccgcttgctgcaa
ctctctcagggccaggcggtgaagggcaatcagctgttgcccgtctcactggtgaaaag
aaaaaccaccctggcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcat
taatgcagctggcacgacaggtttcccgactggaaagcgggcagtgagcgcaacgcaat
taatgtgagttagcgcgaattgatctg

SEQ ID NO:40

pJexpress401-E4IspS-His

ctcatgaccaaaatcccttaacgtgagttacgcgcgcgtcgttccactgagcgt
cagaccccgtagaaaagatcaaaggatcttcttgagatccttttttttctgcgcgtaatc
tgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaaga
gctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactg
ttcttctagtgtagccgtagttagcccaccacttcaagaactctgtagcaccgcctaca
tacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtct
taccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgg
ggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctа
cagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatcc
ggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcct
ggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtga
tgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggtt
cctggccttttgctggccttttgctcacatgttctttcctgcgttatcccctgattctg
tggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagccgaacgacc
gagcgcagcgagtcagtgagcgaggaagcggaaggcgagagtagggaactgccaggcat
caaactaagcagaaggcccctgacggatggccttttgcgtttctacaaactctttctg
tgttgtaaaacgacggccagtcttaagctcgggcccctgggcggttctgataacgagt
aatcgttaatccgcaaataacgtaaaaacccgcttcggcgggtttttttatgggggggag
tttagggaaagagcatttgtcagaatatttaagggcgcctgtcactttgcttgatatat
gagaattatttaaccttataaatgagaaaaaagcaacgcactttaaataagatacgttg
ctttttcgattgatgaacacctataattaaactattcatctattatttatgatttttg
tatatacaatatttctagtttgttaaagagaattaagaaaataaatctcgaaaataata
aagggaaaatcagtttttgatatcaaaattatacatgtcaacgataatacaaaatataa
tacaaactataagatgttatcagtatttattatgcatttagaataaattttgtgtcgcc
cttaattgtgagcggataacaattacgagcttcatgcacagtgaaatcatgaaaaattt
atttgctttgtgagcggataacaattataatatgtggaattgtgagcgctcacaattcc
acaacggtttccctctagaaataatttttgtttaacttttctagaaataatttttgtttaa
ctttaagaggaggtaaaacatatggagggccgtcgcagcgcaaattaccaaccgtctgt
ttggacttataactacttgcagtcgatcgttgcgggtgagggtcgtcaaagccgccgtg
aggttgagcaacagaaagaaaaggttcagattctggaagaagaagtccgtggcgcactg
aatgacgaaaaggcggaaacgttcaccatctttgcaaccgttgacgatattcagcgcct
gggcctgggcgatcatttcgaagaggatattagcaacgcgctgcgtcgttgcgtgtcta
agggtgcggtctttatgagcttgcagaaaagcctgcacggtactgctctgggctttcgt
ctgctgcgccagcacggttatgaagtcagccaagacgtgttcaaaatcttcttggacga
gagcggttcttttgtgaaaacgctgggtggtgacgtgcagggtgttctgtctttgtacg
aggccagccatctggccttcgaagaagagcacatcctgcataaagcgcgtagcttcgcc
atcaagcatttggaaaatctgaacagcgatgttgataaggacctgcaagaccaagttaa
gcacgagctggagctgccgctgcaccgtcgtatgccgttgctggaagcgcgtcgctcta
ttgaggcgtactcccgccgtggctataccaatccgcagatcctggagttggccctgacc
gactttaatgtgagccagagctacttgcaacgtgacttgcaagagatgctgggctggtg
gaacaacacgggtctggcgaaacgcttgagctttgcgcgtgatcgtctgattgagtgtt
tcttttgggcagtgggtatcgcgcacgagccgagcctgagcatttgtcgtaaagctgtc
acgaaggcattcgctttgatcctggttctggatgacgtttacgatgtctttggcacgct
ggaagagctggaactgtttaccgacgcagtgcgccgctgggatctgaatgccgtcgaag
acctgccggtttacatgaaactgtgctacctggccctgtataactccgtcaacgagatg
gcgtatgaaaccctgaaagagaaaggcgagaatgtgattccgtacctggcaaaagcctg
gtacgatctgtgcaaggcattcctgcaagaggcgaagtggagcaacagccgtatcatcc
ctggtgttgaagaatatctgaataacggttgggtcagcagcagcggctccgttatgctg

atccatgcgtactttctggcgagcccgagcatccgtaaagaagaactggagtctctgga
gcactatcacgatctgctgcgtctgccttctctgattttccgcctgaccaatgacattg
cgagcagctccgcagaactggaacgcggtgaaacgaccaactccattcgttgtttcatg
caggagaaaggtatttccgaactggaggctcgtgaatgcgtgaaagaagagatcgacac
cgcttggaaaaagatgaataagtatatggtcgatcgcagcaccttcaatcagagcttcg
tccgtatgacctacaacctggcgcgtatggcgcattgcgtgtatcaggacggtgacgcg
attggtagcccggacgatctgtcctggaaccgcgtccacagcctgatcattaagccgat
tagcccagcggcagaaaacctgtattttcagagcggcagcggcagcggcagcggccatc
accatcaccaccattaactcgagtcgagccccaagggcgacacccctaattagcccgg
gcgaaaggcccagtctttcgactgagcctttcgttttatttgatgcctggcagttccct
actctcgcatggggagtccccacactaccatcggcgctacggcgtttcacttctgagtt
cggcatggggtcaggtgggaccaccgcgctactgccgccaggcaaacaaggggtgttat
gagccatattcaggtataaatgggctcgcgataatgttcagaattggttaattggttgt
aacactgacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatga
gacaataaccctgataaatgcttcaataatattgaaaaaggaagaatatgagccatatt
caacgggaaacgtcgaggccgcgattaaattccaacatggatgctgatttatatgggta
taaatgggctcgcgataatgtcgggcaatcaggtgcgacaatctatcgcttgtatggga
agcccgatgcgccagagttgtttctgaaacatggcaaaggtagcgttgccaatgatgtt
acagatgagatggtcagactaaactggctgacggaatttatgccacttccgaccatcaa
gcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccggaaaaa
cagcgttccaggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctg
gcagtgttcctgcgccggttgcactcgattcctgtttgtaattgtccttttaacagcga
tcgcgtatttcgcctcgctcaggcgcaatcacgaatgaataacggtttggttgatgcga
gtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcat
aaacttttgccattctcaccggattcagtcgtcactcatggtgatttctcacttgataa
ccttatttttgacgaggggaaattaataggttgtattgatgttggacgagtcggaatcg
cagaccgataccaggatcttgccatcctatggaactgcctcggtgagttttctccttca
ttacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgca
gtttcatttgatgctcgatgagtttttctaagcggcgccatcgaatggcgcaaaacc
tttcgcggtatggcatgatagcgcccggaagagagtcaattcagggtggtgaatatgaa
accagtaacgttatacgatgtcgcagagtatgccggtgtctcttatcagaccgtttccc
gcgtggtgaaccaggccagccacgtttctgcgaaaacgcgggaaaaagtggaagcggcg
atggcggagctgaattacattcccaaccgcgtggcacaacaactggcgggcaaacagtc
gttgctgattggcgttgccacctccagtctggccctgcacgcgccgtcgcaaattgtcg
cggcgattaaatctcgcgccgatcaactgggtgccagcgtggtggtgtcgatggtagaa
cgaagcggcgtcgaagcctgtaaagcggcggtgcacaatcttctcgcgcaacgcgtcag
tgggctgatcattaactatccgctggatgaccaggatgccattgctgtggaagctgcct
gcactaatgttccggcgttatttcttgatgtctctgaccagacacccatcaacagtatt
attttctcccatgaggacggtacgcgactgggcgtggagcatctggtcgcattgggtca
ccagcaaatcgcgctgttagcgggcccattaagttctgtctcggcgcgtctgcgtctgg
ctggctggcataaatatctcactcgcaatcaaattcagccgatagcggaacgggaaggc
gactggagtgccatgtccggttttcaacaaaccatgcaaatgctgaatgagggcatcgt
tcccactgcgatgctggttgccaacgatcagatggcgctgggcgcaatgcgcgccatta
ccgagtccgggctgcgcgttggtgcggatatctcggtagtgggatacgacgataccgaa
gatagctcatgttatatcccgccgttaaccaccatcaaacaggattttcgcctgctggg
gcaaaccagcgtggaccgcttgctgcaactctctcagggccaggcggtgaagggcaatc
agctgttgccagtctcactggtgaaaagaaaaaccaccctggcgcccaatacgcaaacc

gcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgact
ggaaagcgggcagtga

SEQ ID NO:41

pCL-SDFi

cccgtcttactgtcgggaattcgcgttggccgattcattaatgcagctggcacg
acaggtttcccgactggaaagcgggcagtgagcgcaacgcaattaatgtgagttagctc
actcattaggcaccccaggctttacactttatgcttccggctcgtatgttgtgtggaat
tgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagcttg
catgcctgcaggtcgactctagaggatccccgcacccaactgatcttcagcatctttta
ctttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggа
ataagggcgacacggaaatgttgaatactcatactcttccttttcaatattattgaag
catttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaata
aacaaaaagagtttgtagaaacgcaaaaaggccatccgtcaggatggccttctgcttaa
tttgatgcctggcagtttatggcgggcgtcctgcccgccaccctccgggccgttgcttc
gcaacgttcaaatccgctcccggcggatttgtcctactcaggagagcgttcaccgacaa
acaacagataaaacgaaaggcccagtctttcgactgagcctttcgttttatttgatgcc
tggcagttccctactctcgcatggggagaccccacactaccatcggcgctacggcgttt
cacttctgagttcggcatggggtcaggtgggaccaccgcgctactgccgccaggcaaat
tctgttttatcagaccgcttctgcgttctgatttaatctgtatcaggctgaaaatcttc
tctcatccgccaaaacagccaagctggagaccgtttaaactcaatgatgatgatgatga
tggtcgacggcgctattcagatcctcttctgagatgagttttgttctagaaagcttcg
aattcttatttaagctgggtaaatgcagataatcgttttctggcttcgcgatttgtcgc
ctgcatcaccatccacggactgaacgcccacggcgtggcatcaataccgtgtaatacat
ctgctaaatcacaccattgataatccatcacttcatcatcattgatctgtaacgcacta
gtggtgcgtgcggcaaataccggacacacttcattttccacaatgccactcggatcggt
ggcgcggtagcgaaagtcaggatagatagattcaggaggcgtaatttccacgccaagct
cataacggcaacggcggatcactgcgtcttcgttgctttctcccagttgtgggtgccca
caaaccgagttagtccacacgccaggccatgctttttgctcagtgcgcggcgggtaac
taataattgtcctttggcattaaacagccaactggagaacgcgagatgtaagcgggtgt
ctgccgtgtgtgcggcatacttttccagcgtacccgtgggaactccctgtgcattcaat
aaaatgacgtgttccgtttgcatgcccatggtatatctccttcttaaagttaaacaaaa
ttatttctagttcattttgttgccttaatgagtagcgccaccgcttcacaggcaatccc
ttccccacgtccggtaaatcccagtttttccgtagtagtggctttcacgttaacatcat
ccatatggcagccgagatcttcggcaataaacacgcgcatttgtggaatgtgcggcaac
atcttcggtgcctgagcgatgatagtgacatcgacgttgccaagggtataacccttcgc
ctgaatacgacgccaggcttcgcgtagcagctcgcggctatcggcacctttaaatgccg
gatcggtatccgggaacagcttgccgatatcccccagcgccgccgcgccaagcaatgca
tcggtcaacgcatggagcgccacgtcgccatcagaatgcgccagcaatccttttcgta
aggaatgcgtacgccaccaatgataattgggccttcaccgccaaaggcatgtacgtcaa
aaccgtgtccaattcgcatgcccatggtatatctccttcttaaagttaaacaaaattat
ttctagtttatgtattctcctgatggatggttcgggtgaggtaaaactcggccagtgcc
aaatcttccgggcgcgtgactttaatgttatccgcacggccttcgaccaactgaggatg
gaatccgcaatattccagcgccgaggcttcgtcggtaatagtcgcgccttcatttagag
cgcgcgtcagacagtcatgtaacagctcacgagggaaaaattgcggcgtcagcgcgtgc
cataagccgttgcgatcaacggtatgagcaatggcatttttgcccggttcggcacgttt

catagtatcgcgcactggtgcggcgaggatccccccgtgcggctggtttcgctcaacg
ccaacaatcgcgcgaggtcatcctgatgcaaacaaggacgagcggcgtcatgcaccaat
acccactgcgcgtcgccagcggctttcagacctgccagcacggaatcggcacgctcatc
accgccatctacaacggtgatttgcggatgattcgccagaggaagttgtgcaaaacggc
tatcgccaggacttatggcaatgacgacacgtttcacccggggatgcgccagcagcgca
tgcaccgagtgttcaagaatggtttgattaccgattgagagatattgcttaggacattc
cgtttgcattcgacggccaaatccggccgccggaaccacggcgcaaacatccaaatgag
tggttgccatgcccatggtatatctccttcttaaagttaaacaaaattatttctagttt
atgccagccaggccttgattttggcttccataccagcggcatcgaggccgagttcggcg
cgcatttcttcctgagttccttgcggaataaagaagtccggcaggccaatgttcagcac
gggtactggtttacgatgggccatcagcacttcgttcacgccgctgcctgcgccgccca
taatggcgttttcttctacggtgaccagcgcttcatggctggcggccatttccagaatt
aacgcttcatcaagcggtttcacaaaacgcatatcgaccagcgtggcgttcagcgattc
ggcgactttcgccgcttctggcatcagcgtaccaaagttaaggatcgccagtttctcgc
cacgacgcttcacaatgcctttgccaattggtagtttttccagcggcgtcagttccacg
ccgaccgcgttgccacgcgggtagcgcaccgctgacgggccatcgttatagtgatagcc
ggtatagagcatctggcgacattcgttttcatcgctcggggtcataatgaccatttccg
gtatgcagcgcaggtaagagagatcaaaagcaccctgatgggtttgaccgtcagcacca
acaatgcccgcgcggtcgatggcgaacaggaccggaagcttttgaatcgccacgtcatg
cagcacctgatcataggcgcgttgcaggaaagtggagtaaatcgcgacaatgggtttgt
acccaccaatcgccagacccgcagcaaaggtcaccgcgtgttgctcggcaattgccacg
tcgaagtagcgatccgggaatttacgtgaaaactcgaccatgccggaaccttcacgcat
cgccggagtaatcgccatcagcttgttgtctttcgctgccgtttcgcacaaccagtcgc
caaagatttttgaatagctcggcaaaccgccgctacttttcggcaaacaaccgctggag
ggatcaaatttaggcacggcgtggaaagtgatcgggtctttttctgccggttcataacc
acgacctttttttggtcatgatatgcaggaactgcgggcctttcaggtcgcgcatgttct
ttagcgtggtgataagccccagcacatcgtgaccgtccaccgggccgatgtagttaaag
cccagctcttcaaacaacgtgccaggcactaccatgcctttaatatgttcttcggtgcg
tttgagcagctctttaattggcggcacgccagagaaaacttttttcccgccttcgcgca
gtgaagagtaaagcttaccggaaagcagctgtgccagatggttgttgagcgcgccgaca
ttttcggaaatcgacatttcattgtcgttgagaatcaccagcatatcaggacggatatc
gcccgcgtgattcatcgcttcaaacgccatgcctgcggtaatcgcgccatcgccaatga
cacagacggtgcggcgatttttgccttcttttttcggcagcaaccgcaataccaattccg
gcactgatggaggttgatgaatgcccgacgcttaatacgtcatattcgctttcgccgcg
ccacgggaacgggtgcagaccgcctttctgacggatggtgccgattttgtcgcggcgtc
cggtcaaaattttatgcggataagcctgatgccccacatcccaaatcaattggtcaaac
ggggtgttgtagacatagtgcagcgccacggtcagttcgaccgtgcccagcccggaggc
gaagtgcccgctggaacggctcacgctgtcgagtaaatagcggcgcagttcgtcgcaga
gtttcggtaaactctctttcggcaacagtcgtaactcctgggtggagtcgaccagtgcc
agggtcgggtatttggcaatatcaaaactcatgcccatggtttattcctccttatttaa
tcgatacattaatatatacctctttaatttttaataataaagttaatcgataattccgg
tcgagtgcccacacagattgtctgataaattgttaaagagcagtgccgcttcgcttttt
ctcagcggcgctgtttcctgtgtgaaattgttatccgctcacaattccacacattatac
gagccggatgattaattgtcaacagctcatttcagaatatttgccagaaccgttatgat
gtcggcgcaaaaaacattatccagaacgggagtgcgccttgagcgacacgaattatgca
gtgatttacgacctgcacagccataccacagcttccgatggctgcctgacgccagaagc
attggtgcagggtaccgagctcgaattcactggccgtcgttttacaacgtcgtgactgg
gaaaaccctggcgttacccaacttaatcgccttgcagcacatccccctttcgccagctg

gcgtaatagcgaagaggcccgcaccgatcgcccttcccaacagttgcgcagcctgaatg
gcgaatggcgcctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgc
atatggtgcactctcagtacaatctgctctgatgccgcatagttaagccagccccgaca
cccgccaacacccgctgacgagcttagtaaagccctcgctagattttaatgcggatgtt
gcgattacttcgccaactattgcgataacaagaaaaagccagcctttcatgatatatct
cccaatttgtgtagggcttattatgcacgcttaaaaataataaaagcagacttgacctg
atagtttggctgtgagcaattatgtgcttagtgcatctaacgcttgagttaagccgcgc
cgcgaagcggcgtcggcttgaacgaattgttagacattatttgccgactaccttggtga
tctcgcctttcacgtagtggacaaattcttccaactgatctgcgcgcgaggccaagcga
tcttcttcttgtccaagataagcctgtctagcttcaagtatgacgggctgatactgggc
cggcaggcgctccattgcccagtcggcagcgacatccttcggcgcgattttgccggtta
ctgcgctgtaccaaatgcgggacaacgtaagcactacatttcgctcatcgccagcccag
tcgggcggcgagttccatagcgttaaggtttcatttagcgcctcaaatagatcctgttc
aggaaccggatcaaagagttcctccgccgctggacctaccaaggcaacgctatgttctc
ttgcttttgtcagcaagatagccagatcaatgtcgatcgtggctggctcgaagatacct
gcaagaatgtcattgcgctgccattctccaaattgcagttcgcgcttagctggataacg
ccacggaatgatgtcgtcgtgcacaacaatggtgacttctacagcgcggagaatctcgc
tctctccaggggaagccgaagtttccaaaaggtcgttgatcaaagctcgccgcgttgtt
tcatcaagccttacggtcaccgtaaccagcaaatcaatatcactgtgtggcttcaggcc
gccatccactgcggagccgtacaaatgtacggccagcaacgtcggttcgagatggcgct
cgatgacgccaactacctctgatagttgagtcgatacttcggcgatcaccgcttccctc
atgatgtttaactttgttttagggcgactgccctgctgcgtaacatcgttgctgctcca
taacatcaaacatcgacccacggcgtaacgcgcttgctgcttggatgcccgaggcatag
actgtaccccaaaaaaacagtcataacaagccatgaaaaccgccactgcgccgttacca
ccgctgcgttcggtcaaggttctggaccagttgcgtgagcgcatacgctacttgcatta
cagcttacgaaccgaacaggcttatgtccactgggttcgtgccttcatccgtttccacg
gtgtgcgtcacccggcaaccttgggcagcagcgaagtcgaggcatttctgtcctggctg
gcgaacgagcgcaaggtttcggtctccacgcatcgtcaggcattggcggccttgctgtt
cttctacggcaaggtgctgtgcacggatctgccctggcttcaggagatcggaagacctc
ggccgtcgcggcgcttgccggtggtgctgaccccggatgaagtggttcgcatcctcggt
tttctggaaggcgagcatcgtttgttcgcccagcttctgtatggaacgggcatgcggat
cagtgagggtttgcaactgcgggtcaaggatctggatttcgatcacggcacgatcatcg
tgcgggagggcaagggctccaaggatcgggccttgatgttacccgagagcttggcaccc
agcctgcgcgagcaggggaattaattcccacgggttttgctgcccgcaaacgggctgtt
ctggtgttgctagtttgttatcagaatcgcagatccggcttcagccggtttgccggctg
aaagcgctatttcttccagaattgccatgatttttttccccacgggaggcgtcactggct
cccgtgttgtcggcagctttgattcgataagcagcatcgcctgtttcaggctgtctatg
tgtgactgttgagctgtaacaagttgtctcaggtgttcaatttcatgttctagttgctt
tgttttactggtttcacctgttctattaggtgttacatgctgttcatctgttacattgt
cgatctgttcatggtgaacagctttgaatgcaccaaaaactcgtaaaagctctgatgta
tctatctttttacaccgttttcatctgtgcatatggacagttttccctttgatatgta
acggtgaacagttgttctacttttgtttgttagtcttgatgcttcactgatagatacaa
gagccataagaacctcagatccttccgtatttagccagtatgttctctagtgtggttcg
ttgtttttgcgtgagccatgagaacgaaccattgagatcatacttactttgcatgtcac
tcaaaaattttgcctcaaaactggtgagctgaatttttgcagttaaagcatcgtgtagt
gtttttcttagtccgttatgtaggtaggaatctgatgtaatggttgttggtattttgtc
accattcatttttatctggttgttctcaagttcggttacgagatccatttgtctatcta
gttcaacttggaaaatcaacgtatcagtcgggcggcctcgcttatcaaccaccaatttc

atattgctgtaagtgtttaaatctttacttattggtttcaaaacccattggttaagcct
tttaaactcatggtagttattttcaagcattaacatgaacttaaattcatcaaggctaa
tctctatatttgccttgtgagttttcttttgtgttagttcttttaataaccactcataa
atcctcatagagtatttgttttcaaaagacttaacatgttccagattatattttatgaa
ttttttttaactggaaaagataaggcaatatctcttcactaaaaactaattctaatttttt
cgcttgagaacttggcatagtttgtccactggaaaatctcaaagcctttaaccaaagga
ttcctgatttccacagttctcgtcatcagctctctggttgctttagctaatacaccata
agcattttccctactgatgttcatcatctgagcgtattggttataagtgaacgataccg
tccgttctttccttgtagggttttcaatcgtggggttgagtagtgccacacagcataaa
attagcttggtttcatgctccgttaagtcatagcgactaatcgctagttcatttgcttt
gaaaacaactaattcagacatacatctcaattggtctaggtgattttaatcactatacc
aattgagatgggctagtcaatgataattactagtccttttcctttgagttgtgggtatc
tgtaaattctgctagacctttgctggaaaacttgtaaattctgctagaccctctgtaaa
ttccgctagacctttgtgtgttttttttgtttatattcaagtggttataatttatagaa
taaagaaagaataaaaaaagataaaaagaatagatcccagccctgtgtataactcacta
ctttagtcagttccgcagtattacaaaaggatgtcgcaaacgctgtttgctcctctaca
aaacagaccttaaaaccctaaaggcttaagtagcaccctcgcaagctcgggcaaatcgc
tgaatattcctttttgtctccgaccatcaggcacctgagtcgctgtctttttcgtgacat
tcagttcgctgcgctcacggctctggcagtgaatgggggtaaatggcactacaggcgcc
ttttatggattcatgcaaggaaactacccataatacaagaaaagcccgtcacgggcttc
tcagggcgttttatggcgggtctgctatgtggtgctatctgactttttgctgttcagca
gttcctgccctctgattttccagtctgaccacttcggattatcccgtgacaggtcattc
agactggctaatgcacccagtaaggcagcggtatcatcaacaggctta

We claim:

1. An isolated nucleic acid sequence comprising a sequence encoding an isoprene synthase variant of a parent eucalyptus isoprene synthase, said sequence operably linked to a promoter, wherein the isoprene synthase variant is capable of catalyzing a reaction that synthesizes isoprene from DMADP (dimethylallyl disphosphate) and said isoprene synthase variant is truncated at the N-terminus as compared to the parent eucalyptus isoprene synthase.

2. The nucleic acid sequence of claim 1, wherein the parent eucalyptus isoprene synthase is *E. globulus* of SEQ ID NO: 6.

3. The nucleic acid sequence of any of claims 1-2, wherein said isoprene synthase variant has a sequence that has at least five amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase.

4. The nucleic acid sequence of any of claims 1-3, wherein said isoprene synthase variant has a sequence that has five to thirty five amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase.

5. The nucleic acid sequence of any of claims 1-4, wherein said isoprene synthase variant has a sequence that has ten to thirty five amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase.

6. The nucleic acid sequence of any of claims 1-5, wherein said isoprene synthase variant has a sequence that has fifteen to thirty five amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase.

7. The nucleic acid sequence of any of claims 1-6, wherein said isoprene synthase variant has a sequence that has twenty to thirty five amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase.

8. The nucleic acid sequence of any of claims 1-7, wherein said isoprene synthase variant has a sequence that has twenty five to thirty five amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase.

9. The nucleic acid sequence of any of claims 1-8, wherein said isoprene synthase variant has a sequence that has thirty to thirty five amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase.

10. The nucleic acid sequence of any of claims 1-9, wherein said isoprene synthase variant has a sequence that has thirty five amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase.

11. The nucleic acid sequence of any of claims 1-2, wherein said isoprene synthase variant has a sequence that has less than thirty five amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase.

12. The nucleic acid sequence of any of claims 1-11, wherein the promoter is a prokaryotic promoter.

13. The nucleic acid sequence of any of claims 1-12, wherein the promoter is a pTrc promoter.

14. An expression vector comprising the nucleic acid sequence of any of claims 1-13.

15. An isolated host cell comprising a heterologous nucleic acid sequence according to any of claims 1-13.

16. An isolated host cell comprising a vector according to claim 14.

17. The isolated host cell of claim 15 or 16 being a bacterial cell.

18. The isolated host cell of claim 17, wherein the bacterial cell is *Escherichia coli*.

19. The isolated host cell of any of claims 15-18, wherein the host cell further comprises one or more recombinant nucleic acid sequence of a MEP pathway gene.

20. The isolated host cell of claim 19, wherein the MEP pathway gene is selected from *dxs*, *ispD*, *ispF*, and *idi*.

21. An isolated isoprene synthase variant of a parent eucalyptus isoprene synthase, wherein said variant comprises a truncation in the N-terminal portion of isoprene synthase as compared to the parent eucalyptus isoprene synthase and wherein said variant is capable of catalyzing a reaction that synthesizes isoprene from DMADP (dimethylallyl disphosphate) .

22. A method of producing isoprene comprising: (a) providing a host cell comprising an expression vector according to claim 14; and (b) culturing the host cell under conditions suitable for producing isoprene.

23. The method of claim 22, further comprising (c) recovering the isoprene.

24. The method of claim 23, further comprising (d) polymerizing the isoprene.

25. An isolated nucleic acid sequence comprising a sequence encoding a variant of a parent solanum phellandrene synthase, said sequence operably linked to a promoter, wherein the isoprene synthase variant is capable of catalyzing a reaction that synthesizes isoprene from DMADP (dimethylallyl disphosphate) and said isoprene synthase variant is truncated at the N-terminus as compared to the parent solanum phellandrene synthase.

26. The nucleic acid sequence of claim 25, wherein the parent eucalyptus isoprene synthase is *S. lycopersicum* of SEQ ID NO: 23.

27. The nucleic acid sequence of any of claims 25-26, wherein said variant has a sequence that has at least five amino acids truncated from the N-terminus as compared to the parent solanum phellandrene synthase.

28. The nucleic acid sequence of any of claims 25-27, wherein said variant has a sequence that has five to thirty six amino acids truncated from the N-terminus as compared to the parent solanum phellandrene synthase.

29. The nucleic acid sequence of any of claims 25-28, wherein said variant has a sequence that has ten to thirty six amino acids truncated from the N-terminus as compared to the parent solanum phellandrene synthase.

30. The nucleic acid sequence of any of claims 25-29, wherein said variant has a sequence that has fifteen to thirty six amino acids truncated from the N-terminus as compared to the parent solanum phellandrene synthase.

31. The nucleic acid sequence of any of claims 25-30, wherein said variant has a sequence that has twenty to thirty six amino acids truncated from the N-terminus as compared to the parent solanum phellandrene synthase.

32. The nucleic acid sequence of any of claims 25-31, wherein said variant has a sequence that has twenty five to thirty six amino acids truncated from the N-terminus as compared to the parent solanum phellandrene synthase.

33. The nucleic acid sequence of any of claims 25-32, wherein said variant has a sequence that has thirty to thirty six amino acids truncated from the N-terminus as compared to the parent solanum phellandrene synthase.

34. The nucleic acid sequence of any of claims 25-33, wherein said variant has a sequence that has thirty six amino acids truncated from the N-terminus as compared to the parent solanum phellandrene synthase.

35. The nucleic acid sequence of any of claims 25-26, wherein said variant has a sequence that has less than thirty six amino acids truncated from the N-terminus as compared to the parent eucalyptus isoprene synthase.

36. The nucleic acid sequence of any of claims 25-35, wherein the promoter is a prokaryotic promoter.

37. The nucleic acid sequence of any of claims 25-36, wherein the promoter is a pTrc promoter.

38. An expression vector comprising the nucleic acid sequence of any of claims 25-37.

39. An isolated host cell comprising a heterologous nucleic acid sequence according to any of claims 25-37.

40. An isolated host cell comprising a vector according to claim 38.

41. The isolated host cell of claim 39 or 40 being a bacterial cell.

42. The isolated host cell of claim 41, wherein the bacterial cell is *Escherichia coli*.

43. The isolated host cell of any of claims 39-42, wherein the host cell further comprises one or more recombinant nucleic acid sequence of a MEP pathway gene.

44. The isolated host cell of claim 43, wherein the MEP pathway gene is selected from *dxs*, *ispD*, *ispF*, and *idi*.

45. A variant of a parent solanum phellandrene synthase, wherein said variant comprises a truncation in the N-terminal portion of isoprene synthase as compared to the parent eucalyptus isoprene synthase and wherein said variant is

capable of catalyzing a reaction that synthesizes isoprene from DMADP (dimethylallyl disphosphate) .

46. A method of producing isoprene comprising: (a) providing a host cell comprising an expression vector according to claim 38; and (b) culturing the host cell under conditions suitable for producing isoprene.

47. The method of claim 46, further comprising (c) recovering the isoprene.

48. The method of claim 47, further comprising (d) polymerizing the isoprene.

Figure 1

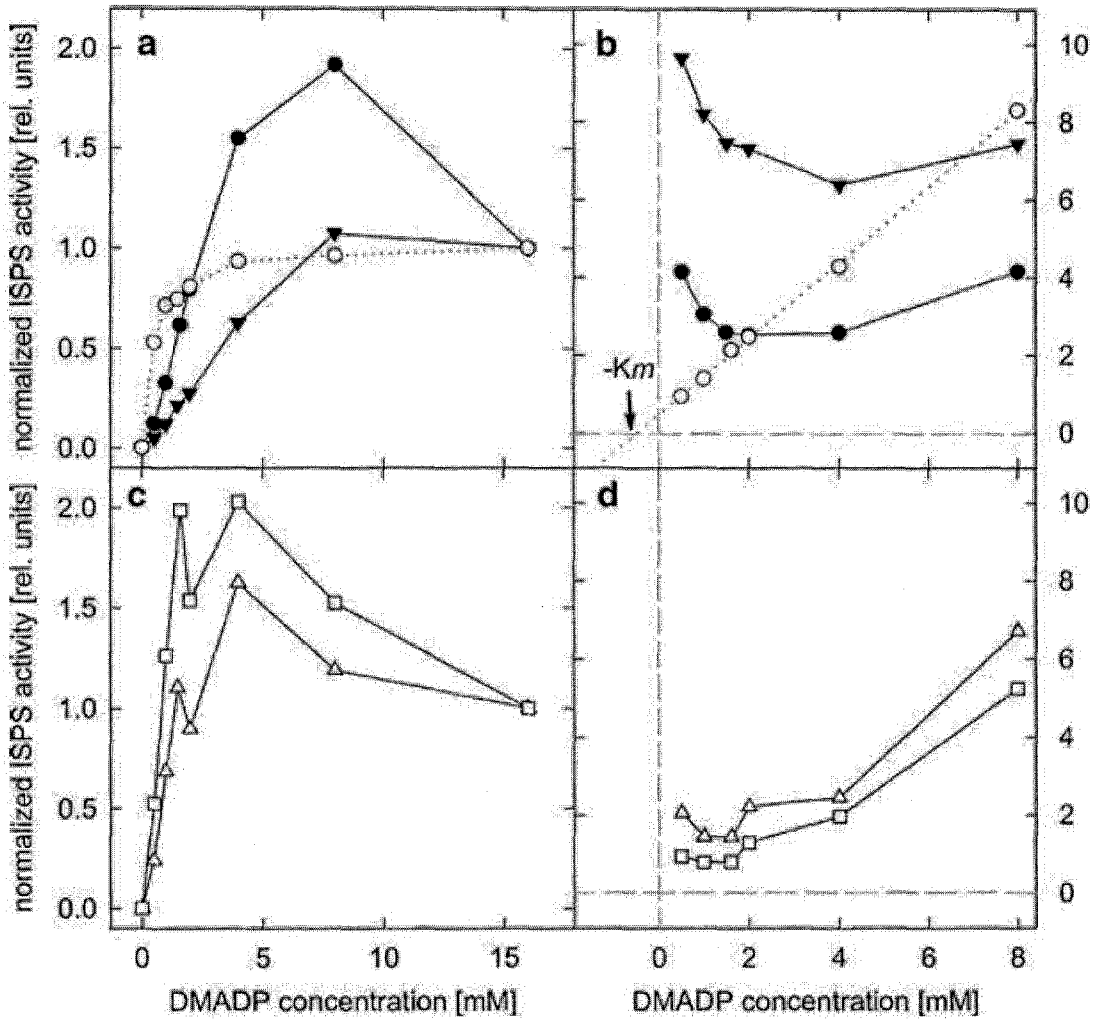

a
b
c
d
normalized ISPS activity [rel. units]
normalized ISPS activity [rel. units]
-Km
DMADP concentration [mM]
DMADP concentration [mM]
0.0
0.5
1.0
1.5
2.0
0
2
4
6
8
10
0
5
10
15

Figure 2

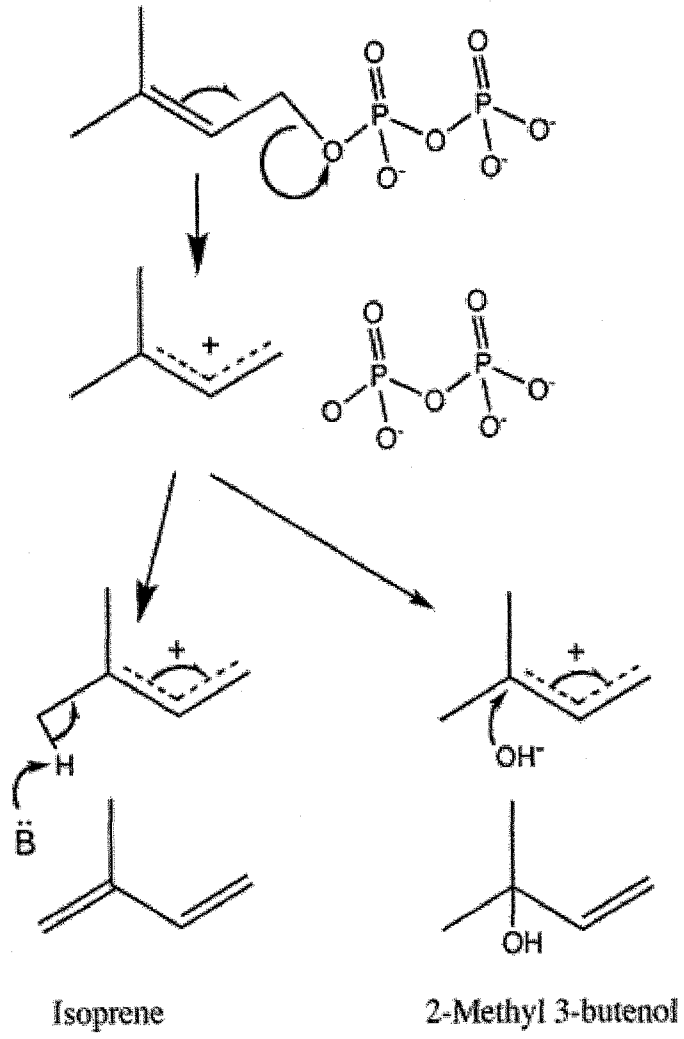


Figure 3

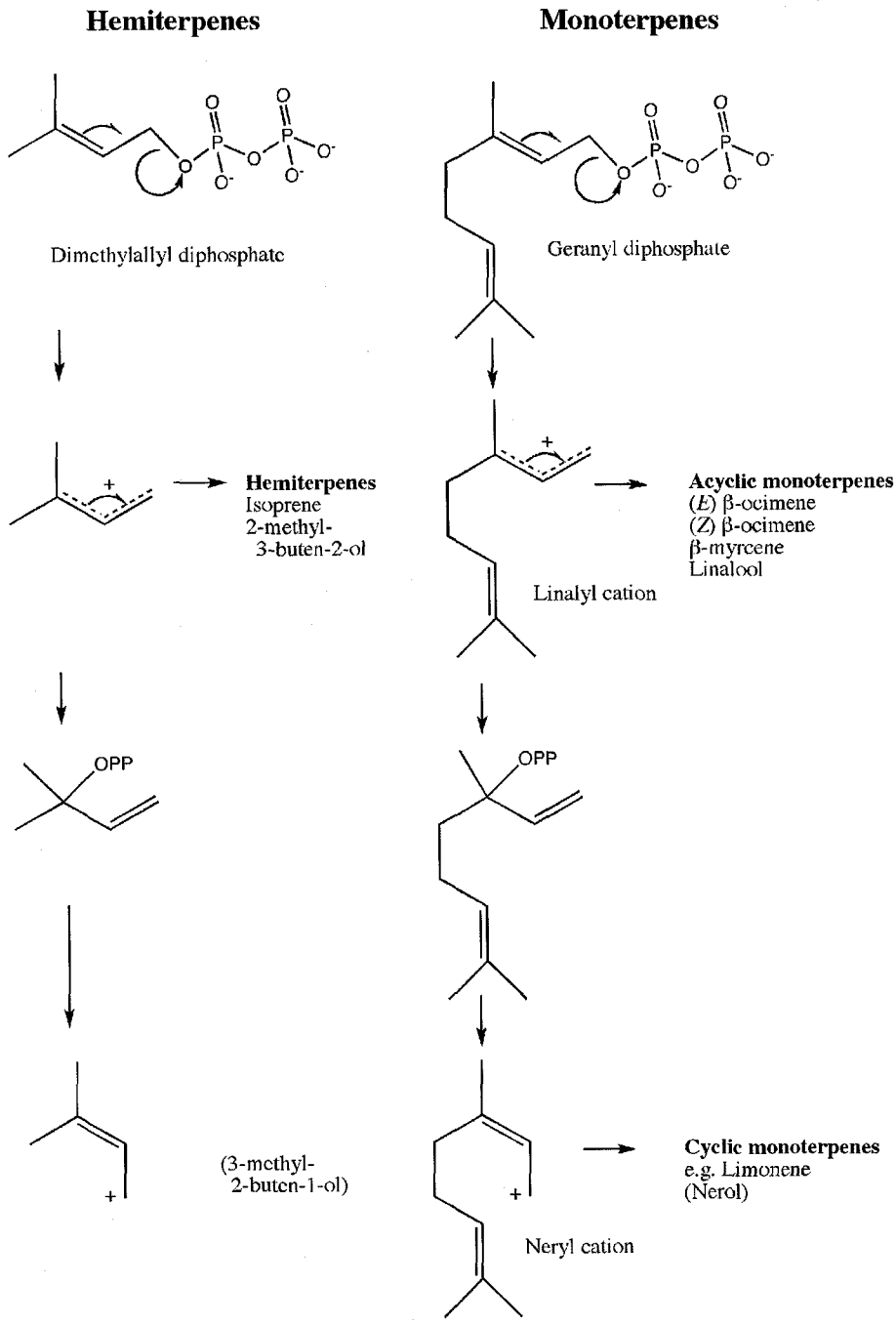

Hemiterpenes
Monoterpenes
Dimethylallyl diphosphate
Geranyl diphosphate
Hemiterpenes
Isoprene
2-methyl-
3-buten-2-ol
Acyclic monoterpenes
(E) β-ocimene
(Z) β-ocimene
β-myrcene
Linalool
Linalyl cation
OPP
OPP
(3-methyl-
2-buten-1-ol)
Cyclic monoterpenes
e.g. Limonene
(Nerol)
Neryl cation

Figure 4

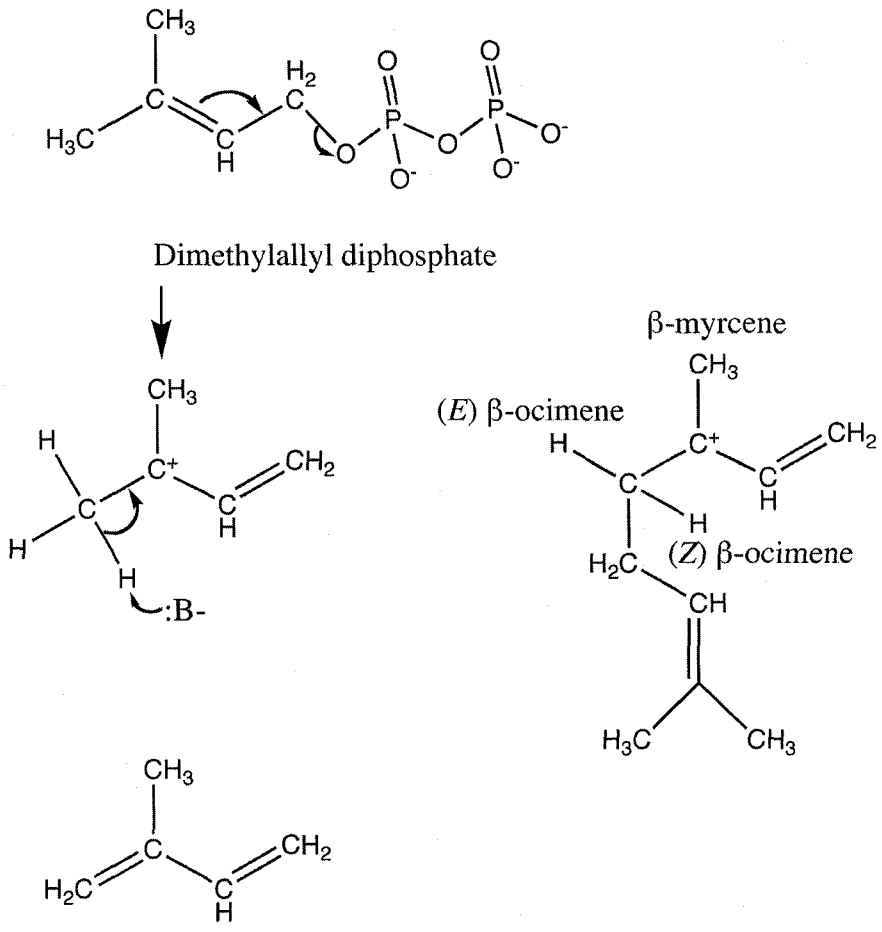

Dimethylallyl diphosphate
:B-
β-myrcene
(E) β-ocimene
(Z) β-ocimene

Figure 5

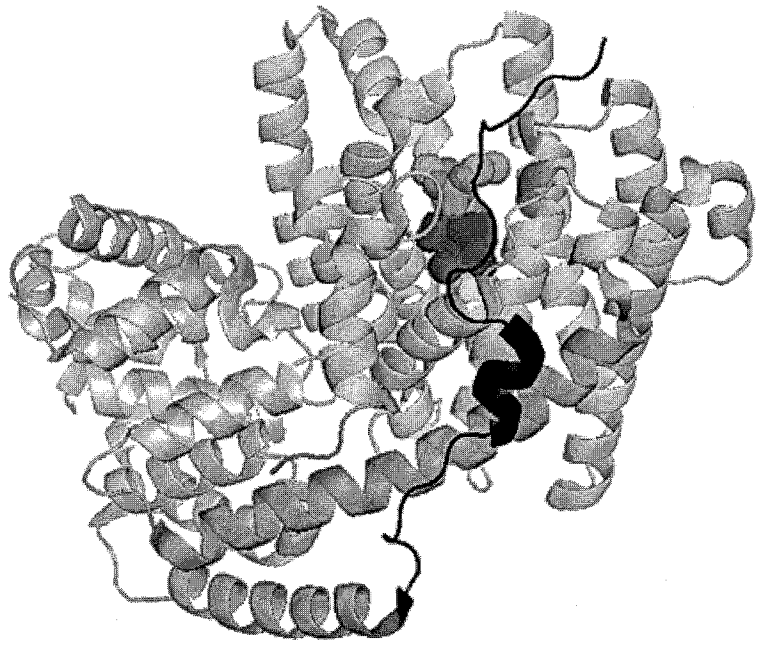


Figure 6

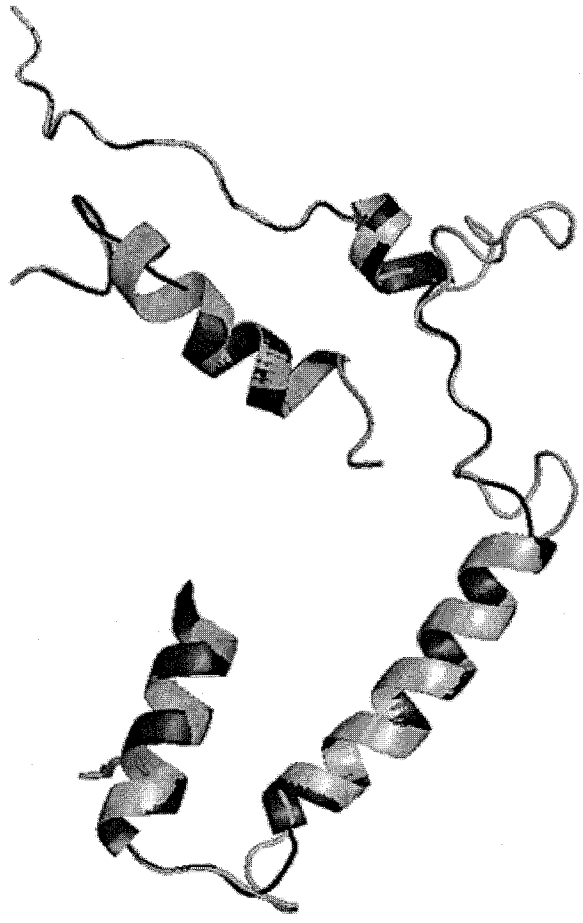


Figure 7

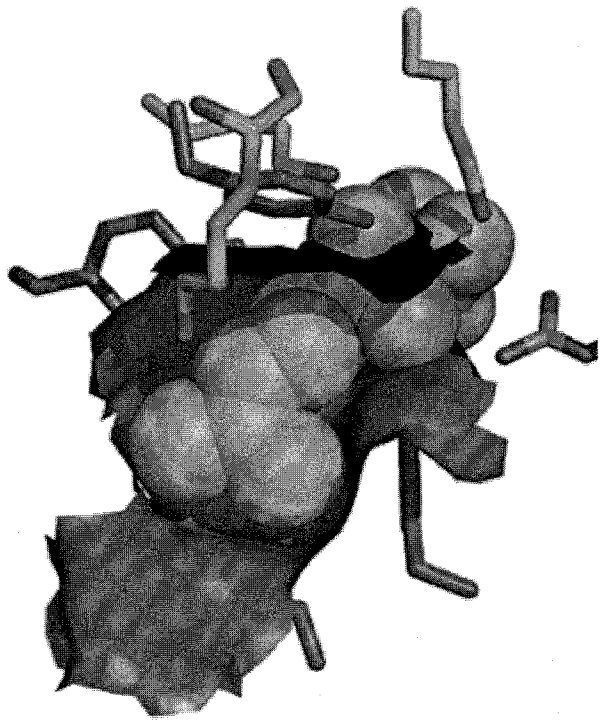


Figure 8

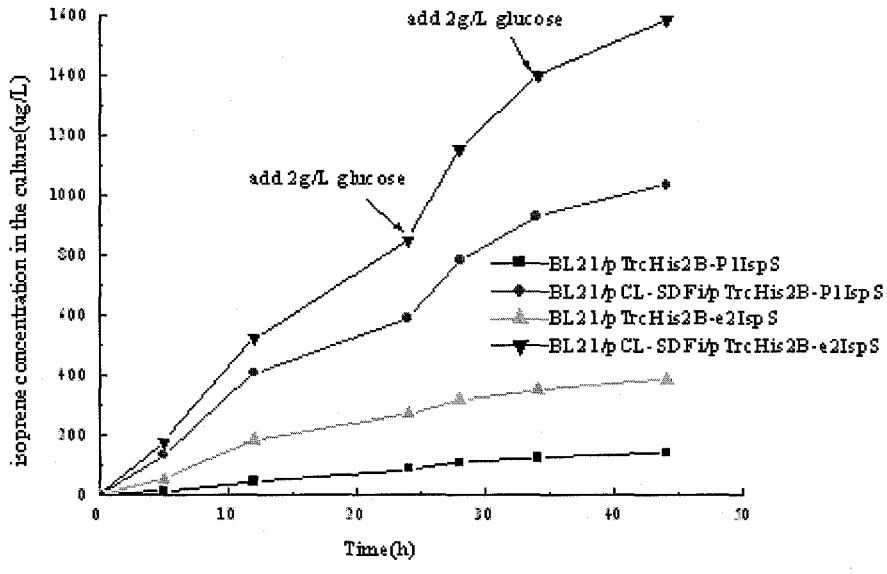

add 2g/L glucose
add 2g/L glucose
isoprene concentration in the culture(ug/L)
Time(h)
BL21/pTrcHis2B-P1IspS
BL21/pCL-SDFi/pTrcHis2B-P1IspS
BL21/pTrcHis2B-e2IspS
BL21/pCL-SDFi/pTrcHis2B-e2IspS

Figure 9

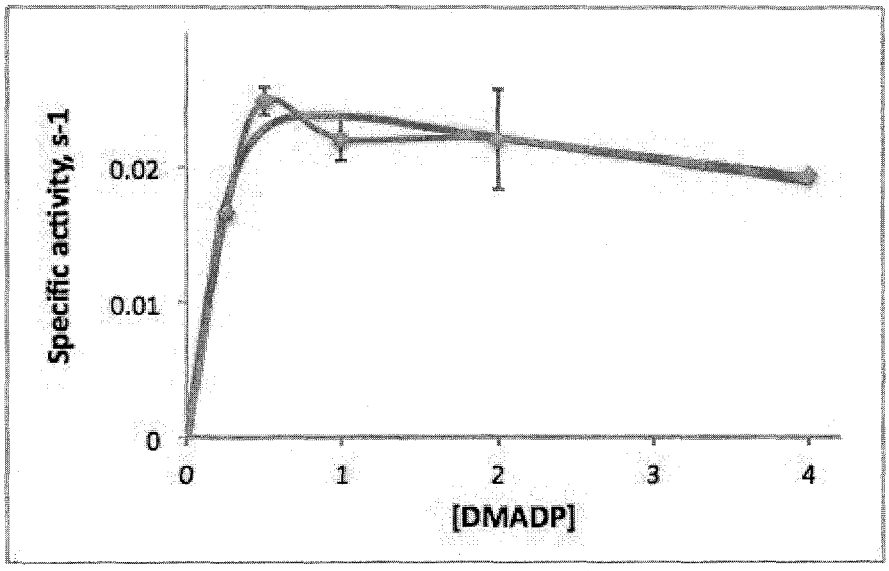


Figure 10

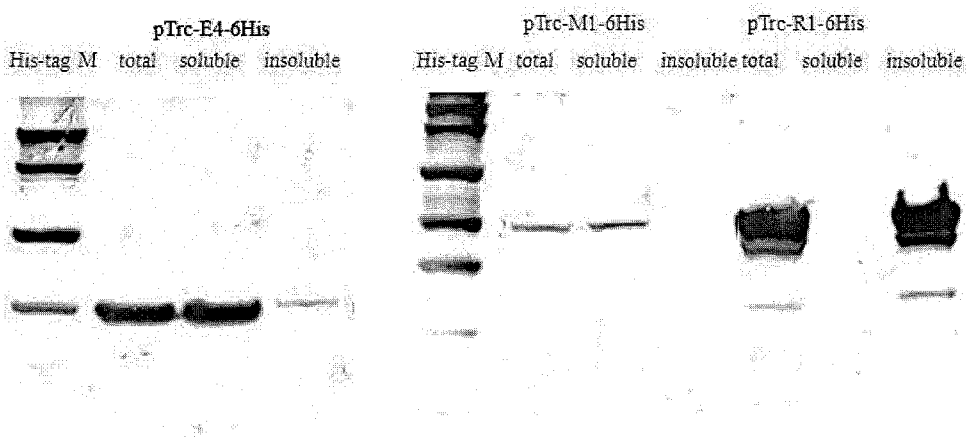
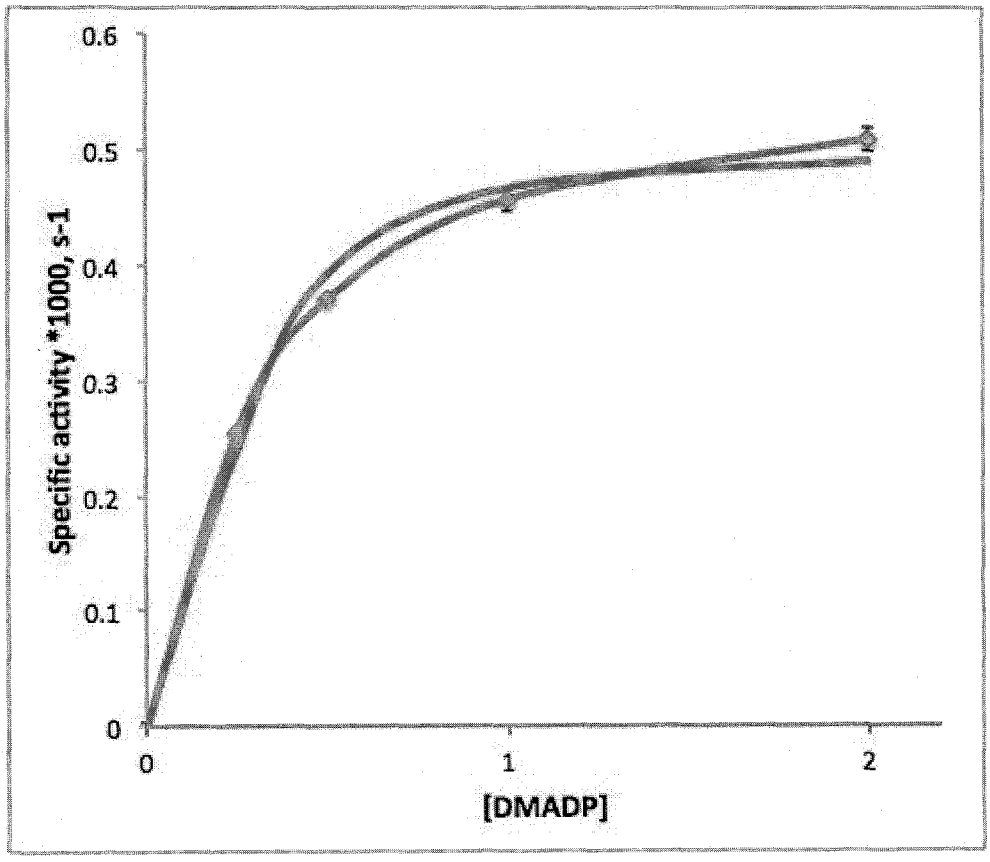

pTrc-E4-6His
His-tag M total soluble insoluble
pTrc-M1-6His
His-tag M total soluble insoluble
pTrc-R1-6His
total soluble insoluble

Figure 11



INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/48422

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12P 5/02; C07H 21/00, 21/04; C12N 15/00, 15/09 (2012.01)
USPC - 435/69.1, 167; 536/23.1, 23.2, 23.6
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8) - C12P 5/02; C07H 21/00, 21/04; C12N 15/00, 15/09 (2012.01)
USPC - 435/69.1, 167; 536/23.1, 23.2, 23.6

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
USPC - 435/471, 320.1, 252.3
(Text Search)

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PubWEST (UPST, EPAB, JPAB, PGPB) and Google Scholar.
Search terms: isoprene synthase (IspS), phellandrene synthase (PHS1), dimethylallyl diphosphate (DMADP or DMAPP), recombinant, heterologous, transit peptide, N-terminal, eucalyptus (e. globulus), tomato (S. lycopersicum), terpene, delet$, truncat$.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	SCHILMILLER, A.L. et al. Monoterpenes in the glandular trichomes of tomato are synthesized from a neryl diphosphate precursor rather than geranyl diphosphate. Proc Nat Acad Sci. 30 June 2009, Vol 106, No 26, Pages 10865-10870: Especially Fig 4A, 4B, 6; Abstract; pg 10865, col 1, para 1; pg 10868 col 1 para 1, para 2; pg 10869, col 2, para 5; Supporting Information, pages 4-5, Fig S3, S4	25-27, 35 and 45
Y	SASAKI, K. et al. Gene expression and characterization of isoprene synthase from Populus alba. FEBS Letters. 07 April 2005, Vol 579, Pages 2514-2518: Especially pg 2515 col 2 para 3; pg 2516 col 2 para 1-2.	1-3, 11 and 21
Y	YANO, S. et al. Eucalyptus globulus mts-1 mRNA for monoterpene synthase, complete cds. GENBANK Submission Accession Number AB266390 [online]. 11 August 2006 [retrieved on 29 September 2012]. Retrieved from the Internet: <URL:http://www.ncbi.nlm.nih.gov/nuccore/AB266390>, pg 1-2	1-3, 11 and 21
Y	US 2010/0003716 A1 (CERVIN et al.) 07 January 2010 (07.01.2010) entire document	1-3, 11, 21, 25-27, 35 and 45
A	US 2008/0038805 A1 (MELIS) 14 February 2008 (14.02.2008) Especially para [0005-0015].	1-3, 11, 21, 25-27, 35 and 45

☐ Further documents are listed in the continuation of Box C. ☐

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
04 October 2012 (04.10.2012)	16 OCT 2012
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-3201	Authorized officer: Lee W. Young PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (July 2009)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 12/48422

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 4-10, 12-20, 22-24, 28-34, 36-44 and 46-48
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.