



(86) Date de dépôt PCT/PCT Filing Date: 2012/11/21  
(87) Date publication PCT/PCT Publication Date: 2013/05/30  
(85) Entrée phase nationale/National Entry: 2014/05/09  
(86) N° demande PCT/PCT Application No.: CA 2012/050837  
(87) N° publication PCT/PCT Publication No.: 2013/075239  
(30) Priorité/Priority: 2011/11/21 (US61/562,280)

(51) Cl.Int./Int.Cl. *C12N 15/60* (2006.01),  
*C12N 9/88* (2006.01), *C12P 41/00* (2006.01),  
*C12P 5/00* (2006.01), *C12P 7/02* (2006.01)  
(71) Demandeur/Applicant:  
THE UNIVERSITY OF BRITISH COLUMBIA, CA  
(72) Inventeurs/Inventors:  
BOHLMANN, JOERG, CA;  
ZERBE, PHILIPP, CA  
(74) Agent: GOWLING LAFLEUR HENDERSON LLP

(54) Titre : DITERPENES SYNTHASES ET PROCEDE DE PRODUCTION DE DITERPENOIDES  
(54) Title: DITERPENE SYNTHASES AND METHOD FOR PRODUCING DITERPENOIDS

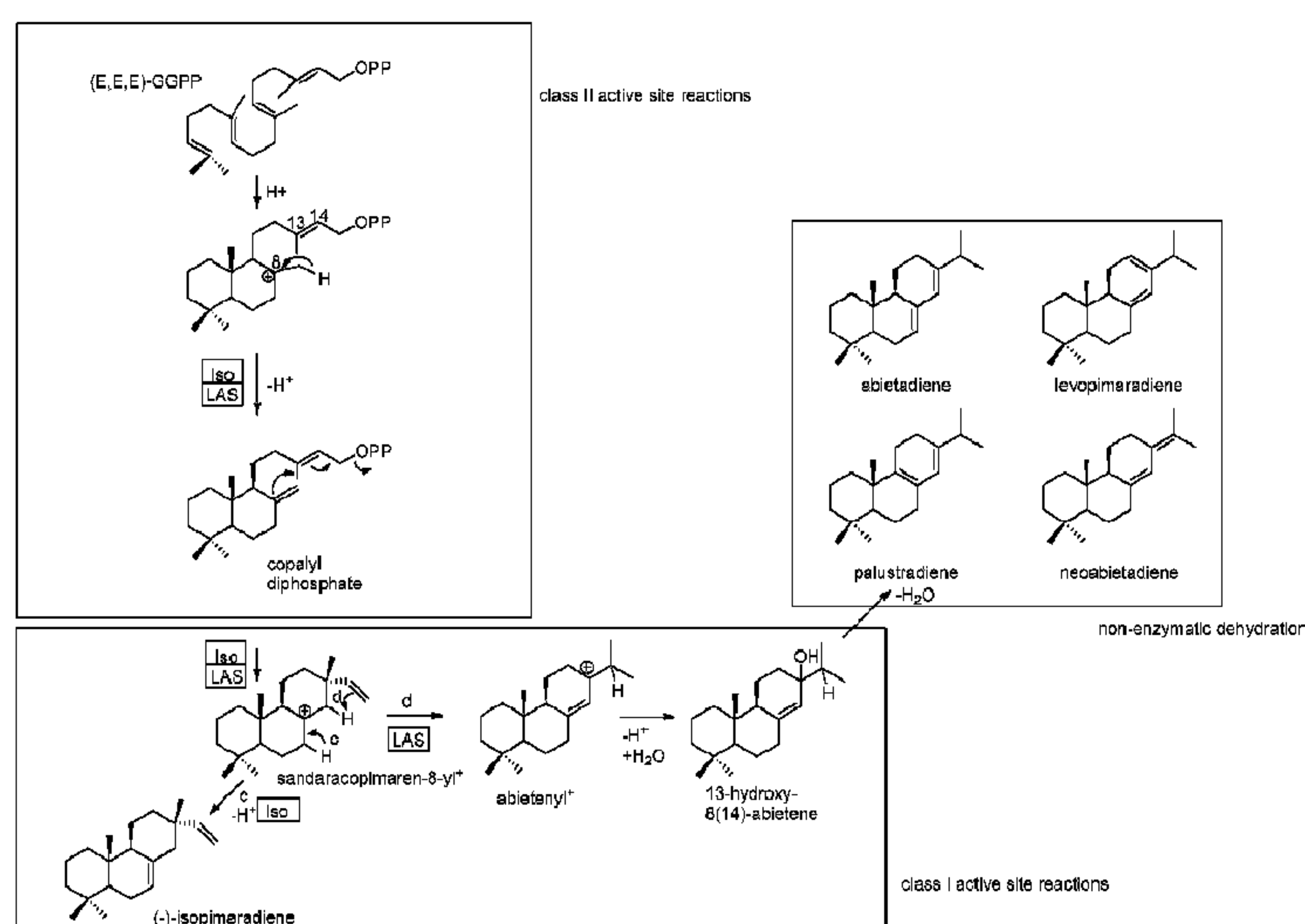


FIGURE 1A

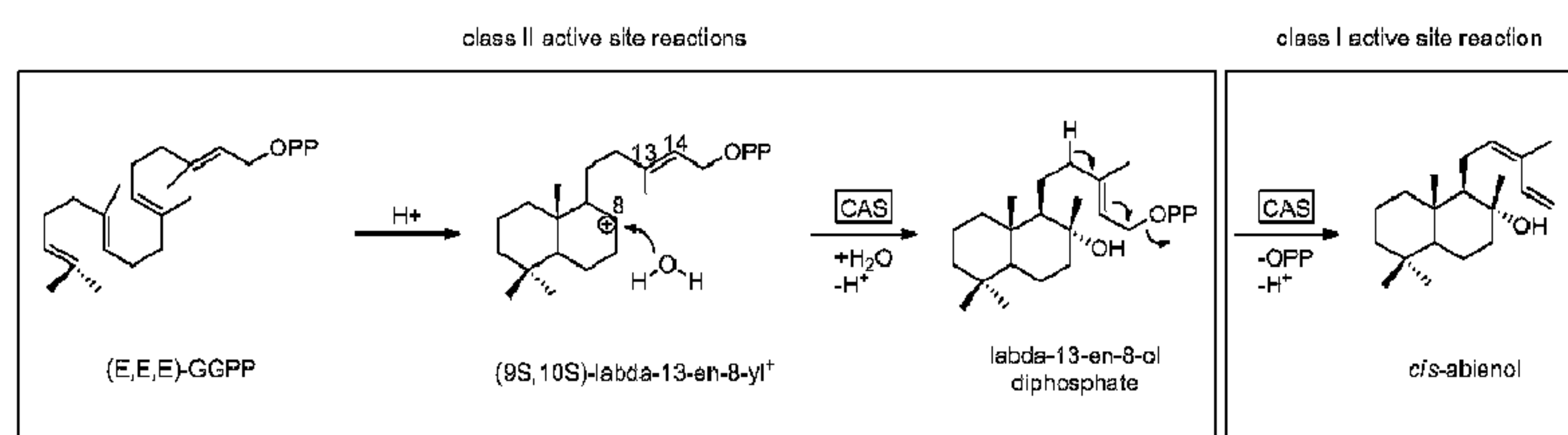


FIGURE 1B

(57) **Abrégé/Abstract:**

Provided herein are diterpene synthases (diTPS) and methods for producing diterpenoids. Also provided herein are nucleic acid sequences encoding diTPS, diTPS amino acid sequences, diTPS proteins, vectors, cells, transgenic organisms, uses, compositions, methods, processes, and kits thereof.



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

(19) World Intellectual Property  
Organization  
International Bureau(10) International Publication Number  
**WO 2013/075239 A1**(43) International Publication Date  
30 May 2013 (30.05.2013)

## (51) International Patent Classification:

C12N 15/60 (2006.01) C12P 5/00 (2006.01)  
C12N 9/88 (2006.01) C12P 7/02 (2006.01)  
C12P 41/00 (2006.01)

## (21) International Application Number:

PCT/CA2012/050837

## (22) International Filing Date:

21 November 2012 (21.11.2012)

## (25) Filing Language:

English

## (26) Publication Language:

English

## (30) Priority Data:

61/562,280 21 November 2011 (21.11.2011) US

(71) Applicant: **THE UNIVERSITY OF BRITISH COLUMBIA** [CA/CA]; #103-6190 Agronomy Road, Vancouver, British Columbia V6T 1Z3 (CA).(72) Inventors: **BOHLMAN, Joerg**; 3706 W29th Street, Vancouver, British Columbia V6S 1T4 (CA). **ZERBE, Philipp**; 3732 Mount Seymour Parkway, North Vancouver, British Columbia V7G 1C3 (CA).(74) Agents: **SECHLEY, Konrad** et al.; Suite 2300, 550 Burrard Street, Vancouver, British Columbia V6C 2B5 (CA).

(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, PA, 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, 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).

[Continued on next page]

## (54) Title: DITERPENE SYNTHASES AND METHOD FOR PRODUCING DITERPENOIDS

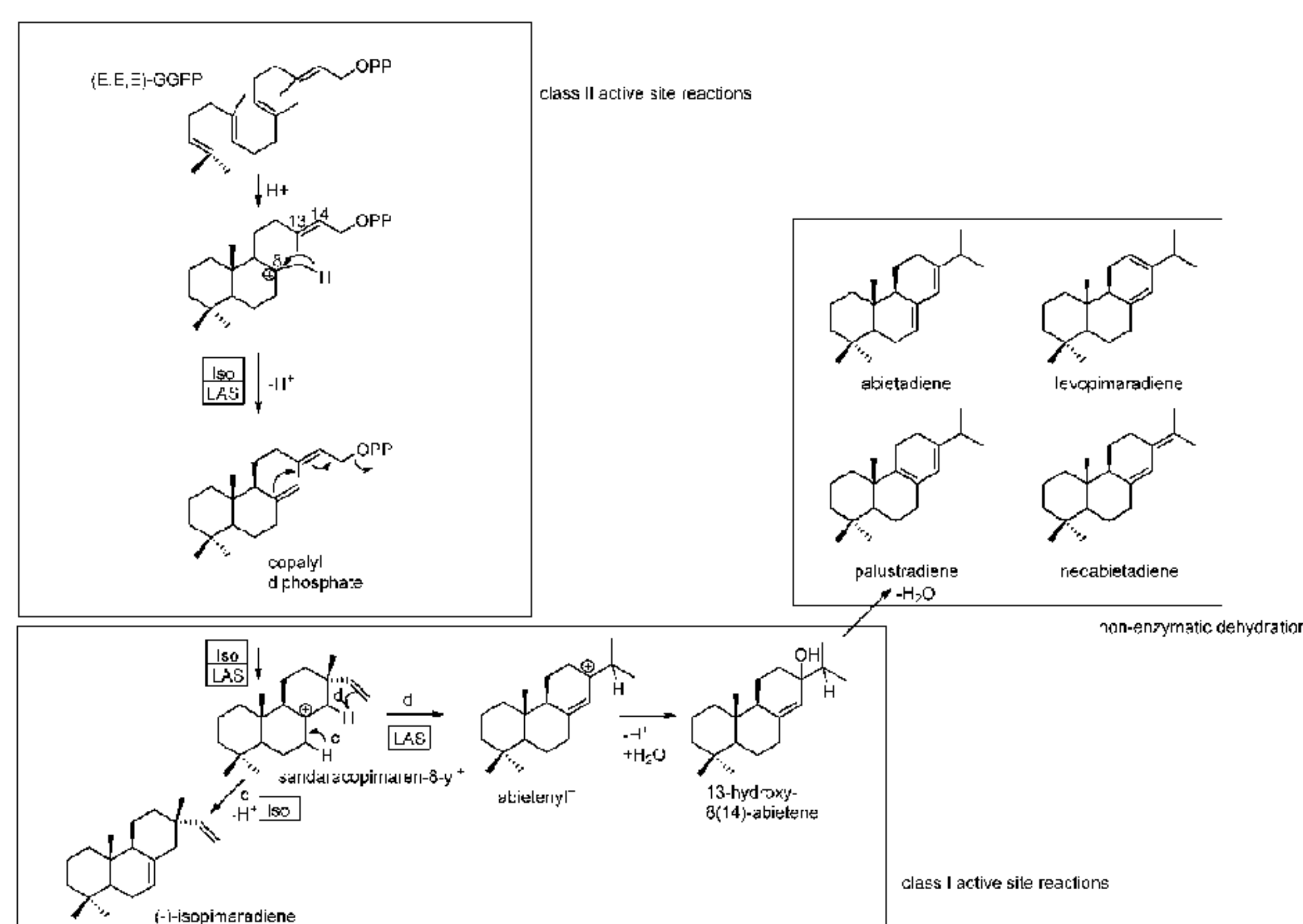


FIGURE 1A

(57) Abstract: Provided herein are diterpene synthases (diTPS) and methods for producing diterpenoids. Also provided herein are nucleic acid sequences encoding diTPS, diTPS amino acid sequences, diTPS proteins, vectors, cells, transgenic organisms, uses, compositions, methods, processes, and kits thereof.

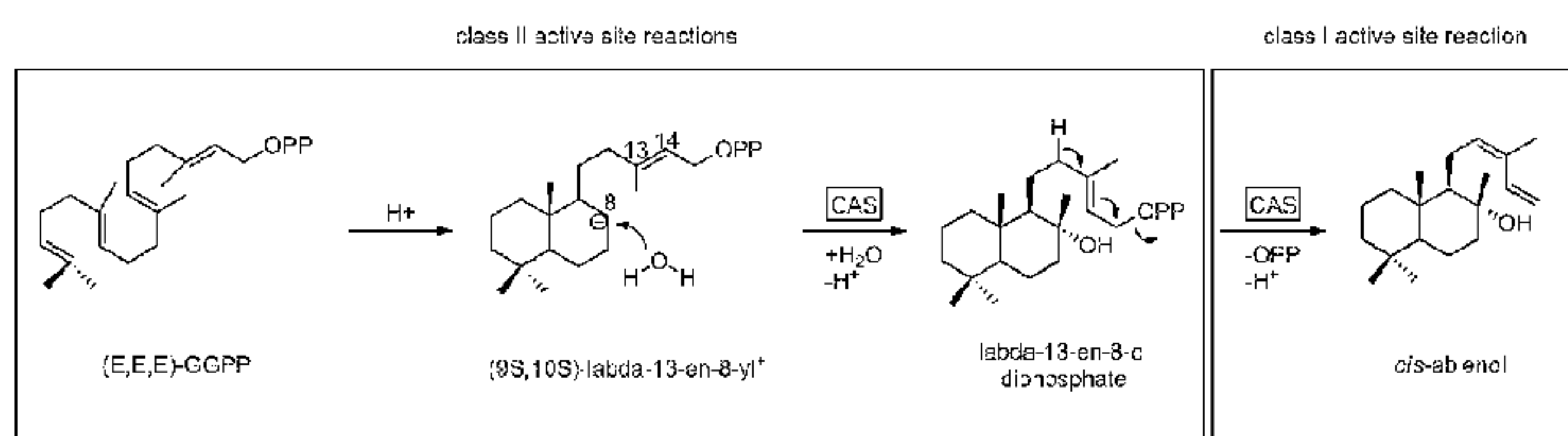


FIGURE 1B

# WO 2013/075239 A1

---

**Published:**

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

— *with sequence listing part of description (Rule 5.2(a))*

- 1 -

## **DITERPENE SYNTHASES AND METHOD FOR PRODUCING DITERPENOIDS**

### **RELATED APPLICATIONS**

Benefit of priority is claimed to U.S. Provisional Application Serial No.  
5 61/562,280, filed November 21, 2011, entitled "Diterpene Synthases and Method for  
Producing Diterpenoids." The subject matter of the above-noted application is  
incorporated by reference in its entirety.

This application is related to United States Patent Application No. (Attorney  
Dkt. No. 33322-00234.US01/234), filed the same day herewith, entitled "Diterpene  
10 Synthases and Method for Producing Diterpenoids," which claims priority to U.S.  
Provisional Application Serial No. 61/562,280.

The subject matter of each of the above-noted applications is incorporated by  
reference in its entirety.

### **Incorporation by reference of sequence listing provided electronically**

15 An electronic version of the Sequence Listing is filed herewith, the contents of  
which are incorporated by reference in their entirety. The electronic file is 200  
kilobytes in size, and titled 234SEQPC1.txt.

### **FIELD OF THE INVENTION**

Provided herein are diterpene synthases (diTPS) and methods for producing  
20 diterpenoids. Also provided herein are nucleic acid sequences encoding diTPS,  
diTPS amino acid sequences, diTPS proteins, vectors, cells, transgenic organisms,  
uses, compositions, methods, processes, and kits thereof.

### **BACKGROUND**

Conifers produce a diverse array of diterpenoids as major oleoresin  
25 components that play a role in the chemical defense against herbivores and pathogens,  
such as bark beetles and their associated fungi (Trapp *et al.* (2001) *Annu. Rev. Plant  
Physiol. Plant Mol. Biol.*, 52:689-724; Keeling *et al.* (2006) *Phytochemistry*, 67:2415-  
2423; Keeling *et al.* (2006) *New Phytol.*, 170:657-675). Oleoresin diterpenoids are  
also used as large-volume, renewable raw material for the production of a suite of  
30 industrial resins and coatings, and other bioproducts (Bohlmann *et al.* (2008) *Plant J.*,  
54:656-669; Hillwig *et al.* (2011) *Biopolymer*, 95:71-76). There is a need for  
alternative methods for producing diterpenoids.

- 2 -

**SUMMARY**

Provided herein are methods for producing diterpenoids. Also provided herein are nucleic acids encoding diterpene synthase (diTPS), diTPS amino acid sequences, and vectors, hosts containing diTPS and that can be used to produce diterpenoids.

5 More specifically, the methods provided herein relate to the production of bicyclic tertiary diterpene alcohols such as for example *cis*-abienol and the production of diterpene olefins such as for example abietadiene, levopimaradiene, palustradiene or neoabietadiene. The production can be *in vivo* or *in vitro*. Also provided herein are methods for expression of an enzyme with diTPS activity in a host organism.

10 Also provided herein are diTPS nucleic acid sequences, diTPS amino acid sequences, proteins, vectors, cells, transgenic organisms, uses, compositions, methods, processes, and kits thereof. For example, a diterpene synthase (diTPS) polypeptide having the sequence set forth in SEQ ID NO:1; SEQ ID NO:3, SEQ ID NO: 5, SEQ ID NO: 7, an active fragment thereof, or sequence substantially identical  
 15 thereto are provided. Furthermore, provided herein is a nucleic acid encoding the polypeptide sequence set forth in SEQ ID NO:1; SEQ ID NO:3, SEQ ID NO: 5 or SEQ ID NO: 7. The nucleic acid can have the sequence set forth in SEQ ID NO: 2; SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8 or a sequence substantially identical thereto. Also provided herein is a nucleic acid encoding the polypeptide sequence as  
 20 set forth in SEQ ID NO: 7, an active fragment thereof, or sequence substantially identical thereto, wherein the nucleotide at position 621 is replaced by A, the nucleotide at position 402 is replaced by A, the nucleotide at position 404 is replaced by A or a combination thereof. Also provided herein is a diterpene synthase with an active site as disclosed in Figure 6.

25 Provided herein is a method for producing diterpenoids that includes providing a host containing a nucleotide sequence encoding a diterpene synthase (diTPS), the nucleotide sequence operatively linked with a regulatory region that is active in the host, and growing the host thereby expressing the nucleotide sequence and producing diterpenoids.

30 Also provided herein is a method for producing one or more diterpenoids by introducing into a host capable of producing (E,E,E)-geranylgeranyl diphosphate (GGPP) one or more nucleotide sequence encoding a diterpene synthase (diTPS), the

- 3 -

nucleotide sequence operatively linked with a regulatory region that is active in the host, and growing the host thereby expressing the nucleotide sequence and producing diterpenoids. The diTPS can contain a bifunctional class I/II diTPS.

Also provided is a method for producing one or more diterpenoids by  
5 contacting (E,E,E)-geranylgeranyl diphosphate (GGPP) with at least one polypeptide having a diterpene synthase (diTPS) activity under conditions effective to produce diterpenoids. Optionally, the diterpenoids produced can be isolated. The diTPS can contain a bifunctional class I/II diTPS.

Also provided herein is a method for producing diterpenoids by contacting  
10 (E,E,E)-geranylgeranyl diphosphate (GGPP) with at least one polypeptide having a diterpene synthase (diTPS) activity under conditions effective to produce diterpenoids and optionally, isolating the produced diterpenoids.

In one example, provided herein is an isolated nucleic acid molecule containing a sequence of nucleotides encoding a bifunctional class I/II cis-abienol  
15 synthase (CAS) polypeptide or an active fragment, wherein the encoded polypeptide or active fragment comprises a class II active site comprising a DxDD motif and a class I active site comprising a DDxxD motif; and the encoded polypeptide or active fragment catalyzes the formation of cis-abienol from geranylgeranyl diphosphate (GGPP). For example, the isolated nucleic acid molecule has a sequence of  
20 nucleotides encoding a cis-abienol synthase (CAS) polypeptide set forth in SEQ ID NO:7, an active fragment thereof, or a sequence that exhibits at least 75% sequence identity to SEQ ID NO:7, wherein the encoded polypeptide or active fragment catalyzes the formation of cis-abienol from geranylgeranyl diphosphate (GGPP). The sequence of nucleotides that encodes a CAS polypeptide exhibits at least 80%, 85%,  
25 90%, 95% or 98% sequence identity to SEQ ID NO:7. In some examples, the sequence of nucleotides encodes a polypeptide that has the sequence of amino acids set forth in SEQ ID NO:7 or an active fragment thereof. In other examples, the sequence of nucleotides encodes a polypeptide that is the sequence of amino acids set forth in SEQ ID NO:7 or an active fragment thereof. In any of the above examples,  
30 the active fragment is a pseudomature form.

For example, in any of the examples of an isolated nucleic acid molecule encoding a bifunctional class I/II cis-abienol synthase (CAS) polypeptide or an active

- 4 -

fragment, the nucleic acid molecule contains the sequence of nucleotides set forth in SEQ ID NO:8 or a sequence of nucleotides that exhibits at least 75% sequence identity to SEQ ID NO:8 or the complement thereof, wherein the sequence of nucleotides encodes a polypeptide or active fragment that catalyzes the formation of *cis*-abienol from geranylgeranyl diphosphate (GGPP). For example, the sequence of nucleotides exhibits at least 80%, 85%, 90%, 95% or 98% sequence identity to SEQ ID NO:8 or the complement thereof. In some examples, the nucleic acid molecule contains the sequence of nucleotides set forth in SEQ ID NO:8, an active fragment thereof or the complement thereof. In other examples, the nucleic acid molecule has the sequence of nucleotides set forth in SEQ ID NO:8, an active fragment thereof or the complement thereof.

Also provided herein is a *cis*-abienol synthase (CAS) polypeptide encoded by any of the nucleic acid molecules provided herein above encoding a bifunctional class I/II *cis*-abienol synthase (CAS) polypeptide or an active fragment thereof.

Also provided herein is a vector containing any of the nucleic acid molecules provided herein above encoding a bifunctional class I/II *cis*-abienol synthase (CAS) polypeptide or an active fragment thereof. Also provided is a host cell containing such a vector. The host cell can be a prokaryotic or eukaryotic host cell. For example, the host cell is a bacterial, fungal, plant, insect, amphibian or animal cell, such as an *E. coli* cell or a yeast cell. In examples herein, the host cell is one that produces GGPP. For example, the host cell is transformed with one or more sequences that result in production of GGPP.

Also provided herein are method for producing a bicyclic tertiary diterpene alcohol, by i) contacting (E,E,E)-geranylgeranyl diphosphate (GGPP) with a bifunctional class I/II synthase polypeptide having a diterpene synthase (diTPS) activity under conditions effective to produce bicyclic tertiary diterpene alcohols, wherein the bifunctional class I/II synthase is a conifer synthase; and ii) optionally, isolating the diterpenoids produced in step i). Also provided herein are methods for producing *cis*-abienol, by i) contacting (E,E,E)-geranylgeranyl diphosphate (GGPP) with a bifunctional class I/II synthase polypeptide having a diterpene synthase (diTPS) activity under conditions effective to produce *cis*-abienol; ii) optionally, isolating the *cis*-abienol produced in step i). In the above provided methods, the

- 5 -

bifunctional class I/II synthase has a class II active site containing a DxDD motif and a class I active site containing a DDxxD motif. In particular examples of any of the above provided methods, the bifunctional class I/II synthase is a *cis*-abienol synthase (CAS) polypeptide or active fragment that catalyzes the formation of *cis*-abienol from geranylgeranyl diphosphate (GGPP).

For example, in the above provided methods, the bifunctional class I/II synthase is a CAS polypeptide or active fragment that contains the sequence of amino acids set forth in SEQ ID NO:7, an active fragment thereof, or a sequence that exhibits at least 75% sequence identity to SEQ ID NO:7, wherein the encoded polypeptide or active fragment catalyzes the formation of *cis*-abienol from geranylgeranyl diphosphate (GGPP). For example, the sequence of amino acids exhibits at least 80%, 85%, 90%, 95% or 98% sequence identity to SEQ ID NO:7. In some examples, the CAS polypeptide contains the sequence of amino acids set forth in SEQ ID NO:7 or an active fragment thereof. In other examples, the CAS polypeptide is the sequence of amino acids set forth in SEQ ID NO:7 or an active fragment thereof. In any of the above example, the the active fragment is a pseudomature form.

In any of the above examples of methods herein, the CAS polypeptide is encoded by the sequence of nucleotides set forth in SEQ ID NO:8 or a sequence of nucleotides that exhibits at least 75% sequence identity to SEQ ID NO:8 or the complement thereof, wherein the sequence of nucleotides encodes a polypeptide or active fragment that catalyzes the formation of *cis*-abienol from geranylgeranyl diphosphate (GGPP). For example, the sequence of nucleotides exhibits at least 80%, 85%, 90%, 95% or 98% sequence identity to SEQ ID NO:8 or the complement thereof. In some examples, the sequence of nucleotides contains the sequence of nucleotides set forth in SEQ ID NO:8, an active fragment thereof or the complement thereof. In other examples, the sequence of nucleotides is set forth in SEQ ID NO:8, an active fragment thereof or the complement thereof.

In any of the above method of producing a bicyclic tertiary diterpene alcohol or *cis*-abienol, the step of contacting E,E,E)-geranylgeranyl diphosphate (GGPP) with a bifunctional class I/II synthase polypeptide is effected *in vitro* or *in vivo*. In

- 6 -

examples of methods herein, the bicyclic tertiary diterpene alcohol or *cis*-abienol is further isolated. For example, the *cis*-abienol is used to produce (-)-ambroxide.

Also provided herein is an isolated nucleic acid molecule containing the sequence of nucleotides encoding a levopimaradiene/abietadiene synthase (LAS) set forth in SEQ ID NO: 1, an active fragment thereof, or a sequence that exhibits greater than 99% sequence identity to SEQ ID NO:1, wherein the encoded polypeptide or active fragment catalyzes the formation of epimers of 13-hydroxy-8(14)-abietene from geranylgeranyl diphosphate (GGPP), such as any one or more of abietadine, levopimaradiene, neoabietadiene and plaustradiene. For example, the sequence of nucleotides encodes a polypeptide that contains the sequence of amino acids set forth in SEQ ID NO: 1 or an active fragment thereof. In other examples, the sequence of nucleotides encodes a polypeptide that has the sequence of amino acids set forth in SEQ ID NO:1 or an active fragment thereof. In any of the above examples, the active fragment is a pseudomature form.

For example, in any of the examples of an isolated nucleic acid molecule encoding a levopimaradiene/abietadiene synthase (LAS), the nucleic acid molecule contains the sequence of nucleotides set forth in SEQ ID NO:2 or a sequence of nucleotides that exhibits greater than 99% sequence identity to SEQ ID NO:2 or the complement thereof, wherein the sequence of nucleotides encodes a polypeptide or active fragment that catalyzes the formation of epimers of 13-hydroxy-8(14)-abietene from geranylgeranyl diphosphate (GGPP). In one example, the isolated nucleic acid molecule contains the sequence of nucleotides set forth in SEQ ID NO:2, an active fragment thereof or the complement thereof. In another example, the isolated nucleic acid molecule has the sequence of nucleotides set forth in SEQ ID NO:1, an active fragment thereof or the complement thereof.

Also provided herein is a levopimaradiene/abietadiene synthase (LAS) polypeptide encoded by any of the nucleic acid molecules provided above encoding a levopimaradiene/abietadiene synthase (LAS) or an active fragment thereof.

Also provided herein is a vector containing any of the nucleic acid molecules provided herein above encoding a levopimaradiene/abietadiene synthase (LAS) or an active fragment thereof. Also provided is a host cell containing such a vector. The host cell can be a prokaryotic or eukaryotic host cell. For example, the host cell is a

- 7 -

bacterial, fungal, plant, insect, amphibian or animal cell, such as an *E. coli* cell or a yeast cell. In examples herein, the host cell is one that produces GGPP. For example, the host cell is transformed with one or more sequences that result in production of GGPP.

5           Also provided herein is a method of producing an epimer of 13-hydroxy-8(14)-abietene, such as any one or more of abietadine, levopimaradiene, neoabietadiene and plaustadiene, by i) contacting (E,E,E)-geranylgeranyl diphosphate (GGPP) with a levopimaradiene/abietadiene synthase (LAS) polypeptide encoded by any of the nucleic acid molecules provided herein encoding a  
10   levopimaradiene/abietadiene synthase (LAS) or an active fragment thereof under conditions effective to produce an epimer of 13-hydroxy-8(14)-abietene; and ii) optionally, isolating the epimer of 13-hydroxy-8(14)-abietene produced in step i), such as any one or more of abietadine, levopimaradiene, neoabietadiene and plaustadiene. In examples of the methods provided herein, the step of contacting  
15   E,E,E)-geranylgeranyl diphosphate (GGPP) with a levopimaradiene/abietadiene synthase (LAS) polypeptide is effected *in vitro* or *in vivo*. In examples of the methods herein, the method further includes isolating one or more of the diterpenes selected from among abietadine, levopimaradiene, neoabietadiene and plaustadiene.

          Also provided herein is an isolated nucleic acid molecule containing the  
20   sequence of nucleotides encoding a isopimaradiene synthase (Iso) polypeptide set forth in SEQ ID NO: 3, an active fragment thereof, or a sequence that exhibits at least 95% sequence identity to SEQ ID NO:3, wherein the encoded polypeptide or active fragment catalyzes the formation of isopimaradiene from geranylgeranyl diphosphate (GGPP). For example, the sequence of nucleotides encodes a polypeptide that  
25   exhibits at least 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:3. In one example, the sequence of nucleotides encodes a polypeptide that contains the sequence of amino acids set forth in SEQ ID NO: 3 or an active fragment thereof. In another example, the sequence of nucleotides encodes a polypeptide that has the sequence of amino acids set forth in SEQ ID NO:3 or an active fragment thereof. In  
30   any of the above examples, the active fragment is a pseudomature form.

          For example, in any of the examples of an isolated nucleic acid molecule encoding isopimaradiene synthase (Iso) polypeptide, the nucleic acid molecule

- 8 -

contains the sequence of nucleotides set forth in SEQ ID NO:4 or a sequence of nucleotides that exhibits at least 95% sequence identity to SEQ ID NO:4 or the complement thereof, wherein the sequence of nucleotides encodes a polypeptide or active fragment that catalyzes the formation of isopimaradiene from geranylgeranyl diphosphate (GGPP). For example, the sequence of nucleotides exhibits at least 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:4. In one example, the isolated nucleic acid molecule contains the sequence of nucleotides set forth in SEQ ID NO:4, an active fragment thereof or the complement thereof. In another example, the isolated nucleic acid molecule has the sequence of nucleotides set forth in SEQ ID NO:4, an active fragment thereof or the complement thereof.

Also provided herein is a isopimaradiene synthase (Iso) polypeptide encoded by any of the nucleic acid molecules provided above encoding a isopimaradiene synthase (Iso) polypeptide or an active fragment thereof.

Also provided herein is a vector containing any of the nucleic acid molecules provided herein above encoding isopimaradiene synthase (Iso) polypeptide or an active fragment thereof. Also provided is a host cell containing such a vector. The host cell can be a prokaryotic or eukaryotic host cell. For example, the host cell is a bacterial, fungal, plant, insect, amphibian or animal cell, such as an *E. coli* cell or a yeast cell. In examples herein, the host cell is one that produces GGPP. For example, the host cell is transformed with one or more sequences that result in production of GGPP.

Also provided herein is a method of producing isopimaradiene by i) contacting (E,E,E)-geranylgeranyl diphosphate (GGPP) with a isopimaradiene synthase (Iso) polypeptide encoded by the nucleic acid molecule of any of claims 57-65 under conditions effective to produce isopimaradiene; and ii) optionally, isolating the isopimaradiene produced in step i). In examples of the methods provided herein, the step of contacting E,E,E)-geranylgeranyl diphosphate (GGPP) with a isopimaradiene synthase (Iso) polypeptide is effected *in vitro* or *in vivo*. In examples of the methods provided herein, the method further includes isolating isopimaradiene.

It is understood that this summary of subject matter provided herein does not necessarily describe all features provided herein.

#### **BRIEF DESCRIPTION OF THE FIGURES**

- 9 -

These and other features provided herein will become more apparent from the following description in which reference is made to the appended drawings wherein: **Figure 1** shows a schematic of the proposed biosynthesis of diterpene olefins and alcohols by class I/II bifunctional diTPSs in conifers. The activity of LAS- and Iso-type diTPSs in the formation of tricyclic diterpenes involves the stepwise cyclization of GGPP via (9S,10S)-CPP (*i.e.*, CPP of normal or (+) stereochemistry). **Figure 1A.** Protonation initiated cyclization of GGPP to CPP is catalyzed by the class II active site of LAS- and Iso-type diTPSs. At the class I active site of Iso enzymes, ionization-dependent secondary cyclization of CPP and deprotonation of the resulting sandaracpoimaren-8-yl cation lead to the formation of isopimaradiene. Alternatively, in the class I active site of LAS enzymes, additional proton transfer and methyl migration afford the tertiary abietenyl cation and further deprotonation and hydroxylation (*i.e.*, water capture) result in the formation of the instable 13-hydroxy-8(14)abietene product of the LAS activity. Dehydration of this tertiary diterpene alcohol yields several different diterpene olefins, including abietadiene, levopimaradiene, palustradiene, and neoabietadiene. **Figure 1B.** A different reaction sequence is proposed for the formation of the bicyclic diterpene alcohol *cis*-abienol by the class I/II bifunctional *AbCAS* enzyme. The class II activity of *AbCAS* converts GGPP to labda-13-en-8-ol diphosphate via water capture at the C-8 carbon of the labda-13-en-8-yl<sup>+</sup> cation. Subsequent ionization of the allylic diphosphate at the class I active site, without additional cyclization, yields *cis*-abienol.

**Figure 2** shows the activity of balsam fir diTPSs *AbdiTPS1* (*AbLAS*; SEQ ID NO:1), *AbdiTPS2* (*AbIso*; SEQ ID NO:3) and *AbdiTPS4* (*AbCAS*; SEQ ID NO:7). **Figure 2A.** Total ion chromatograms (TIC) of reactions products from *in vitro* assays with purified recombinant enzymes, as described in Example 3. IS, internal standard 1.6 μM 1-eicosene; **1**, palustradiene; **2**, levopimaradiene; **3**, abietadiene; **4**, neoabietadiene; **5** and **6**, epimers of 13-hydroxy-8(14)-abietadiene; **7**, isopimaradiene; **8**, *cis*-abienol. **Figure 2B** Mass spectrum of *cis*-abienol, produced by recombinant *AbdiTPS4* with GGPP as substrate, and comparison to the reference mass spectrum of *cis*-abienol.

**Figure 3** shows LC-MS analysis of diterpenol products of *AbdiTPS1* (*AbLAS*; SEQ ID NO:1) and *AbdiTPS4* (*AbCAS*; SEQ ID NO:7). Reaction products are shown as

- 10 -

extracted ion chromatograms (EIC) of the base peak  $m/z$  273, including **3**, abietadiene; **5/6**, 13-hydroxy-8(14)-abietadiene epimers; **8**, *cis*-abienol.

**Figure 4** shows phylogeny of balsam fir diTPSs of the LAS-, Iso- and CAS-type in the context of the TPS-d3 subfamily of conifer diTPSs, including *PpCPS/KS*,

- 5 *Physcomitrella patens ent*-copalyl diphosphate/*ent*-kaurene synthase (NCBI accession no. BAF61135; SEQ ID NO:42); *TbrTS*, *Taxus brevifolia* taxadiene synthase (NCBI accession no. AAC49310; SEQ ID NO:43); *TcuTS*, *T. cuspidata* taxadiene synthase (NCBI accession no. ABW82997; SEQ ID NO:44); *TxmTS*, *T. x media* taxadiene synthase (NCBI accession no. AAS18603; SEQ ID NO:45); *PaIso*, *Picea abies*  
10 isopimaradiene synthase (NCBI accession no. AAS47690; SEQ ID NO:35); *PsIso*, *Picea sitchensis* isopimaradiene synthase (NCBI accession no. ADZ45512; SEQ ID NO:36); *PaLAS*, *P. abies* (NCBI accession no. AAS47691; SEQ ID NO:37); *PsLAS*, *P. sitchensis* levopimaradiene/abietadiene synthase (NCBI accession no. ADZ45517; SEQ ID NO:38); *PtLAS*, *Pinus taeda* levopimaradiene synthase (NCBI accession no.  
15 AY779541; SEQ ID NO:39); *AgAS*, *Abies grandis* abietadiene synthase (NCBI accession no. AAK83563; SEQ ID NO:40); and *GbLS*, *Ginkgo biloba* levopimaradiene synthase (NCBI accession no. AAL09965; SEQ ID NO:41).

Asterisks indicate nodes supported by > 90% bootstrap values.

- Figure 5** shows characterization of *AbCAS* protein variants. **Figure 5A.** Total ion  
20 chromatograms (TIC) of reactions products. IS, internal standard 1.6  $\mu$ M 1-eicosene; PC, plasticizer contamination; **8**, *cis*-abienol; **9**, geranylgeraniol; **10**, (9*S*,10*S*)-copalol; **11**, epi-manoyl oxide; **12**, manoyl oxide; **13**, labd-13-en-8,15-diol. **Figure 5B.** Characteristic mass spectra of (9*S*,10*S*)-copalol (upper panel, compound **10**) and labd-13-en-8,15-diol (lower panel, compound **13**), produced by *PaLAS*:D611A and  
25 *AbCAS*:D621A, respectively.

- Figure 6** depicts the unique active site residues of *AbCAS* (SEQ ID NO:7) relative to *AbLAS* (SEQ ID NO:1) and *AbIso* (SEQ ID NO:3) and implication for the formation of *cis*-abienol. **Figure 6A.** Superimposed homology models of *AbCAS*, *AbLAS* and *AbIso* resembling the common  $\alpha$ -helical folding pattern, comprised of the N-terminal  
30  $\beta$  domain and  $\gamma$  domain and C-terminal  $\alpha$  domain. Labda-13-en-8-ol diphosphate is shown in the active sites, with the diphosphate group bound to the  $Mg^{2+}$ -cluster, which in return is coordinated by the DDxxD (SEQ ID NO:13) and NDxxTxxxE

- 11 -

(SEQ ID NO:14) motifs. **Figure 6B.** Superimposition of putative catalytic residues in the class II active site of *AbLAS*, *AbIso* and *AbCAS*. Asp348 is located at the posterior of the class II active site, opposite of the DIDD (SEQ ID NO:16) motif, with its side chain protruding towards C-8 of labda-13-en-8-ol diphosphate. **Figure 6C.**

5 Leu617 and Phe696 are located in the class I active site cavity upstream of the DDLYD motif (SEQ ID NO:17), creating an expansion of the hydrophobic pocket relative to *AbLAS* and *AbIso*. **Figure 6D.** Gly723 and Val724 contribute to a change in the hinge region between helix G1 and G2.

**Figure 7** shows amino acid sequence alignment of bifunctional class I/II conifer  
10 diTPSs, including AbdiTPS1 (SEQ ID NO:57), AbdiTPS2 (SEQ ID NO:3), AbdiTPS3 (SEQ ID NO:5) and AbdiTPS4 (SEQ ID NO:58). Grey shading indicates strictly conserved residues. The catalytically relevant aspartate-rich motifs (*i.e.*, DxDD, DDxxD, NDxxTxxxE, (SEQ ID NOS:12-14)) are highlighted. Plastidial transit peptides are underlined. Amino acids identified in a 7Å radius around the  
15 docked reaction intermediates in the N-terminal class II (light asterisks) and C-terminal class I (dark asterisks) active site are highlighted. Additional bifunctional class I/II conifer diTPSs include *PaIso*, *Picea abies* isopimaradiene synthase (NCBI accession no. AAS47690; SEQ ID NO:35); *PsIso*, *Picea sitchensis* isopimaradiene synthase (NCBI accession no. ADZ45512; SEQ ID NO:36); *PaLAS*, *P. abies* (NCBI  
20 accession no. AAS47691; SEQ ID NO:37); *PsLAS*, *P. sitchensis* levopimaradiene/abietadiene synthase (NCBI accession no. ADZ45517; SEQ ID NO:38); *PtLAS*, *Pinus taeda* levopimaradiene synthase (NCBI accession no. AY779541; SEQ ID NO:39); *AgAS*, *Abies grandis* abietadiene synthase (NCBI accession no. AAK83563; SEQ ID NO:40); and *GbLS*, *Ginkgo biloba*  
25 levopimaradiene synthase (NCBI accession no. AAL09965; SEQ ID NO:41).

**FIG. 8.** **Figure 8A** shows the synthesis of (-)-ambroxide from cis-abienol. **Figure 8B** shows the synthesis of (-)-ambroxide from geranylgeranyl diphosphate. **Figure 8C** shows the synthesis of (-)-ambroxide from cis-abienol through the intermediate sclareolide.

## 30 DETAILED DESCRIPTION

### Outline

#### A. Definitions

- 12 -

**B. Bifunctional Class I/II Diterpene Synthase Polypeptides and Diterpenoid Products**

1. **Bifunctional Class I/II Diterpene Synthases in Balsam Fir**
2. **Biosynthesis of Oleoresin Diterpenoids by diTPSs**
  - a. **Biosynthesis of Tricyclic Diterpene**
  - b. **Biosynthesis of Bicyclic Diterpene**
3. **Methods of Producing Diterpenoids**

**C. cis-Abienol Synthase and Production of Cis-Abienol**

1. **Nucleic Acid and Encoded *AbCAS* Polypeptides**
  2. **Methods of producing *cis*-abienol and ambroxide**
- Production of (-)-ambroxide**

**D. LAS-type bifunctional class I/II Synthase**

**E. Isopimaradiene Synthase**

**F. Methods of Producing or Generating Diterpene Synthases, Vectors & Host Cells**

1. **Isolation of nucleic acid encoding diterpene synthases**
2. **Vectors and Cells**
3. **Expression Systems**
  - a. **Prokaryotic cells**
  - b. **Yeast cells**
  - c. **Plants and plant cells**
  - d. **Insects and insect cells**
  - e. **Mammalian expression**
4. **Purification**
5. **Fusion Proteins**

**G. Examples**

**A. DEFINITIONS**

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as is commonly understood by one of skill in the art to which the invention(s) belong. All patents, patent applications, published applications and publications, Genbank sequences, databases, websites and other published materials referred to throughout the entire disclosure herein, unless noted otherwise, are incorporated by reference in their entirety. In the event that there are a plurality of definitions for terms herein, those in this section prevail. Where reference is made to a URL or other such identifier or address, it understood that such identifiers can change and particular information on the internet can come and go, but equivalent information can be found by searching the internet. Reference thereto evidences the availability and public dissemination of such information.

As used herein, a diterpenoid is an unsaturated hydrocarbon based on the isoprene unit ( $C_5H_8$ ), and having a general formula  $C_{5x}H_{8x}$ . A diterpene contain 20

- 13 -

carbon atoms, and hence is made up of four isoprene units. A diterpenoid also is a type of diterpene. A diterpenoid can derive from geranylgeranyl pyrophosphate (GGPP). Diterpenoids include diterpene olefins and diterpene alcohols.

As used herein, “diterpene synthase” or “diTPS” as used herein, refers to  
5 bifunctional diterpene synthase that is capable of synthesizing diterpene olefins and alcohols by sequential cycloisomerisation of the substrate (E,E,E)-geranylgeranyl diphosphate (GGPP).

As used herein, a bifunctional class I/II diTPS refers to a synthase that contains a class II active site that has a DxDD motif (SEQ ID NO:12) and a class I  
10 active site that has a DDxxD motif (SEQ ID NO:13).

As used herein, an active fragment of a synthase polypeptide refers to a contiguous sequence of amino acids of a synthase polypeptide that exhibits synthase activity (*e.g.* *cis*-abienol synthase activity, levopimaradiene/abietadiene synthase activity or isopimaradiene synthase activity), but that does not include the full-  
15 sequence of the synthase polypeptide. For purposes herein, the active fragment typically includes the class I site and class II site, and thereby contains a class II active site that has a DxDD motif (SEQ ID NO:12) and a class I active site that has a DDxxD motif (SEQ ID NO:13). The active fragment generally contains at least 300, 400, 500, 600, 700, 800 or more amino acid residues.

20 As used herein, “*cis*-abienol synthase (CAS) activity” or “AbCAS activity” refers to a synthase polypeptide or an active fragment of a synthase polypeptide that catalyzes the formation of *cis*-abienol from geranylgeranyl diphosphate (GGPP).

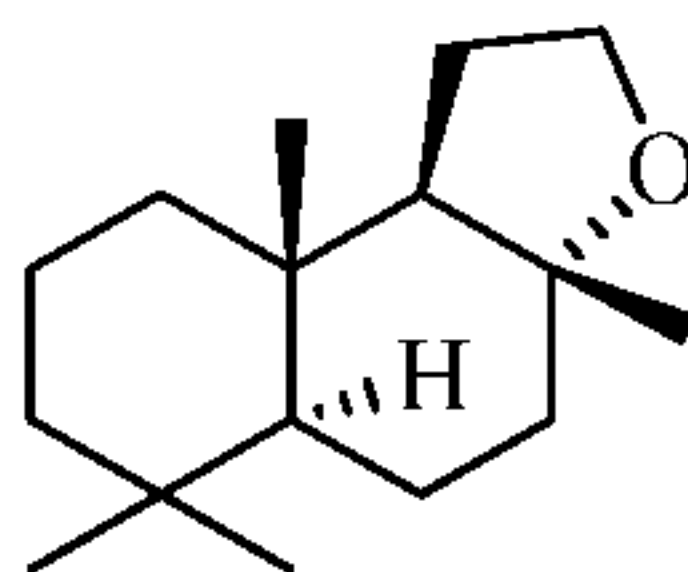
As used herein, “levopimaradiene/abietadiene synthase (LAS) activity” or “AbLAS activity” refers to a synthase polypeptide or an active fragment of a synthase  
25 polypeptide that catalyzes the formation of 13-hydroxy-8(14)-abietene from geranylgeranyl diphosphate (GGPP) and the generation of epimers thereof as dehydration products, including abietadine, levopimaradiene, neoabietadiene and plaustadiene.

As used herein, “isopimaradiene synthase activity” or “AbIso” refers to a  
30 synthase polypeptide or an active fragment of a synthase polypeptide that catalyzes the formation of isopimaradiene from geranylgeranyl diphosphate (GGPP).

- 14 -

As used herein, a pseudomature polypeptide with reference to a synthase refers to a polypeptide that lacks one or more amino acid residues from the N-terminus of the preprotein, and typically at least 10, 20, 30, 40, 50, 60, 70, 80, 90 or more N-terminal amino acid residues. Typically, a pseudomature polypeptide lacks the plastidial transit peptide. For example, with reference to *AbCAS*, the plastidial transit polypeptide corresponds to amino acid residues 1-50 of SEQ ID NO:7. Hence, a pseudomature *AbCAS* polypeptide lacks at least 50, 55, 60, 65, 70, 75, 80, 90 or more N-terminal amino acid residues of the preprotein set forth in SEQ ID NO:7.

As used herein, ambroxide is the compound having the following structure or a mixture of isomers thereof:



As used herein, corresponding residues refers to residues that occur at aligned loci. Related or variant polypeptides are aligned by any method known to those of skill in the art. Such methods typically maximize matches, and include methods such as manual alignments and those produced by the numerous alignment programs available (for example, BLASTP) and others known to those of skill in the art. By aligning the sequences of polypeptides, one skilled in the art can identify corresponding residues, using conserved and identical amino acid residues as guides. Corresponding positions also can be based on structural alignments, for example by using computer simulated alignments of protein structure.

As used herein, nucleic acids or nucleic acid molecules include DNA, RNA and analogs thereof, including peptide nucleic acids (PNA) and mixtures thereof. Nucleic acids can be single or double-stranded.

As used herein, the term polynucleotide means a single- or double-stranded polymer of deoxyribonucleotides or ribonucleotide bases read from the 5' to the 3' end. Polynucleotides include RNA and DNA, and can be isolated from natural sources, synthesized *in vitro*, or prepared from a combination of natural and synthetic molecules. The length of a polynucleotide molecule is given herein in terms of nucleotides (abbreviated "nt") or base pairs (abbreviated "bp"). The term nucleotides is used for single- and double-stranded molecules where the context permits. When

- 15 -

the term is applied to double-stranded molecules it is used to denote overall length and will be understood to be equivalent to the term base pairs. It will be recognized by those skilled in the art that the two strands of a double-stranded polynucleotide can differ slightly in length and that the ends thereof can be staggered; thus all nucleotides  
 5 within a double-stranded polynucleotide molecule cannot be paired. Such unpaired ends will, in general, not exceed 20 nucleotides in length.

As used herein, a peptide refers to a polypeptide that is greater than or equal to 2 amino acids in length, and less than or equal to 40 amino acids in length.

As used herein, the amino acids which occur in the various sequences of  
 10 amino acids provided herein are identified according to their known, three-letter or one-letter abbreviations (Table 1). The nucleotides which occur in the various nucleic acid fragments are designated with the standard single-letter designations used routinely in the art.

As used herein, an “amino acid” is an organic compound containing an amino  
 15 group and a carboxylic acid group. A polypeptide contains two or more amino acids. For purposes herein, amino acids include the twenty naturally-occurring amino acids, non-natural amino acids and amino acid analogs (*i.e.*, amino acids wherein the  $\alpha$ -carbon has a side chain).

As used herein, “amino acid residue” refers to an amino acid formed upon  
 20 chemical digestion (hydrolysis) of a polypeptide at its peptide linkages. The amino acid residues described herein are presumed to be in the “L” isomeric form. Residues in the “D” isomeric form, which are so designated, can be substituted for any L-amino acid residue as long as the desired functional property is retained by the polypeptide. NH<sub>2</sub> refers to the free amino group present at the amino terminus of a polypeptide.  
 25 COOH refers to the free carboxy group present at the carboxyl terminus of a polypeptide. In keeping with standard polypeptide nomenclature described in J. Biol. Chem. 243:3557-3559 (1968), and adopted 37 C.F.R. §§ 1.821-1.822, abbreviations for amino acid residues are shown in Table 1:

**Table 1 – Table of Correspondence**

| SYMBOL   |          |            |
|----------|----------|------------|
| 1-Letter | 3-Letter | AMINO ACID |
| Y        | Tyr      | Tyrosine   |
| G        | Gly      | Glycine    |

- 16 -

| SYMBOL   |          |                  |
|----------|----------|------------------|
| 1-Letter | 3-Letter | AMINO ACID       |
| F        | Phe      | Phenylalanine    |
| M        | Met      | Methionine       |
| A        | Ala      | Alanine          |
| S        | Ser      | Serine           |
| I        | Ile      | Isoleucine       |
| L        | Leu      | Leucine          |
| T        | Thr      | Threonine        |
| V        | Val      | Valine           |
| P        | Pro      | Proline          |
| K        | Lys      | Lysine           |
| H        | His      | Histidine        |
| Q        | Gln      | Glutamine        |
| E        | Glu      | Glutamic acid    |
| Z        | Glx      | Glu and/or Gln   |
| W        | Trp      | Tryptophan       |
| R        | Arg      | Arginine         |
| D        | Asp      | Aspartic acid    |
| N        | Asn      | Asparagine       |
| B        | Asx      | Asn and/or Asp   |
| C        | Cys      | Cysteine         |
| X        | Xaa      | Unknown or other |

All amino acid residue sequences represented herein by formulae have a left to right orientation in the conventional direction of amino-terminus to carboxyl-terminus. In addition, the phrase “amino acid residue” is defined to include the amino acids listed in the Table of Correspondence (Table 1) and modified and unusual amino acids, such as those referred to in 37 C.F.R. §§ 1.821-1.822, and incorporated herein by reference. Furthermore, it should be noted that a dash at the beginning or end of an amino acid residue sequence indicates a peptide bond to a further sequence of one or more amino acid residues, to an amino-terminal group such as NH<sub>2</sub> or to a carboxyl-terminal group such as COOH.

As used herein, “naturally occurring amino acids” refer to the 20 L-amino acids that occur in polypeptides.

As used herein, “non-natural amino acid” refers to an organic compound containing an amino group and a carboxylic acid group that is not one of the naturally-occurring amino acids listed in Table 1. Non-naturally occurring amino acids thus include, for example, amino acids or analogs of amino acids other than the 20 naturally-occurring amino acids and include, but are not limited to, the D-

- 17 -

isostereomers of amino acids. Exemplary non-natural amino acids are known to those of skill in the art.

As used herein, modification is in reference to modification of the primary sequence of amino acids of a polypeptide or a sequence of nucleotides in a nucleic acid molecule and includes deletions, insertions, and replacements and rearrangements of amino acids and nucleotides. Modifications can be made by making conservative amino acid replacements and also non-conservative amino acid substitutions as well as by insertions and other such changes in primary sequence. Modifications also can include post-translational modifications or other changes to the molecule that can occur due to conjugation or linkage, directly or indirectly, to another moiety, but when such modifications are contemplated they are referred to as post-translational modifications or conjugates or other such term as appropriate. Methods of modifying a polypeptide are routine to those of skill in the art, and can be performed by standard methods, such as site directed mutations, amplification methods, and gene shuffling methods.

As used herein, amino acid replacements or substitutions contemplated include, but are not limited to, conservative substitutions, including, but not limited to, those set forth in Table 2. Suitable conservative substitutions of amino acids are known to those of skill in the art and can be made generally without altering the conformation or activity of the polypeptide. Those of skill in this art recognize that, in general, single amino acid substitutions in non-essential regions of a polypeptide do not substantially alter biological activity (see, *e.g.*, Watson *et al. Molecular Biology of the Gene*, 4th Edition, 1987, The Benjamin/Cummings Pub. co., p.224). Conservative amino acid substitutions are made, for example, in accordance with those set forth in Table 2 as follows:

**TABLE 2**

| <b>Original residue</b> | <b>Conservative substitution</b> |
|-------------------------|----------------------------------|
| Ala (A)                 | Gly; Ser                         |
| Arg (R)                 | Lys                              |
| Asn (N)                 | Gln; His                         |
| Cys (C)                 | Ser                              |
| Gln (Q)                 | Asn                              |
| Glu (E)                 | Asp                              |
| Gly (G)                 | Ala; Pro                         |

- 18 -

| Original residue | Conservative substitution |
|------------------|---------------------------|
| His (H)          | Asn; Gln                  |
| Ile (I)          | Leu; Val                  |
| Leu (L)          | Ile; Val                  |
| Lys (K)          | Arg; Gln; Glu             |
| Met (M)          | Leu; Tyr; Ile             |
| Phe (F)          | Met; Leu; Tyr             |
| Ser (S)          | Thr                       |
| Thr (T)          | Ser                       |
| Trp (W)          | Tyr                       |
| Tyr (Y)          | Trp; Phe                  |
| Val (V)          | Ile; Leu; Met             |

Other conservative substitutions also are permissible and can be determined empirically or in accord with known conservative substitutions.

As used herein, a DNA construct is a single or double stranded, linear or  
5 circular DNA molecule that contains segments of DNA combined and juxtaposed in a manner not found in nature. DNA constructs exist as a result of human manipulation, and include clones and other copies of manipulated molecules.

As used herein, a DNA segment is a portion of a larger DNA molecule having specified attributes. For example, a DNA segment encoding a specified polypeptide is  
10 a portion of a longer DNA molecule, such as a plasmid or plasmid fragment, which, when read from the 5' to 3' direction, encodes the sequence of amino acids of the specified polypeptide.

As used herein, "primary sequence" refers to the sequence of amino acid residues in a polypeptide.

As used herein, "similarity" between two proteins or nucleic acids refers to the  
15 relatedness between the sequence of amino acids of the proteins or the nucleotide sequences of the nucleic acids. Similarity can be based on the degree of identity and/or homology of sequences of residues and the residues contained therein. Methods for assessing the degree of similarity between proteins or nucleic acids are  
20 known to those of skill in the art. For example, in one method of assessing sequence similarity, two amino acid or nucleotide sequences are aligned in a manner that yields a maximal level of identity between the sequences. "Identity" refers to the extent to which the amino acid or nucleotide sequences are invariant. Alignment of amino acid sequences, and to some extent nucleotide sequences, also can take into account

- 19 -

conservative differences and/or frequent substitutions in amino acids (or nucleotides). Conservative differences are those that preserve the physico-chemical properties of the residues involved. Alignments can be global (alignment of the compared sequences over the entire length of the sequences and including all residues) or local  
5 (the alignment of a portion of the sequences that includes only the most similar region or regions).

As used herein, “sequence identity” refers to the number of identical or similar amino acids or nucleotide bases in a comparison between a test and a reference polypeptide or polynucleotide. Sequence identity can be determined by sequence  
10 alignment of nucleic acid or protein sequences to identify regions of similarity or identity. For purposes herein, sequence identity is generally determined by alignment to identify identical residues. The alignment can be local or global. Matches, mismatches and gaps can be identified between compared sequences. Gaps are null amino acids or nucleotides inserted between the residues of aligned sequences so that  
15 identical or similar characters are aligned. Generally, there can be internal and terminal gaps. Sequence identity can be determined by taking into account gaps as the number of identical residues/ length of the shortest sequence x 100. When using gap penalties, sequence identity can be determined with no penalty for end gaps (*e.g.* terminal gaps are not penalized). Alternatively, sequence identity can be determined  
20 without taking into account gaps as the number of identical positions/length of the total aligned sequence x 100.

Amino acid sequence similarity or identity can be computed by using the BLASTP and TBLASTN programs which employ the BLAST (basic local alignment search tool) 2.0 algorithm. Techniques for computing amino acid sequence similarity  
25 or identity are well known to those skilled in the art, and the use of the BLAST algorithm is described in ALTSCHUL *et al.* 1990, *J Mol. Biol.* 215: 403- 410 and ALTSCHUL *et al.* (1997), *Nucleic Acids Res.* 25: 3389-3402.

As used herein, the term “identity” represents a comparison between a test and a reference polypeptide or polynucleotide. In one non-limiting example, “at least 90  
30 % identical to” refers to percent identities from 90 to 100 % relative to the reference polypeptides. Identity at a level of 90 % or more is indicative of the fact that, assuming for exemplification purposes a test and reference polypeptide length of 100

- 20 -

amino acids are compared, no more than 10 % (*i.e.*, 10 out of 100) of amino acids in the test polypeptide differs from that of the reference polypeptides. Similar comparisons can be made between a test and reference polynucleotides. Such differences can be represented as point mutations randomly distributed over the entire  
5 length of an amino acid sequence or they can be clustered in one or more locations of varying length up to the maximum allowable, *e.g.*, 10/100 amino acid difference (approximately 90 % identity). Differences also can be due to deletions or truncations of amino acid or nucleotide residues. Differences are defined as nucleic acid or amino acid substitutions, insertions or deletions. Depending on the length of the compared  
10 sequences, at the level of homologies or identities above about 85-90 %, the result reasonably independent of the program and gap parameters set; such high levels of identity can be assessed readily, often without relying on software.

As used herein, a substantially similar sequence is an amino acid sequence that differs from a reference sequence only by one or more conservative substitutions.  
15 Such a sequence can, for example, be functionally homologous to another substantially similar sequence. It will be appreciated by a person of skill in the art the aspects of the individual amino acids in a peptide provided herein that can be substituted.

As used herein, an aligned sequence refers to the use of homology (similarity  
20 and/or identity) to align corresponding positions in a sequence of nucleotides or amino acids. Typically, two or more sequences that are related by about or 50 % or more identity are aligned. An aligned set of sequences refers to 2 or more sequences that are aligned at corresponding positions and can include aligning sequences derived from RNAs, such as ESTs and other cDNAs, aligned with genomic DNA sequence.

25 As used herein, substantially pure means sufficiently homogeneous to appear free of readily detectable impurities as determined by standard methods of analysis, such as thin layer chromatography (TLC), gel electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the  
30 physical and chemical properties, such as enzymatic and biological activities, of the substance. Methods for purification of the compounds to produce substantially chemically pure compounds are known to those of skill in the art. A substantially

- 21 -

chemically pure compound can, however, be a mixture of stereoisomers or isomers. In such instances, further purification might increase the specific activity of the compound.

As used herein, isolated or purified polypeptide or protein or biologically-  
5 active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue from which the protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. Preparations can be determined to be substantially free if they appear free of readily detectable impurities as determined by standard methods of analysis, such as thin  
10 layer chromatography (TLC), gel electrophoresis and high performance liquid chromatography (HPLC), used by those of skill in the art to assess such purity, or sufficiently pure such that further purification would not detectably alter the physical and chemical properties, such as proteolytic and biological activities, of the substance. Methods for purification of the compounds to produce substantially chemically pure  
15 compounds are known to those of skill in the art. A substantially chemically pure compound, however, can be a mixture of stereoisomers. In such instances, further purification might increase the specific activity of the compound.

As used herein, substantially free of cellular material includes preparations of diTPSs or diterpene products in which the synthase or product is separated from  
20 cellular components of the cells from which it is isolated or produced. In one embodiment, the term substantially free of cellular material includes preparations of having less than about or less than 30%, 20%, 10%, 5% or less (by dry weight) of non-diTPS or diterpene product, including cell culture medium. When the synthase is recombinantly produced, it also is substantially free of culture medium, *i.e.*, culture  
25 medium represents less than about or at 20%, 10% or 5% of the volume of the synthase protein preparation.

As used herein, the term substantially free of chemical precursors or other chemicals includes preparations of synthase proteins or diterpene products that is separated from chemical precursors or other chemicals that are involved in the  
30 synthesis thereof. The term includes preparations of synthase proteins or diterpene products having less than about or less than 30% (by dry weight), 20%, 10%, 5% or less of chemical precursors or non- synthase chemicals or components.

- 22 -

As used herein, synthetic, with reference to, for example, a synthetic nucleic acid molecule or a synthetic gene or a synthetic peptide refers to a nucleic acid molecule or polypeptide molecule that is produced by recombinant methods and/or by chemical synthesis methods.

5           As used herein, production by recombinant methods by using recombinant DNA methods refers to the use of the well known methods of molecular biology for expressing proteins encoded by cloned DNA.

          As used herein, vector (or plasmid) refers to discrete DNA elements that are used to introduce heterologous nucleic acid into cells for either expression or  
10          replication thereof. The vectors typically remain episomal, but can be designed to effect integration of a gene or portion thereof into a chromosome of the genome. Also contemplated are vectors that are artificial chromosomes, such as bacterial artificial chromosomes, yeast artificial chromosomes and mammalian artificial chromosomes. Selection and use of such vehicles are well known to those of skill in the art.

15           As used herein, expression refers to the process by which nucleic acid is transcribed into mRNA and translated into peptides, polypeptides, or proteins. If the nucleic acid is derived from genomic DNA, expression can, if an appropriate eukaryotic host cell or organism is selected, include processing, such as splicing of the mRNA.

20           As used herein, an expression vector includes vectors capable of expressing DNA that is operatively linked with regulatory sequences, such as promoter regions, that are capable of effecting expression of such DNA fragments. Such additional segments can include promoter and terminator sequences, and optionally can include one or more origins of replication, one or more selectable markers, an enhancer, a  
25          polyadenylation signal, and the like. Expression vectors are generally derived from plasmid or viral DNA, or can contain elements of both. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector that, upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well  
30          known to those of skill in the art and include those that are replicable in eukaryotic cells and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome.

- 23 -

As used herein, vector also includes “virus vectors” or “viral vectors.” Viral vectors are engineered viruses that are operatively linked to exogenous genes to transfer (as vehicles or shuttles) the exogenous genes into cells. Viral vectors include, but are not limited to, adenoviral vectors, retroviral vectors and vaccinia virus vectors.

5 As used herein, operably or operatively linked when referring to DNA segments means that the segments are arranged so that they function in concert for their intended purposes, *e.g.*, transcription initiates downstream of the promoter and upstream of any transcribed sequences. The promoter is usually the domain to which the transcriptional machinery binds to initiate transcription and proceeds through the  
10 coding segment to the terminator.

As used herein, the term assessing or determining includes quantitative and qualitative determination in the sense of obtaining an absolute value for the activity of a product, and also of obtaining an index, ratio, percentage, visual or other value indicative of the level of the activity. Assessment can be direct or indirect.

15 As used herein, recitation that a polypeptide “consists essentially” of a recited sequence of amino acids means that only the recited portion, or a fragment thereof, of the full-length polypeptide is present. The polypeptide can optionally, and generally will, include additional amino acids from another source or can be inserted into another polypeptide

20 As used herein, the singular forms “a”, “an” and “the” include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to polypeptide, comprising “an amino acid replacement” includes polypeptides with one or a plurality of amino acid replacements.

As used herein, ranges and amounts can be expressed as “about” a particular  
25 value or range. About also includes the exact amount. Hence “about 5 %” means “about 5 %” and also “5 %.”

As used herein, “optional” or “optionally” means that the subsequently described event or circumstance does or does not occur, and that the description includes instances where said event or circumstance occurs and instances where it  
30 does not. For example, an optional step of isolating a diterpenoid (*e.g. cis*-abienol) means that the diterpenoid (*e.g. cis*-abienol) is isolated or is not isolated.

- 24 -

As used herein, the abbreviations for any protective groups, amino acids and other compounds, are, unless indicated otherwise, in accord with their common usage, recognized abbreviations, or the IUPAC-IUB Commission on Biochemical Nomenclature (see, (1972) *Biochem.* 11:1726).

5 For clarity of disclosure, and not by way of limitation, the detailed description is divided into the subsections that follow.

**B. Bifunctional Class I/II Diterpene Synthase Polypeptides and Diterpenoid Products**

The present disclosure provides, in part, diterpene synthase (diTPS) nucleic acid molecules and polypeptides that can, for example, be used in the production of diterpenoids. The disclosure also relates, in part, to vectors containing such sequences, transformed cells, cell lines, and transgenic organisms. The present disclosure also provides methods for producing a diterpene. The present disclosure further relates to a method for producing diterpenoids in a host. The present disclosure also provides compositions, uses, and kits comprising diTPS.

**1. Bifunctional Class I/II Diterpene Synthases in Balsam fir**

Varied mechanisms of diTPS-catalyzed cycloisomerization of (*E,E,E*)-geranylgeranyl diphosphate (GGPP) contribute substantially to the chemical diversity of diterpene metabolites found in the oleoresin of conifers, and in nature, in general. Introduction of hydroxyl functions further increases the diversity of naturally occurring diTPS products. Hydroxylation of conifer diterpenes can result from activity of P450s acting on products of diTPSs (Ro *et al.* (2005) *Proc. Natl. Acad. Sci. U.S.A.*, 102:8060-8065; Hamberger *et al.* (2011) *Plant Physiology*, 157:1677-1695), or can result from capture of water by carbocation intermediates during the diTPS reaction (Keeling *et al.* (2011) *J. Biol. Chem.*, 286:21145-21153). For example, in general, biosynthesis of oleoresin diterpenoids in conifers involves the sequential cycloisomerization of GGPP, catalyzed by bifunctional class I/II diterpene synthases (diTPSs) of the TPS-d subfamily. Oxygen functionality of conifer oleoresin diterpenes can be introduced by diTPSs (Keeling *et al.* (2011) *J. Biol. Chem.*, 286:21145-21153) and/or by separate activity of cytochrome P450 dependent monooxygenases of the CYP720B subfamily (Ro *et al.* (2005) *Proc. Natl. Acad. Sci.*

- 25 -

U.S.A., 102:8060-8065; Hamberger *et al.* (2006) *Biochem. Soc. Trans.*, 34:1209-1214; Hamberger *et al.* (2011) *Plant Physiol*, 157:1677-1695).

Relatively little is known about diTPSs catalyzing cyclohydration reactions. Reported examples are the biosynthesis of copal-8-ol by a monofunctional class II diTPS from *Cistus creticus* (Falara *et al.* (2010) *Plant Physiol.*, 154: 301-310), the formation of *ent*-16 $\alpha$ -hydroxy-kaurene as a product of the bifunctional class I/II *ent*-copalyl diphosphate/*ent*-kaurene synthase (CPS/KS) from the non-vascular plants *Physcomitrella patens* and *Jungermannia subulata* (Hayashi *et al.* (2006) *FEBS Lett.*, 580: 6175-6181; Kawaide *et al.* (2011) *FEBS J.*, 278:123-133), and the biosynthesis of labda-7,13-dien-15-ol catalyzed by a CPS/KS-like diTPS from the lycophyte *Sellaginella moellendorffii* (Mafu *et al.* (2011) *Chembiochem.*, 12:1984-1987). In addition, a recent study on the product specificity of Norway spruce PaLAS (Keeling *et al.* (2011) *J. Biol. Chem.*, 286:21145-21153), demonstrated a tricyclic, tertiary diterpene alcohol as the initial, but highly unstable, product of this diTPS. In the case of PaLAS, the allylic diterpenol product dehydrates to a set of diterpene olefins.

Previously characterized conifer diTPSs produce either isopimaradiene and minor amounts of sandaracopimaradiene (Iso-type diTPSs) or epimeric forms of 13-hydroxy-8(14)-abietene (LAS-type diTPSs) (Figure 1A; *see also* Stofer Bogel *et al.* (1996) *J. Biol. Chem.*, 271:23262-23268; Peters *et al.* (2000) *Biochemistry*, 39:15592-15602; Peters *et al.* (2001) *J. Am. Chem. Soc.*, 123:8974-8978; Martin *et al.* (2004) *Plant Physiol.*, 135:1908-1927; Ro *et al.* (2006) *Phytochemistry*, 67:1572-1578; Keeling *et al.* (2008) *Proc. Natl. Acad. Sci. U.S.A.*, 105:1085-1090; Keeling *et al.* (2011) *BMC Plant Biol.*, 11:43; and Keeling *et al.* (2011) *J. Biol. Chem.*, 286:21145-21153). *In vitro*, and perhaps *in vivo*, 13-hydroxy-8(14)-abietene is readily dehydrated, resulting in a mixture of abietadiene, levopimaradiene, neoabietadiene, and palustradiene (Keeling *et al.* (2011) *J. Biol. Chem.*, 286:21145-21153).

As described herein from development of a 454-transcriptome resource for balsam fir, which was used in conjunction with terpenoid metabolite profiling, diTPSs were identified that resembled bifunctional class I/II gymnosperm diTPSs of the TPS-d group containing the characteristic DxDD, DDxxD and NSE/DTE motifs (SEQ ID NOS:12-14, respectively) (Martin *et al.* (2004) *Plant Physiol* 135:1908-1927;

- 26 -

Keeling *et al.* (2011) *BMC Plant Biol* 11:43; Chen *et al.* (2011) *The Plant Journal* 66:212-229; Peters *et al.* (2003) *Biochemistry* 42:2700-2707; Peters & Croteau (2002) *Biochemistry* 41:1836-1842; Zhou & Peters (2009) *Phytochemistry* 70:366-369). The identified diTPS synthases are denoted AbdiTPS1 (set forth in SEQ ID NO:2 and  
 5 encoding the sequence of amino acids set forth in SEQ ID NO:1), AbdiTPS2 (set forth in SEQ ID NO: 4 and encoding the sequence of amino acids set forth in SEQ ID NO:3), AbdiTPS3 (set forth in SEQ ID NO:6 and encoding the sequence of amino acids set forth in SEQ ID NO:5) and AbdiTPS4 (set forth in SEQ ID NO:8 and encoding the sequence of amino acids set forth in SEQ ID NO:7. Identification and  
 10 functional characterization of pseudomature forms lacking the plastidial transit peptide revealed that AbdiTPS4 is a class I/II gymnosperm *cis*-abienol synthase (CAS). The same sequence resource also revealed the two additional bifunctional class I/II diTPSs AbdiTPS2 and AbdiTPS1 represent the balsam fir Iso and LAS enzymes, respectively, which form tricyclic diterpenes.

15 The present disclosure relates to one, or more than one, diterpene synthase (diTPS) nucleic acid molecule and one, or more than one, diTPS polypeptide. The one or more than one, diTPS polypeptides can be a class I/II diTPS. More specifically the one or more than one diTPS polypeptides can be a bifunctional class I/II diTPS. The diTPS can therefore contain a class II active site that has a DxDD motif and/or a  
 20 class I active site that has a DDxxD motif. The present disclosure provides a nucleic acid containing a nucleotide sequence encoding diterpene synthase (diTPS), for example, LAS diTPS, Iso-type diTPS and CAS diTPS. The nucleotide sequence encoding diTPS can be operatively linked to a regulatory region active in a host.

Also provided herein are variants of any of the nucleic acid sequences  
 25 provided herein exhibiting substantially the same properties as the sequences provided herein. By this it is meant that nucleic acid sequences need not be identical to the sequence disclosed herein. Variations can be attributable to single or multiple base substitutions, deletions, or insertions or local mutations involving one or more nucleotides not substantially detracting from the properties of the nucleic acid  
 30 sequence as encoding an enzyme having the activity of the diTPS as provided herein.

One, or more than one, nucleic acid encoding a diTPS are provided. The nucleic acid encoding a diTPS, such as is used in any of the described embodiments

- 27 -

herein, can contain a nucleotide sequence that is at least 50% identical to any of SEQ ID NO: 2, 4, 6 or 8, a portion thereof that encodes an active fragment that exhibits diTPS activity or to the complement thereof. For example, the nucleic acid contains a nucleotide sequence that is at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, such as generally at least 95% or at least 98% identical to any of SEQ ID NO: 2, 4, 6 or 8, a portion thereof that encodes an active fragment that exhibits diTPS activity or the complement thereof. The present disclosure provides nucleic acid sequences encoding for a polypeptide having a sequence selected from SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, an active fragment thereof or sequences substantially identical thereto. For example, the provided nucleic acid sequence encodes a pseudomature form of any of SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, or an active fragment thereof. The one, or more than one, nucleic acid can contain the sequence set forth in SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 6, SEQ ID NO: 8, a portion thereof that encodes an active fragment that exhibits diTPS activity, combinations thereof, or sequences substantially similar thereto. The sequence of the nucleic acid can be changed, for example, to account for codon preference in a particular host cell. In particular examples, the nucleic acid encoding a diTPS contains a nucleotide sequence set forth in any of SEQ ID NO: 2, 4, 6 or 8, a portion thereof that encodes an active fragment or the complement thereof. In other examples, the nucleic acid encoding a diTPS is set forth in any of SEQ ID NO: 2, 4, 6 or 8, a portion thereof that encodes an active fragment or the complement thereof.

Also provided are one, or more than one diTPS polypeptides. The polypeptide having a diTPS activity, such as intended for use in aspects of the methods provided herein, is a polypeptide having an amino acid sequence that is at least 50% identical to any of SEQ ID NO: 1, 3, 5, or 7 or an active fragment thereof that exhibits a diTPS activity. Such polypeptides include pseudomature forms lacking the transit peptide. For examples, among polypeptides provided herein are any that have an amino acid sequence that is at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, such as generally at least 95% or at least 98% identical to any of SEQ ID NO: 1, 3, 5, or 7 or an active fragment thereof that exhibits diTPS activity. The one, or more than one diTPS polypeptides can contain the

- 28 -

sequence set forth in SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 or an active fragment thereof that exhibits diTPS activity, or sequences having at least about 80-100% sequence similarity thereto, including any percent similarity within these ranges, such as or at least or greater than 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% sequence similarity thereto. The present disclosure provides nucleic acid sequences encoding for a polypeptide having a sequence selected from SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7, an active fragment thereof that exhibits diTPS activity or sequences substantially identical thereto. In examples herein, the polypeptide contains the sequence of amino acids set forth in any of SEQ ID NO: 1, 3, 5, or 7 or an active fragment thereof that exhibits diTPS activity. In other examples, the amino acid sequence for a polypeptide provided herein is set forth in SEQ ID NO: 1, 3, 5, or 7 or an active fragment thereof that exhibits diTPS activity. Also provided herein are pseudomature forms of any of SEQ ID NOS:1, 3, 5 or 7 lacking the transit peptide.

15           The present study highlights the powers of combined metabolite profiling, tissue-specific deep transcriptome sequencing, and functional (*i.e.*, biochemical) genomics for the successful discovery and characterization of new enzymes of natural product biosynthesis (*see also* Bleeker *et al.* (2011) *Plant. Mol. Biol.*, 77:323-336). Of fundamental importance for the success of a genomics- or transcriptomics-based approach to natural products enzyme discovery is the traditional or new knowledge of the occurrence of specialized metabolites associated with particular plant species or particular tissues. Balsam fir was ideally suited for the discovery of CAS, since this species accumulates large amounts of *cis*-abienol in the oleoresin of bark/phloem tissue (Example 1). Xylem/wood or needles of balsam fir do not contain *cis*-abienol in relevant amounts, and therefore were not included in the transcriptome sequencing for CAS discovery.

          Beyond the selection of species and tissues via metabolite profiling prior to 454 sequencing, another element in the successful discovery strategy of *AbCAS* was the application of a directed BlastX search of the *de novo* transcriptome assembly. Specifically, the BlastX search of the *de novo* transcriptome assembly was directed against a comprehensive sequence database of characterized TPSs. This strategy allowed for the curator-optimized annotation of TPSs genes, which substantially

- 29 -

streamlined the efforts of functional gene characterization towards four candidate *AbdiTPS*s. Of these four candidates, the corresponding enzymes of *AbdiTPS1* (*AbLAS*), *AbdiTPS2* (*AbIso*), and *AbdiTPS4* (*AbCAS*) were successfully characterized. Together, the set of three types of bifunctional diTPSs, *AbIso*, *AbLAS* and *AbCAS*, account for the majority of diterpene structures found in the specialized diterpene metabolism of balsam fir bark tissue.

## 2. Biosynthesis of Oleoresin Diterpenoids by diTPSs

Bifunctional class I/II diTPSs of conifer specialized metabolism are members of the TPS-d group (Martin *et al.* (2004) *Plant Physiol.*, 135:1908-1927; Keeling *et al.* (2011) *BMC Plant Biol.*, 11:43; Chen *et al.* (2011) *The Plant Journal*, 66:212-229). Within this group, new genes and enzyme functions appear to have evolved from a common ancestor by repeated gene duplication. Neo- or sub-functionalization of members of this gene family involved mutation of specific active site residues (Keeling *et al.* (2008) *Proc. Natl. Acad. Sci., U.S.A.*, 105:1085-1090; Wilderman *et al.* (2007) *J. Am. Chem. Soc.*, 129:15736-15737). As shown herein, using mutational analysis of *AbCAS* and comparative structural analysis of *AbCAS* with *AbIso* and *AbLAS*, distinct catalytic mechanisms were identified. Also, unique residues in the *AbCAS* class I and class II active sites that are associated with product specificity also were revealed.

The initial bicyclization of GGPP, resulting in a labda-13-en-8-yl<sup>+</sup> intermediate of (9S,10S)-stereochemistry, is likely a common feature of the class II reactions of CAS-, LAS-, and Iso-type diTPSs (Figure 1). Previous work showed that naturally occurring *cis*-abienol is optically pure and represents the (9S,10S)-stereoisomer (Carman *et al.* (1993) *Aust. J. Chem.*, 46: 1105-1114). Based on the relatively high protein sequence identity of more than 90% with *Abies grandis* abietadiene synthase (*AgAS*), a similar catalytic function was suspected for *AbdiTPS1* and *AbdiTPS2*, while the lower sequence identity of *AbdiTPS4* (75%) was indicative of a distinct function. Characterization by sequence identity was confirmed by functional characterization. In a distinct pathway of the class II reaction, it is shown herein that *AbCAS* then promotes water capture at the C-8 carbon of the carbocation to form labda-13-en-8-ol diphosphate, while the LAS- and Iso-type diTPSs form (9S,10S)-CPP through deprotonation at the C-8 position (Figure 1). In contrast to the

- 30 -

well characterized biosynthesis of tricyclic diterpenes by conifer diTPSs of the Iso- and LAS-type, until now a conifer gene or enzyme for the biosynthesis of a bicyclic diterpene *cis*-abienol has not yet been reported. Thus, *AbdiTPS4* (*AbCAS*) represents a previously unknown diTPS function.

#### 5                                    **a.        Biosynthesis of Tricyclic Diterpene**

As shown in Figure 1A, the reaction mechanism of the Iso- and LAS-type diTPSs involves the initial protonation-initiated bicyclization of GGPP at the class II active site resulting in copalyl diphosphate (CPP) of (9*S*,10*S*) configuration (*i.e.*, normal or (+)-configuration). (9*S*,10*S*)-CPP then translocates to the class I active site  
10 where it undergoes a secondary ionization-dependent cyclization and enzyme-specific rearrangement of intermediate carbocations. In Iso-type diTPSs, direct deprotonation of the sandaracopimarenyl cation results in the formation of isopimaradiene, whereas LAS-type diTPSs catalyze additional rearrangement and water capture at C-13 resulting in 13-hydroxy-8(14)-abietane as the initial product (Keeling *et al.* (2011) *J. Biol. Chem.*, 286: 21145-21153). This reaction mechanism has been exemplified in  
15 the examples herein for the Iso- and LAS-type diTPS provided herein from balsam fir.

The functional characterization of *AbdiTPS1* (*AbLAS*) with 13-hydroxy-8(14)-abietene as the initial reaction product, and multiple olefin compounds occurring upon dehydration of the alcohol, substantiates a report of the formation of a  
20 tricyclic tertiary C-13 alcohol as the primary product of Norway spruce *PaLAS* (Keeling *et al.* (2011) *J. Biol. Chem.*, 286:21145-21153). Likewise, *AbdiTPS2* (*AbIso*) is the first Iso-type diTPS to be characterized outside of the spruce genus (*Picea*), with phylogenetic patterns indicating that functional divergence of LAS- and Iso-type diTPS occurred independently in the spruce and fir (*Abies*) lineages (Figure 4).

#### 25                                    **b.        Biosynthesis of Bicyclic Diterpene**

A special feature of *AbCAS* from other conifer diTPSs is the formation of a bicyclic, tertiary diterpenol, *cis*-abienol, at the class II active site of a bifunctional class I/II diTPS. As shown herein, the reaction sequence of *cis*-abienol formation catalyzed by the bifunctional conifer class I/II diTPS (*AbCAS*) proceeds via water  
30 capture of a carbocation intermediate at carbon C-8 and subsequent ionization of the allylic diphosphate group without further cycloisomerization (Figure 1B). For the fragrance industry, bicyclic hydroxylated diterpenes, such as *cis*-abienol and sclareol,

- 31 -

are of particular value as plant-derived precursors for the sustainable production of Ambrox<sup>®</sup>, which replaces the controversial use of animal-derived ambergris in perfume formulations.

Structural modeling and molecular docking of labda-13-en-8-ol diphosphate in the class I and class II active sites of *AbCAS* (see Figure 6) revealed only one unique amino acid in the class II active site in proximity of the docked intermediate that seems likely to control this particular hydroxylation reaction, namely Asp348 with reference to the sequence set forth in SEQ ID NO:58 (corresponding to Asp349 of SEQ ID NO:7). Located at the posterior of the active site opposite of the DIDD motif and Trp358, which have previously been reported to contribute to the *AgAS*-catalyzed class II reaction (Peters *et al.* (2002) *Biochemistry*, 41:1836-1842), the negatively charged side chain of Asp348 appears to stabilize the positive charge at C-8 for water quenching to occur in the formation of labda-13-en-8-ol diphosphate (see Figure 6B). This local negative charge is not present in any LAS- or Iso-type enzymes, which contain a conserved histidine in this position (Figure 7).

Comparative homology modeling and site directed mutagenesis revealed unique features of *cis*-abienol biosynthesis that demonstrate a functional divergence in the evolution of CAS relative to the paralogous Iso- and LAS-type conifer diTPSs associated with substitution of key residues in the active sites. Mutational analysis and complementation assays indicate that the class I active site of *AbCAS* is selective for labda-13-en-8-ol diphosphate as a substrate, but does not convert (9S,10S)-CPP, which is the class I active site substrate of LAS-, and Iso-type diTPSs (Figure 5). Unlike LAS- and Iso-enzymes, the class I active site of *AbCAS* catalyzes the ionization of the diphosphate group without cyclization of a C-ring (Figure 1). Several unique residues were found in a radius of 7Å around the hydroxy group of labda-13-en-8-ol diphosphate docked within the class I active site of *AbCAS* that indicate a role in the *AbCAS* reaction (Figure 6 C and D). Among these, Leu617, Phe696 and Gly723 with reference to the sequence set forth in SEQ ID NO:58 are of particular interest as they contribute to contour of the active site cavity (corresponding to residues Leu618, Phe697 and Gly724 of SEQ ID NO:7). Leu617 and Phe696 are located in the class I active site cavity upstream of the DDxxD motif on helix D, creating an expansion of the hydrophobic pocket relative to *AbLAS* and *AbIso*.

- 32 -

Gly723 and Val724 account for a change in the hinge region between helix G1 and G2. These residues are positioned to contribute to the release of a bicyclic product rather than facilitating a secondary cyclization. Interestingly, residues corresponding to Leu617, Phe696 and Gly723 have previously been shown to be critical for the catalytic plasticity of conifer diTPSs (Keeling *et al.* (2008) *Proc. Natl. Acad. Sci. U.S.A.*, 105:1085-1090; Wilderman *et al.* (2007) *J. Am. Chem. Soc.*, 129:15736-15737; Peters *et al.* (2002) *Proc. Natl. Acad. Sci., U.S.A.*, 99:580-584; Leonard *et al.* (2010) *Proc. Natl. Acad. Sci. U.S.A.*, 107:13654-13659) and thus appear to represent key positions for the functional evolution of these enzymes.

While tricyclic diterpene resin acids are ubiquitously abundant in the pine family (*Pinaceae*), the oleoresin of balsam fir (*Abies balsamea* (L.) Mill.) contains *cis*-abienol, a bicyclic, tertiary diterpene alcohol, as the major diterpenoid (Gray *et al.* (1964) *J. Chem. Soc.*, 5822-5825). The generation of bicyclic, tertiary diterpene alcohols has been demonstrated in non-conifer clades. For example, recently, a monofunctional angiosperm diTPS, catalyzing the formation of a bicyclic oxygen-containing diterpenoid, copal-8-ol diphosphate synthase, from *Cistus creticus* (*Cistaceae*) has been reported (Falara *et al.* (2010) *Plant Physiol.*, 154:301-310). This enzyme represents a class II diTPS that catalyzes the protonation-initiated cyclization of GGPP to form the hydroxylated CPP compound. In addition, Mafu *et al.* ((2011), *Chembiochem*, 12:1984-1987) described a bifunctional class I/II diTPS, *SmCPSKSL1*, from the lycophyte *Sellaginella moellendorffii* for the formation of labda-7,13-dien-15-ol, where the primary hydroxyl group is introduced in the reaction of the class I active site.

Despite extensive efforts of TPS gene discovery in species of spruce (Martin *et al.* (2004) *Plant Physiol.*, 135:1908-1927; Keeling *et al.* (2011) *BMC Plant Biol.*, 11:43; Ralph *et al.* (2008) *BMC Genomics*, 9:484) and in grand fir (Bohlmann *et al.* (1999) *Arch. Biochem. Biophys.*, 368:232-243), there is no known gene in any plant species that is closely related (> 70% protein identity) to *AbCAS*. It is therefore possible that a CAS-type bifunctional class I/II diTPS is unique to balsam fir, which is the first conifer species for which three types of functionally distinct diTPSs are now known. A phylogenetic position of *AbCAS* between *Ginkgo biloba* *GbLS* and other known conifer diTPSs of specialized metabolism in spruce, firs, and pine (Figure 4)

- 33 -

indicates an evolutionary origin of *AbCAS* by gene duplication prior to speciation within the pine family with its possible loss in some or all of the other species of this comparison.

### 3. Methods of Producing Diterpenoids

5 Plant TPSs are useful enzymes for the metabolic engineering of bioproducts and biofuels in yeast and *E. coli* (Bohlmann *et al.* (2008) *Plant J.*, 54:656-669; Peralta-Yahya *et al.* (2011) *Nat. Commun.*, 2:483). US Patent Application 2011/0041218 discloses a method for the production of sclareol, a compound useful in the fields of perfumery and flavoring. US Patent Application 2008/0281135  
10 discloses a method for producing terpenes of interest in plants having glandular trichomes. The plants contain a sequence encoding a heterologous terpene synthase under the control of a promoter permitting it to be specifically expressed in the trichomes. Moreover, the pathway for producing endogenous diterpenes is blocked in the trichomes of the plants, to increase the flow in the heterologous pathway. WO  
15 2008/007031 discloses a protein having a syn-copalyl-8-ol diphosphate synthase activity, the nucleotide sequence encoding said protein, as well as a vector and a transgenic non-human organism containing the nucleic acid.

Provided herein are methods of producing diterpenoids *in vitro* or *in vivo* using the bifunctional class I/II diTPSs provided herein. Depending on the diTPS  
20 used, the diterpenoid that can be produced by the present methods are for example *cis*-abienol, abietadiene, levopimaradiene, palustradiene, neoabietadiene and/or isopimaradiene.

In one example, the method for producing diterpenoids is carried out *in vitro*. In this case, (E,E,E)-geranylgeranyl diphosphate (GGPP) is contacted with at least  
25 one polypeptide having a diterpene synthase (diTPS) activity under conditions effective to produce diterpenoids. In performing the methods, GGPP can be added to a suspension or solution containing a diterpene synthase polypeptide, which is then incubated at optimal temperature, for example between 15 and 40° C, such as between 25 and 35° C, or at 30° C. The produced diterpenoid can optionally be isolating by  
30 methods known in the art. For example, after incubation, the one or more than one diterpene produced can be isolated from the incubated solution by standard isolation procedures, such as solvent extraction and distillation, optionally after removal of

- 34 -

polypeptides from the solution. For example, extraction can be effected with pentane, diethyl ether, methyl tertiary butyl ether or other organic solvent. Production and quantification of the amount of the diterpene product (*e.g.* any one or more of *cis*-abienol, abietadiene, levopimaradiene, palustradiene, neoabietadiene and/or isopimaradiene) can then be determined using any method known in the art, such as column chromatography, for example liquid chromatography (*e.g.* LC-MS or HPLC) or gas chromatography (*e.g.* GC-MS), using an internal standard. For detection of diphosphate intermediates, reaction products can be dephosphorylated prior to extraction by incubation with alkaline phosphatase.

10 In another example, the method for producing diterpenoids is carried out *in vivo*. In this case, the method involves introducing into a host capable of producing GGPP, a nucleotide sequence encoding a diterpene synthase (diTPS) operatively linked with a regulatory region active in the host, and growing that host under conditions that permit the expression of the nucleic acid, thereby producing the diterpenoids. Any host cell can be used for expressing the diTPS, such as any host cell described in Section F. For example, the host cell can be a eukaryotic or prokaryotic host cell that produces GGPP or is modified to produce GGPP. Exemplary of host cells are bacterial host cells (*e.g.* *E.coli*) or fungal host cells (*e.g.* yeast). In such an example, it is possible to carry out the method *in vivo* without previously isolating the polypeptide. The reaction occurs directly within the organism or cell transformed to express said nucleic acid. The diterpene product (*e.g.* any one or more of *cis*-abienol, abietadiene, levopimaradiene, palustradiene, neoabietadiene and/or isopimaradiene) then can be extracted from the cell culture medium with an organic solvent and subsequently isolated or purified by any known method, such as column chromatography, such as liquid chromatography (*e.g.* LC-MS or HPLC) or gas chromatography (*e.g.* GC-MS), and the amount and purity of the recovered product are assessed.

For example *cis*-abienol, abietadiene, levopimaradiene, palustradiene, neoabietadiene and/or isopimaradiene can be obtained by these *in vitro* and *in vivo* methods. Other compounds that can be formed are labda-13-en-8-ol diphosphate, (+)-copalyl diphosphate, sandaracopimaren-8-yl, abientenyl, 13-hydroxy-8(14)labietene

- 35 -

and (-)-isopimaradiene, (see Figure 1). The exact product profile is dependent on the conditions in which the method is carried out and the particular diTPS employed.

The quantity of diterpene produced, such as for example *cis*-abienol, abietadiene, levopimaradiene, palustradiene, neoabietadiene and/or isopimaradiene, can be determined by any known standard chromatographic technique useful for separating and analyzing organic compounds. For example, production can be assayed by any known chromatographic technique useful for the detection and quantification of hydrocarbons, including, but not limited to, gas chromatography mass spectrometry (GC-MS), gas chromatography using a flame ionization detector (GC-FID), capillary GC-MS, liquid chromatography mass spectrometry (LC-MS), high performance liquid chromatography (HPLC) and column chromatography. Typically, these techniques are carried out in the presence of known internal standards which are used to quantify the amount of the terpene produced. For example, diterpenes can be identified by comparison of retention times and mass spectra to those of authentic standards for the particular diterpene in gas chromatography with mass spectrometry detection. In other examples, quantification can be achieved by gas chromatography with flame ionization detection based upon calibration curves with known amounts of authentic standards and normalization to the peak area of an internal standard. These chromatographic techniques allow for the identification of any terpene or diterpene present in the organic layer.

Exemplary provided diTPSs, and methods of use thereof, are described in the following Sections.

### C. *cis*-Abienol Synthase and Production of *Cis*-Abienol

Provided herein is a *cis*-Abienol Synthase (*Ab*TPS4 or *Ab*CAS) polypeptide or active fragment thereof that catalyzes the formation of *cis*-abienol from geranylgeranyl diphosphate (GGPP). The *Ab*CAS is a bifunctional class I/II dTPS that exhibits class I and class II enzymatic activities to produce a diterpenoid. The nucleic acid or encoded *Ab*CAS polypeptide can be used in methods of producing *cis*-abienol, which can be used to generate Ambrox<sup>®</sup> ((-)-ambroxide).

In addition to balsam fir, a gymnosperm tree, only a few angiosperm plant sources such as tobacco (*Nicotiana tabacum*; family *Solanaceae*) trichomes (Guo *et al.* (1994) *Archives of Biochemistry and Biophysics*, 308:103-108; Guo *et al.* (1995)

- 36 -

*Planta*, 197:627-632) or the tuberous roots of Bolivian sunroot (*Polymnia sonchifolia*; family *Asteraceae*) (Miyazawa *et al.* (2008) *J. Essent. Oil Res.*, 20:12-14) are known to produce *cis*-abienol in amounts that are relevant for industrial-scale extraction.

Metabolic engineering of microbial hosts or plants to produce oxygenated  
 5 diterpenoids such as *cis*-abienol or sclareol, can provide a sustainable production platform for these compounds for the fragrance industry and other applications.

The importance of *cis*-abienol as a plant-derived precursor for the fragrance and bioproducts industry has spawned an interest in the discovery of the relevant biosynthetic genes and enzymes and their future application in metabolically  
 10 engineered microbial or plant production systems. Genes for *cis*-abienol and sclareol biosynthesis have been cloned from angiosperms as described in the patent literature (see e.g. International PCT Application Nos. WO2008/007031; WO2009/044336; WO2009/095366; WO2009/101126) . Unlike the bifunctional conifer class I/II diTPSs provided herein, the known angiosperm diTPSs are exclusively  
 15 monofunctional enzymes (Chen *et al.* (2011) *The Plant Journal*, 66:212-229). Thus, using the angiosperm enzymes for pathway engineering would require the dual expression and optimization of separate class I and class II enzymes.

In contrast, the use of a bifunctional class I/II diTPS such as *AbCAS* provided herein requires only expression of a single gene. Optimization of protein expression is  
 20 also likely to be less complicated for the bifunctional class I/II diTPS (*i.e.*, *AbCAS*), since the two active sites required for the conversion of GGPP to *cis*-abienol will be in equal amounts as part of the same protein and will be in very close physical proximity of each other. In essence, for applications of metabolic engineering, the bifunctional class I/II *AbCAS* offers a scaffolded arrangement of the two active sites.  
 25 This scaffold evolved in nature for the high volume production of *cis*-abienol in the oleoresin of balsam fir. Since *AbCAS* is only very distantly related to angiosperm diTPSs (less than 30% protein sequence identity), this enzyme also is contemplated as the enzyme used for metabolic engineering of *cis*-abienol production in industrial crops, such as tobacco, as co-suppression effects on endogenous diTPSs are unlikely.

### 30 1. Nucleic Acid and Encoded *AbCAS* Polypeptides

Provided herein are nucleic acid molecules encoding a *cis*-Abienol Synthase (*AbCAS*) polypeptide or an active fragment thereof , including pseudomature forms

- 37 -

lacking the plastidial transit peptide, and the encoded polypeptides. The *AbCAS* polypeptide or active fragment thereof catalyzes the formation of *cis*-abienol from geranylgeranyl diphosphate (GGPP). The polypeptide having an *AbCAS* activity, such as intended for use in aspects of the methods provided herein, is a polypeptide  
 5 having an amino acid sequence that is at least 50% identical to SEQ ID NO: 7 or an active fragment thereof.

The *AbCAS* or active fragment thereof provided herein is a diTPS that is bifunctional and contains a class II active site that has a DxDD motif (SEQ ID NO:12) and a class I active site that has a DDxxD (SEQ ID NO:13) motif. The class  
 10 II active site is located in the N-terminal beta-gamma domain of the diTPS, encompassing for example residues corresponding to residues Asn87-Glu545 of AbdiTPS4 as set forth in SEQ ID NO: 7 (corresponding to residues Asn86-Glu544 as set forth in SEQ ID NO:58). The DxDD motif corresponds to amino acid residues Asp403-Asp406 as set forth in SEQ ID NO:7 (corresponding to residues Asp402-  
 15 Asp405 as set forth in SEQ ID NO:58). The class I active site is located in the C-terminal alpha domain of the diTPS, encompassing for example residues corresponding to residues Ser576-Thr867 of AbdiTPS4 as set forth in SEQ ID NO: 7 (corresponding to residues Ser575-Thr866 as set forth in SEQ ID NO:58). The DDxxD motif corresponds to amino acid residues Asp622-Asp626 as set forth in SEQ  
 20 ID NO:7 (corresponding to residues Asp621-Asp625 as set forth in SEQ ID NO:58). In a one example, a diTPS provided herein is an *AbCAS* polypeptide or active fragment thereof that contains an Asp349 amino acid in the class II active site with reference to SEQ ID NO:7 (corresponding to Asp348 as set forth in SEQ ID NO:58). Furthermore, in some examples, a diTPS provided herein is an *AbCAS* that contains a  
 25 Leu618, Phe697 and a Gly724 amino acid in the class I site with reference to amino acid residues set forth in SEQ ID NO:7 (corresponding to Leu617, Phe696 and a Gly723, respectively, as set forth in SEQ ID NO:58).

For examples, among polypeptides provided herein are any that have an amino acid sequence that is at least 55%, at least 60%, at least 65%, at least 70%, at least  
 30 75%, at least 80%, at least 85%, at least 90%, such as generally at least 95% or at least 98% identical to SEQ ID NO: 7 or an active fragment thereof. The diTPS that is an *AbCAS* polypeptide can contain the sequence set forth in SEQ ID NO: 7 or an

- 38 -

active fragment thereof, or sequences having at least about 80-100% sequence similarity thereto, including any percent similarity within these ranges, such as or at least or greater than 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% sequence similarity thereto. In particular examples, the polypeptide  
5 contains the sequence of amino acids set forth in SEQ ID NO: 7 or an active fragment thereof. In other examples, the amino acid sequence for a polypeptide provided herein is set forth in SEQ ID NO: 7 or an active fragment thereof.

Hence, also provided herein are nucleic acid molecules that encode for a polypeptide having a sequence that is at least 50% identical to SEQ ID NO:7 or that  
10 has a sequence set forth in SEQ ID NO: 7 or sequences substantially identical thereto, or an active fragment thereof. The nucleic acid encoding a diTPS that is an *AbCAS*, such as is used in any of the described methods herein, can contain a nucleotide sequence that is at least 50% identical to SEQ ID NO: 8 or a portion thereof that encodes an active fragment having *AbCAS* activity, or to the complement thereof. For  
15 example, the nucleic acid contains a nucleotide sequence that is at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, such as generally at least 95% or at least 98% identical to SEQ ID NO: 8, or a portion thereof that encodes an active fragment having *AbCAS* activity or the complement thereof. The nucleic acid can contain the sequence set forth in SEQ ID NO: 8 or a  
20 portion thereof encoding an active fragment, or sequences substantially similar thereto. The sequence of the nucleic acid can be changed, for example, to account for codon preference in a particular host cell. In particular examples, the nucleic acid encoding a diTPS that is an *AbCAS* contains a nucleotide sequence set forth in SEQ ID NO: 8, or a portion thereof that encodes an active fragment or the complement  
25 thereof. In other examples, the nucleic acid encoding a diTPS that is an *AbCAS* is set forth in SEQ ID NO: 8 or is a portion thereof that encodes an active fragment or the complement thereof.

For example, *AbCAS* polypeptides provided herein also include active forms that are pseudomature forms that lack the plastidial transit peptide and that exhibit  
30 *AbCAS* activity to catalyze the formation of *cis*-abienol from geranylgeranyl diphosphate (GGPP). The plastidial transit peptide of *AbCAS* corresponds to amino acid residues 1-50 of SEQ ID NO:7 (corresponding to residues 1-49 of SEQ ID

- 39 -

NO:58; see also Figure 7). Exemplary of pseudomature forms are forms that include all or part of the conserved N/KRx6W motif (SEQ ID NO:15), such as the N-terminal sequence K/NREFPP (SEQ ID NO:11; see also Figure 7). For example, an exemplary pseudomature form of an *AbCAS* polypeptide provided herein is a

5 polypeptide having an amino acid sequence that is at least 50% identical to SEQ ID NO: 50 or an active fragment thereof. For example, among polypeptides provided herein are any that have an amino acid sequence that is at least or greater than 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 50 or an active

10 fragment thereof.

It is understood that the pseudomature form can include fewer or greater amino acids at the N- terminus or less amino acids at the C-terminus as compared to the polypeptide set forth in SEQ ID NO:50 so long as the resulting polypeptide exhibits *AbCAS* activity to catalyze the formation of *cis*-abienol from geranylgeranyl

15 diphosphate (GGPP). For example, the *AbCAS* polypeptide can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60 or more amino acids longer or shorter than the *AbCAS* polypeptide set forth in SEQ ID NO:50. Also, one or more amino acid variations can occur in the N-terminal or C-terminal domain of the pseudomature

20 form of the *AbCAS* polypeptide depending on such factors as the cloning procedures employed (*e.g.* the addition or deletion of amino acids to facilitate cloning procedures), the presence or absence of an N-terminal methionine (*e.g.* for translation initiation), the presence or absence of a tag or other moiety and other factors well within the knowledge of one of skill in the art. Shortened or lengthened variants with

25 insertions or deletions of amino acids, particularly at either terminus that retain an activity readily can be prepared and the loci for corresponding mutations identified. For example, provided herein is a pseudomature form of *AbCAS* having an amino acid sequence that is at least 50% identical to SEQ ID NO: 55 or an active fragment thereof. For examples, among polypeptides provided herein are any that have an

30 amino acid sequence that is at least or greater than 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 55 or an active fragment thereof.

- 40 -

Hence, also provided herein are nucleic acid molecules that encode an *AbCAS* polypeptide having a sequence of amino acids that is at least 50% identical to the sequence set forth in SEQ ID NO: 50 or SEQ ID NO:55, or an active fragment thereof. The nucleic acid encoding a diTPS that is an *AbCAS*, such as is used in any of the described methods herein, can contain a nucleotide sequence that is at least 50% identical to SEQ ID NO: 54 or 56 or a portion thereof that encodes an active fragment having *AbCAS* activity, or to the complement thereof. For example, the nucleic acid contains a nucleotide sequence that is at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, such as generally at least 95% or at least 98% identical to SEQ ID NO: 54 or 56, or a portion thereof that encodes an active fragment having *AbCAS* activity or the complement thereof. The one, or more than one, nucleic acid can contain the sequence set forth in SEQ ID NO: 54 or 56 or a portion thereof encoding an active fragment, or sequences substantially similar thereto. The sequence of the nucleic acid can be changed, for example, to account for codon preference in a particular host cell. In particular examples, the nucleic acid encoding a diTPS that is an *AbCAS* contains a nucleotide sequence set forth in SEQ ID NO: 54 or 56, or a portion thereof that encodes an active fragment or the complement thereof. In other examples, the nucleic acid encoding a diTPS that is an *AbCAS* is set forth in SEQ ID NO: 54 or 56 or is a portion thereof that encodes an active fragment or the complement thereof.

In examples provided herein, amino acid residues in the class I and/or the class II active site of the the *AbCAS* polypeptides or active fragment thereof provided herein can be altered, for example by mutation as compared to any of the sequences set forth in SEQ ID NO: 7, 50 or 55. For example the diTPS polypeptide that is an *AbCAS* polypeptide or active fragment thereof can contain the sequence as set forth in SEQ ID NO: 7, an active fragment thereof (*e.g.* such as a pseudomature form as set forth in SEQ ID NO:50 or SEQ ID NO:55), or sequence substantially identical thereto, wherein the amino acid at position 622 is replaced by A, the amino acid at position 403 is replaced by A, the amino acid at position 405 is replaced by A or a combination thereof with reference to amino acid numbering set forth in SEQ ID NO:7. With reference to SEQ ID NO:58, such replacements correspond to amino acids at position 621 replaced by A, amino acid at position 402 replaced by A, amino

- 41 -

acid at position 404 replaced by A, or a combination thereof (see for example Figure 5). Corresponding amino acid residues can be identified by one of skill in the art in other sequence forms of the AbCAS polypeptide by alignment of residues with SEQ ID NO:7 or SEQ ID NO:58. In other examples provided herein, the above amino acid residues are not altered because they are identified herein to change *AbCAS* activity resulting in a loss of enzymatic activity to produce *cis*-abienol from GGPP (see Example 5).

Furthermore, the one or more than one diTPS polypeptides can contain modifications in active site residues as disclosed in Figure 6 (see also Example 6). For example, the diTPS polypeptide that is an *AbCAS* polypeptide or active fragment thereof can contain the sequence as set forth in SEQ ID NO: 7, an active fragment thereof (*e.g.* such as a pseudomature form as set forth in SEQ ID NO:50 or SEQ ID NO:55), or sequence substantially identical thereto, wherein the amino acid at position 618, 697 or 724 or a combination thereof with reference to amino acid numbering set forth in SEQ ID NO:7 is replaced with another amino acid. With reference to SEQ ID NO:58, such replacements correspond to replacement at amino acid position 617, 696 or 723 or a combination thereof with another amino acid. Corresponding amino acid residues can be identified by one of skill in the art in other sequence forms of the *AbCAS* polypeptide by alignment of residues with SEQ ID NO:7 or SEQ ID NO:58.

## 2. Methods of producing *cis*-abienol and ambroxide

Provided herein are *in vitro* and *in vivo* methods of using an *AbCAS* polypeptide or active fragment thereof provided herein for producing *cis*-abienol and related products such as (-)-ambroxide (Ambrox<sup>®</sup>). Among other applications, *cis*-abienol and other oxygen-containing diterpenoids of plant origin (*e.g.*, sclareol and manool) can be used in the fragrance industry to produce Ambrox<sup>®</sup>. Ambrox<sup>®</sup> serves as a sustainable replacement for the use of ambergris in high-end perfume formulations (Barrero et al. (1993) *Tetrahedron*, 49:10405-10412). While Ambrox<sup>®</sup> is produced from plant terpenoids, ambergris is an animal product secreted from the intestines of sperm whales, which are listed as an endangered species.

In one example, the method for producing the diterpenoid *cis*-abienol is carried out *in vitro*. In this case, (E,E,E)-geranylgeranyl diphosphate (GGPP) is

- 42 -

contacted with at least one polypeptide having *AbCAS* activity, such as any described above, under conditions effective to produce *cis*-abienol. In performing the methods, GGPP can be added to a suspension or solution containing an *AbCAS* polypeptide or active fragment thereof, such as any provided herein, which is then incubated at  
5 optimal temperature, for example between 15 and 40° C, such as between 25 and 35°C, or at 30° C. The *cis*-abienol diterpenoid can optionally be isolated by methods known in the art. For example, after incubation, the *cis*-abienol diterpene produced can be isolated from the incubated solution by standard isolation procedures, such as solvent extraction and distillation, optionally after removal of polypeptides from the  
10 solution. For example, isolation can be effected by column chromatography, including liquid chromatography methods (*e.g.* HPLC). Production and quantification of the amount of *cis*-abienol can be determined using methods as described herein, such as gas chromatography-mass spectroscopy (*e.g.* GC-MS) or liquid chromatography-mass spectroscopy (*e.g.* LC-MS), using an internal standard. In some  
15 examples, the internal standard can be a *cis*-abienol authentic standard. In other cases, production can be confirmed by comparison to a reference mass spectrum of *cis*-abienol as described by Vlad *et al.* ((1974) *Khimiya Prirodnikh Soedinenii* 1:30-35) and obtained from the National Institute of Standards and Technology MS library searches (Wiley W9N08).

20 In particular examples of product profiles obtained when the method is carried out *in vitro* with a Ni-affinity purified polypeptide that is a pseudomature form of sequence SEQ ID NO:7 (*e.g.* as set forth in SEQ ID NO:55) are provided in Figure 2. *Cis*-abienol is the direct product of the enzymatic reaction catalyzed by the diTPS that is an *AbCAS* polypeptide or active fragment thereof used in the method provided  
25 herein.

In another example, the method for producing the diterpenoid *cis*-abienol is carried out *in vivo* using any of the nucleic acid molecules encoding an *AbCAS* polypeptide or active fragment thereof provided herein. In this case, the method involves introducing into a host capable of producing GGPP, a nucleotide sequence  
30 encoding an *AbCAS* polypeptide or active fragment thereof operatively linked with a regulatory region active in the host, and growing that host under conditions that permit the expression of the nucleic acid, thereby producing the *cis*-abienol

- 43 -

diterpenoid. Any host cell can be used for expressing an *AbCAS* polypeptide or active fragment thereof, such as any host cell described in Section F. For example, the host cell can be a eukaryotic or prokaryotic host cell that produces GGPP or is modified to produce GGPP. Exemplary of host cells are bacterial host cells (*e.g.* *E.coli*) or fungal host cells (*e.g.* yeast). In such an example, it is possible to carry out the method *in vivo* without previously isolating the polypeptide. The reaction occurs directly within the organism or cell transformed to express the nucleic acid. The *cis*-abienol then can be extracted from the cell culture medium with an organic solvent and subsequently isolated or purified by any known method, such as column chromatography, including liquid chromatography methods (*e.g.* HPLC) or gas chromatography. The amount and purity of the produced *cis*-abienol also can be assessed by any known standard chromatographic technique for the detection and quantification of hydrocarbons, including, but not limited to, gas chromatography mass spectrometry (GC-MS), gas chromatography using a flame ionization detector (GC-FID), capillary GC-MS, liquid chromatography mass spectrometry (LC-MS), high performance liquid chromatography (HPLC) and column chromatography using an internal standard or based on a reference mass spectrum as described above.

#### Production of (-)-ambroxide

The *AbdiTPS4* or *AbCAS* diterpene synthase polypeptides or active fragments thereof provided herein produce *cis*-abienol, which can be used to produce (-)-ambroxide. (-)-Ambroxide is used as a base note in the perfume industry as a substitute for ambergris. Conversion of *cis*-abienol to (-)-ambroxide can be carried out through chemical means (see *e.g.* Barrero *et al.* (1993) *Tetrahedron* 49(45): 10405-10412; Barrero *et al.* (1994) *Tetrahedron* 50:6653-6662; Barrero *et al.*, (2004) *Synthetic Communications* 34(19):3631-3643; and U.S. Pat. No. 5,525,728). In one example, *cis*-abienol is converted to (-)-ambroxide by chemical methods involving ozonolysis of the side chain followed by reduction and cyclization (see Figure 8A, Scheme I). In another example, *cis*-abienol is oxidatively cleaved to form sclareolide (10), which is then converted to (-)-ambroxide by reduction and cyclization (see Figure 8B, Scheme II) (see, Barrero *et al.* (1994) *Tetrahedron* 50:6653-6662; and U.S. Pat. No. 5,525,728). For example, *cis*-abienol can be treated with OsO<sub>4</sub>-NaIO<sub>4</sub> or RuO<sub>4</sub>-NaIO<sub>4</sub> to form sclareolide.

- 44 -

(-)-Ambroxide can be purified from the reaction mixture by extraction with organic solvents, such as ethers and hydrocarbons, including for example, methyl tert-butyl ether, diethylether, n-hexane and toluene, column chromatography, or extraction with an organic solvent followed by column chromatography. (-)-Ambroxide  
5 formation can be confirmed and/or quantified by any of the chromatographic techniques described herein.

**D. LAS-type bifunctional class I/II Synthase**

Provided herein are nucleic acid molecules encoding a levopimaradiene/abietadiene synthase (*AbLAS* or *AbTPS1*) polypeptide or active  
10 fragment thereof, including pseudomature forms lacking the plastidial transit polypeptide, and the encoded polypeptides. The LAS-type synthase is a bifunctional class I/II dTPS that exhibits class I and class II enzymatic activities to produce diterpenoid products. The polypeptide or active fragment thereof, including  
15 pseudomature forms, catalyze the formation of 13-hydroxy-8(14)-abietene from geranylgeranyl diphosphate (GGPP) to effect production of epimers thereof produced as dehydration products of 13-hydroxy-8(14)-abietene, including abietadine,  
levopimaradiene, neoabietadiene and palustradiene. The nucleic acid or encoded *AbLAS* polypeptide can be used in methods to produce a diterpenoid product from  
20 among any one or more of abietadiene, levopimaradiene, neoabietadiene and palustradiene.

The *AbLAS* or active fragment thereof provided herein is a diTPS that is bifunctional and contains a class II active site that has a DxDD motif (SEQ ID NO:12) and a class I active site that has a DDxxD motif (SEQ ID NO:13). The class II active site is located in the N-terminal beta-gamma domain of the diTPS,  
25 encompassing for example residues corresponding to residues Lys84-Glu542 of *AbdiTPS1* as set forth in SEQ ID NO 1 (corresponding to residues Lys21-Glu479 as set forth in SEQ ID NO:57). The DxDD motif corresponds to amino acid residues Asp400-Asp403 as set forth in SEQ ID NO:1 (corresponding to residues Asp337-Asp340 as set forth in SEQ ID NO:57). The class I active site is located in the C-  
30 terminal alpha domain of the diTPS, encompassing for example residues corresponding to amino acid residues Ser573-Ala866 of *AbdiTPS1* as set forth in SEQ ID NO: 1 (corresponding to residues Ser510-Ala803 as set forth in SEQ ID NO:57).

- 45 -

The DDxxD motif corresponds to amino acid residues Asp619-Asp623 as set forth in SEQ ID NO:1 (corresponding to residues Asp556-Asp560 as set forth in SEQ ID NO:57).

For example among the polypeptides provided herein having *AbLAS* activity, such as intended for use in aspects of the methods provided herein, is a polypeptide having an amino acid sequence that is at least 50% identical to SEQ ID NO: 1 or an active fragment thereof. For example, among polypeptides provided herein are any that have an amino acid sequence that is at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, such as generally at least 95% or at least 98% identical to SEQ ID NO: 1 or an active fragment thereof. The diTPS that is an *AbLAS* polypeptide can contain the sequence set forth in SEQ ID NO: 1 or an active fragment thereof, or sequences having at least about 80-100% sequence similarity thereto, including any percent similarity within these ranges, such as or at least or greater than 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% sequence similarity thereto. In particular examples, the polypeptide contains the sequence of amino acids set forth in SEQ ID NO: 1 or an active fragment thereof. In other examples, the amino acid sequence for a polypeptide provided herein is set forth in SEQ ID NO: 1 or an active fragment thereof.

Hence, also provided herein are nucleic acid molecules that encode for a polypeptide having a sequence that is at least 50% identical to SEQ ID NO:1 or that has a sequence set forth in SEQ ID NO:1 or sequences substantially identical thereto, or an active fragment thereof. The nucleic acid encoding a diTPS that is an *AbLAS*, such as is used in any of the described methods herein, can contain a nucleotide sequence that is at least 50% identical to SEQ ID NO: 2, or a portion thereof that encodes an active fragment having *AbLAS* activity, or to the complement thereof. For example, the nucleic acid contains a nucleotide sequence that is at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, such as generally at least 95% or at least 98% identical to SEQ ID NO: 2, or a portion thereof that encodes an active fragment having *AbLAS* activity or the complement thereof. The nucleic acid can contain the sequence set forth in SEQ ID NO: 2 or a portion thereof encoding an active fragment, or sequences substantially similar thereto. The sequence of the nucleic acid can be changed, for example, to

- 46 -

account for codon preference in a particular host cell. In particular examples, the nucleic acid encoding a diTPS that is an *AbLAS* contains a nucleotide sequence set forth in SEQ ID NO: 2, or a portion thereof that encodes an active fragment or the complement thereof. In other examples, the nucleic acid encoding a diTPS that is an *AbLAS* is set forth in SEQ ID NO: 2 or is a portion thereof that encodes an active fragment or the complement thereof.

For example, *AbLAS* polypeptides provided herein include active forms that are pseudomature forms that lack the plastidial transit peptide and that exhibit *AbLAS* activity to catalyze the formation of 13-hydroxy-8(14)-abietene from geranylgeranyl diphosphate (GGPP) to effect production of epimers thereof produced as dehydration products of 13-hydroxy-8(14)-abietene, including abietadine, levopimaradiene, neoabietadiene and/or plaustradiene. Exemplary of pseudomature forms are forms that include all or part of the conserved N/KRx6W motif (SEQ ID NO:15; see also Figure 7). For example, an exemplary pseudomature form of an *AbLAS* polypeptide provided herein is a polypeptide having an amino acid sequence that is at least 50% identical to SEQ ID NO: 47 or an active fragment thereof. For examples, among polypeptides provided herein are any that have an amino acid sequence that is at least or greater than 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 47 or an active fragment thereof. It is understood that the pseudomature form can include fewer or greater amino acids at the N- terminus or less amino acids at the C-terminus as compared to the polypeptide set forth in SEQ ID NO:47 so long as the resulting polypeptide exhibits *AbLAS* activity to catalyze the formation of 13-hydroxy-8(14)-abietene from geranylgeranyl diphosphate (GGPP) to effect production of epimers thereof produced as dehydration products of 13-hydroxy-8(14)-abietene, including abietadine, levopimaradiene, neoabietadiene and/or plaustradiene. For example, the *AbLAS* polypeptide can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60 or more amino acids longer or shorter than the *AbLAS* polypeptide set forth in SEQ ID NO:47. Also, one or more amino acid variations can occur in the N-terminal or C-terminal domain of the pseudomature form of the *AbLAS* polypeptide depending on such factors as the cloning procedures employed (e.g. the

- 47 -

addition or deletion of amino acids to facilitate cloning procedures), the presence or absence of an N-terminal methionine (*e.g.* for translation initiation), the presence or absence of a tag or other moiety and other factors well within the knowledge of one of skill in the art. Shortened or lengthened variants with insertions or deletions of amino acids, particularly at either terminus that retain an activity readily can be prepared and the loci for corresponding mutations identified.

Hence, also provided herein are nucleic acid molecules that encode an *AbLAS* polypeptide having a sequence of amino acids that is at least 50% identical to the sequence set forth in SEQ ID NO: 47, or an active fragment thereof. The nucleic acid encoding a diTPS that is a *AbLAS*, such as is used in any of the described methods herein, can contain a nucleotide sequence that is at least 50% identical to SEQ ID NO: 51 or a portion thereof that encodes an active fragment having *AbLAS* activity, or to the complement thereof. For example, the nucleic acid contains a nucleotide sequence that is at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, such as generally at least 95% or at least 98% identical to SEQ ID NO: 51, or a portion thereof that encodes an active fragment having *AbLAS* activity or the complement thereof. The one, or more than one, nucleic acid can contain the sequence set forth in SEQ ID NO: 51 or a portion thereof encoding an active fragment, or sequences substantially similar thereto. The sequence of the nucleic acid can be changed, for example, to account for codon preference in a particular host cell. In particular examples, the nucleic acid encoding a diTPS that is an *AbLAS* contains a nucleotide sequence set forth in SEQ ID NO: 51, or a portion thereof that encodes an active fragment or the complement thereof. In other examples, the nucleic acid encoding a diTPS that is an *AbLAS* is set forth in SEQ ID NO: 51 or is a portion thereof that encodes an active fragment or the complement thereof.

In examples provided herein, amino acid residues in the class I and/or the class II active site of the the *AbLAS* polypeptides or active fragment thereof provided herein can be altered, for example by mutation as compared to the sequences set forth in SEQ ID NO: 1 or 47.

Also provided herein are *in vitro* and *in vivo* methods of using any *AbLAS* polypeptide or active fragment thereof provided herein for producing an epimer of 13-

- 48 -

hydroxy-8(14)-abietene, such as any one or more of or all of abietadiene, levopimaradiene, neoabietadiene and/or palustradiene or related products.

In one example, the method for producing an *AbLAS* product is carried out *in vitro*. In this case, (E,E,E)-geranylgeranyl diphosphate (GGPP) is contacted with at least one polypeptide having *AbLAS* activity, such as any described above, under conditions effective to produce one or more of abietadiene, levopimaradiene, neoabietadiene and/or palustradiene. In performing the methods, GGPP can be added to a suspension or solution containing an *AbLAS* polypeptide or active fragment thereof, such as any provided herein, which is then incubated at optimal temperature, for example between 15 and 40° C, such as between 25 and 35° C, or at 30° C. The produced *AbLAS* diterpenoid or diterpenoids can optionally be isolating by methods known in the art. For example, after incubation, a diterpene product produced can be isolated from the incubated solution by standard isolation procedures, such as solvent extraction and distillation, optionally after removal of polypeptides from the solution. For example, isolation can be effected by column chromatography, including liquid chromatography methods (*e.g.* HPLC) methods. In such examples, one, two, three or all diterpene products from among any one or more of one or more of abietadiene, levopimaradiene, neoabietadiene and/or palustradiene can be isolated. Methods and procedures for individually isolated such products are known to one of skill in the art, for example, by elution profiles using known chromatographic techniques (*see e.g.* Keeling *et al.* (2011) *J. Biol. Chem.*, 286:21145-53). Production and quantification of the amount of any one or more of abietadiene, levopimaradiene, neoabietadiene and/or palustradiene can be determined using methods known in the art or as described herein, such as gas chromatography-mass spectroscopy (*e.g.* GC-MS) or liquid chromatography-mass spectroscopy (*e.g.* LC-MS) using an internal standard for the particular product.

In another example, the method for producing an *AbLAS* diterpenoid is carried out *in vivo* using any of the nucleic acid molecules encoding an *AbLAS* polypeptide or active fragment thereof provided herein. In this case, the method involves introducing into a host capable of producing GGPP, a nucleotide sequence encoding an *AbLAS* polypeptide or active fragment thereof operatively linked with a regulatory region active in the host, and growing that host under conditions that permit the

- 49 -

expression of the nucleic acid, thereby producing any one or more of abietadiene, levopimaradiene, neoabietadiene and/or palustradiene. Any host cell can be used for expressing an *AbLAS* polypeptide or active fragment thereof, such as any host cell described in Section F. For example, the host cell can be a eukaryotic or prokaryotic host cell that produces GGPP or is modified to produce GGPP. Exemplary of host cells are bacterial host cells (*e.g.* *E.coli*) or fungal host cells (*e.g.* yeast). In such an example, it is possible to carry out the method *in vivo* without previously isolating the polypeptide. The reaction occurs directly within the organism or cell transformed to express said nucleic acid. The isolated diterpene product or products can be optionally isolated from the cell. For example, the products can be extracted from the cell culture medium with an organic solvent and subsequently isolated or purified by any known methods, such as by known chromatographic techniques including column chromatography methods. In such examples, one, two, three or all diterpene products from among any one or more of one or more of abietadiene, levopimaradiene, neoabietadiene and/or palustradiene can be isolated based on differences in elution profiles as described above and known in the art. Production and quantification of the amount of any one or more of abietadiene, levopimaradiene, neoabietadiene and/or palustradiene can be determined using methods known in the art or as described herein, such as by using any chromatographic technique such as gas chromatography-mass spectroscopy (*e.g.* GC-MS) or liquid chromatography-mass spectroscopy (*e.g.* LC-MS) using an internal standard for the particular product.

#### **E. Isopimaradiene Synthase**

Provided herein are nucleic acid molecules encoding an isopimaradiene synthase (*AbIso* or *AbTPS2*) polypeptide or active fragment thereof, including pseudomature forms lacking the plastidial transit polypeptide, and the encoded polypeptides. *AbIso* is a bifunctional class I/II dTPS that exhibits class I and class II enzymatic activities to produce diterpene products. The polypeptide or active fragment thereof, including pseudomature forms, catalyze the formation of isopimaradiene from geranylgeranyl diphosphate (GGPP). The nucleic acid or encoded *AbIso* polypeptide can be used in methods to produce a diterpenoid product that is an isopimaradiene.

- 50 -

The *AbIso* or active fragment thereof provided herein is a diTPS that is bifunctional and contains a class II active site that has a DxDD motif (SEQ ID NO:12) and a class I active site that has a DDxxD motif (SEQ ID NO:13). The class II active site is located in the N-terminal beta-gamma domain of the diTPS, encompassing for example residues corresponding to residues Lys69-Glu527 of AbdiTPS2 as set forth in SEQ ID NO: 3. The DxDD motif corresponds to amino acid residues Asp385-Asp388. The class I active site is located in the C-terminal alpha domain of the diTPS, encompassing for example residues corresponding to residues Ser558-Ala852 of AbdiTPS2 as set forth in SEQ ID NO:3. The DDxxD motif corresponds to amino acid residues Asp604-Asp608 as set forth in SEQ ID NO:3.

For example among the polypeptides provided herein having *AbIso* activity, such as intended for use in aspects of the methods provided herein, is a polypeptide having an amino acid sequence that is at least 50% identical to SEQ ID NO: 3 or an active fragment thereof. For example, among polypeptides provided herein are any that have an amino acid sequence that is at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, such as generally at least 95% or at least 98% identical to SEQ ID NO: 3 or an active fragment thereof. The diTPS that is an *AbIso* polypeptide can contain the sequence set forth in SEQ ID NO: 3 or an active fragment thereof, or sequences having at least about 80-100% sequence similarity thereto, including any percent similarity within these ranges, such as or at least or greater than 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99% sequence similarity thereto. In particular examples, the polypeptide contains the sequence of amino acids set forth in SEQ ID NO: 3 or an active fragment thereof. In other examples, the amino acid sequence for a polypeptide provided herein is set forth in SEQ ID NO: 3 or is an active fragment thereof.

Hence, also provided herein are nucleic acid molecules that encode for a polypeptide having a sequence that is at least 50% identical to SEQ ID NO:3 or that has a sequence set forth in SEQ ID NO:3 or sequences substantially identical thereto or an active fragment thereof. The nucleic acid encoding a diTPS that is an *AbIso*, such as is used in any of the described methods herein, can contain a nucleotide sequence that is at least 50% identical to SEQ ID NO: 4, or a portion thereof that encodes an active fragment having *AbISO* activity, or to the complement thereof. For

- 51 -

example, the nucleic acid contains a nucleotide sequence that is at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, such as generally at least 95% or at least 98% identical to SEQ ID NO: 4, or a portion thereof that encodes an active fragment having *AbISO* activity or the complement thereof. The nucleic acid can contain the sequence set forth in SEQ ID NO: 4 or a portion thereof encoding an active fragment, or sequences substantially similar thereto. The sequence of the nucleic acid can be changed, for example, to account for codon preference in a particular host cell. In particular examples, the nucleic acid encoding a diTPS that is an *AbIso* contains a nucleotide sequence set forth in SEQ ID NO: 4, or a portion thereof that encodes an active fragment or the complement thereof. In other examples, the nucleic acid encoding a diTPS that is an *AbIso* is set forth in SEQ ID NO: 4, or is the portion thereof that encodes an active fragment or the complement thereof.

For example, *AbIso* polypeptides provided herein include active forms that are pseudomature forms that lack the plastidial transit peptide and that exhibit *AbIso* activity to catalyze the formation of isopimaradiene from geranylgeranyl diphosphate (GGPP). Exemplary of pseudomature forms are forms that include all or part of the conserved N/KRx6W motif (SEQ ID NO:15; see also Figure 7). For example, an exemplary pseudomature form of an *AbIso* polypeptide provided herein is a polypeptide having an amino acid sequence that is at least 50% identical to SEQ ID NO: 48 or an active fragment thereof. For examples, among polypeptides provided herein are any that have an amino acid sequence that is at least or greater than 55%, 60%, 65%, 70%, 75%, 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to SEQ ID NO: 48 or an active fragment thereof. It is understood that the pseudomature form can include fewer or greater amino acids at the N- terminus or less amino acids at the C-terminus as compared to the polypeptide set forth in SEQ ID NO:48 so long as the resulting polypeptide exhibits *AbIso* activity to catalyze the formation of isopimaradiene from geranylgeranyl diphosphate (GGPP). For example, the *AbIso* polypeptide can be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60 or more amino acids longer or shorter than the *AbIso* polypeptide set forth in SEQ ID NO:48. Also, one or more

- 52 -

amino acid variations can occur in the N-terminal or C-terminal domain of the pseudomature form of the *AbIso* polypeptide depending on such factors as the cloning procedures employed (*e.g.* the addition or deletion of amino acids to facilitate cloning procedures), the presence or absence of an N-terminal methionine (*e.g.* for translation  
5 initiation), the presence or absence of a tag or other moiety and other factors well within the knowledge of one of skill in the art. Shortened or lengthened variants with insertions or deletions of amino acids, particularly at either terminus that retain an activity readily can be prepared and the loci for corresponding mutations identified.

Hence, also provided herein are nucleic acid molecules that encode an *AbIso*  
10 polypeptide having a sequence of amino acids that is at least 50% identical to the sequence set forth in SEQ ID NO: 48, or an active fragment thereof. The nucleic acid encoding a diTPS that is a *AbIso*, such as is used in any of the described methods herein, can contain a nucleotide sequence that is at least 50% identical to SEQ ID NO: 52 or a portion thereof that encodes an active fragment having *AbIso* activity, or to the  
15 complement thereof. For example, the nucleic acid contains a nucleotide sequence that is at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, such as generally at least 95% or at least 98% identical to SEQ ID NO: 52, or a portion thereof that encodes an active fragment having *AbIso* activity or the complement thereof. The nucleic acid can contain the  
20 sequence set forth in SEQ ID NO: 52 or a portion thereof encoding an active fragment, or sequences substantially similar thereto. The sequence of the nucleic acid can be changed, for example, to account for codon preference in a particular host cell. In particular examples, the nucleic acid encoding a diTPS that is an *AbIso* contains a nucleotide sequence set forth in SEQ ID NO: 52, or a portion thereof that encodes an  
25 active fragment or the complement thereof. In other examples, the nucleic acid encoding a diTPS that is an *AbIso* is set forth in SEQ ID NO: 52 or is a portion thereof that encodes an active fragment or the complement thereof.

In examples provided herein, amino acid residues in the class I and/or the class II active site of the *AbIso* polypeptides or active fragment thereof provided herein can  
30 be altered, for example by mutation as compared to the sequences set forth in SEQ ID NO: 3 or 48.

- 53 -

Also provided herein are *in vitro* and *in vivo* methods of using an *AbIso* polypeptide or active fragment thereof provided herein for producing isopimaradiene or a related product. In one example, the method for producing an isopimaradiene product is carried out *in vitro*. In this case, (E,E,E)-geranylgeranyl diphosphate (GGPP) is contacted with at least one polypeptide having *AbIso* activity, such as any described above, under conditions effective to produce isopimaradiene. In performing the methods, GGPP can be added to a suspension or solution containing an *AbIso* polypeptide or active fragment thereof, such as any provided herein, which is then incubated at optimal temperature, for example between 15 and 40° C, such as between 25 and 35° C, or at 30° C. The produced *AbIso* diterpenoid can optionally be isolated by methods known in the art. For example, after incubation, the diterpene produced can be isolated from the incubated solution by standard isolation procedures, such as solvent extraction and distillation, optionally after removal of polypeptides from the solution. For example, isolation can be effected by column chromatography, including liquid chromatography methods (*e.g.* HPLC). Production and quantification of the amount of isopimaradiene can be determined using methods as known in the art or as described herein, such as gas chromatography-mass spectroscopy (*e.g.* GC-MS) or liquid chromatography-mass spectroscopy (*e.g.* LC-MS) using an internal standard for isopimaradiene.

In another example, the method for producing an *AbIso* diterpenoid is carried out *in vivo* using any of the nucleic acid molecules encoding an *AbIso* polypeptide or active fragment thereof provided herein. In this case, the method involves introducing into a host capable of producing GGPP, a nucleotide sequence encoding an *AbIso* polypeptide or active fragment thereof operatively linked with a regulatory region active in the host, and growing that host under conditions that permit the expression of the nucleic acid, thereby producing isopimaradiene. Any host cell can be used for expressing an *AbIso* polypeptide or active fragment thereof, such as any host cell described in Section F. For example, the host cell can be a eukaryotic or prokaryotic host cell that produces GGPP or is modified to produce GGPP. Exemplary of host cells are bacterial host cells (*e.g.* E.coli) or fungal host cells (*e.g.* yeast). In such an example, it is possible to carry out the method *in vivo* without previously isolating the *AbIso* polypeptide. The reaction occurs directly within the organism or cell

- 54 -

transformed to express said nucleic acid. The produced isopimaradiene can be isolated from the cell. For example, the isopimaradiene then can be extracted from the cell culture medium with an organic solvent and subsequently isolated or purified by any known method, such as column chromatography, including liquid chromatography methods (*e.g.* HPLC) or gas chromatography. The amount and purity of the produced isopimaradiene also can be assessed by any known standard chromatographic technique for the detection and quantification of hydrocarbons, including, but not limited to, gas chromatography mass spectrometry (GC-MS), gas chromatography using a flame ionization detector (GC-FID), capillary GC-MS, liquid chromatography mass spectrometry (LC-MS), high performance liquid chromatography (HPLC) and column chromatography using an internal standard as described above.

#### **F. Methods of Producing or Generating Diterpene Synthases, Vectors & Host Cells**

Provided herein are polynucleotides encoding any of the diTPS provided herein or the encoded diTPSs polypeptide. As described herein, the nucleic acids and encoding polypeptides are derived from *Abies balsamea*. The polypeptide or the nucleic acid can be used in any of the method provided herein for producing a diterpenoid. Also provided herein are vectors and hosts containing the diTPS and that can be used for producing diterpenoids.

The diTPS to be used in methods provided herein also can be generated synthetically. Standard reference works setting forth the general principles of peptide synthesis technology and methods known to those of skill in the art include, for example: Chan et al., *Fmoc Solid Phase Peptide Synthesis*, Oxford University Press, Oxford, United Kingdom, 2005; *Peptide and Protein Drug Analysis*, ed. Reid, R., Marcel Dekker, Inc., 2000; *Epitope Mapping*, ed. Westwood et al., Oxford University Press, Oxford, United Kingdom, 2000; Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 3<sup>rd</sup> ed., Cold Spring Harbor Press, Cold Spring Harbor, NY 2001; and Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing Associates and John Wiley & Sons, NY, 1994.

Also provided is a diTPS kit. The kit can contain one or more diTPS nucleic acid molecules. The kit can contain one or more diTPS polypeptides. The kit can contain a synthetic diTPS gene. The kit can contain a vector containing one or more

- 55 -

diTPS nucleic acids. The kit can contain a host cell capable of expressing one or more than one diTPS polypeptide.

### 1. Isolation of nucleic acid encoding diterpene synthases

The one or more than one polynucleotide sequences encoding the diTPS as  
5 provided herein can be prepared by any method known by the person skilled in the art. For example, the polynucleotide sequence encoding a diTPS can be amplified from a cDNA template, by polymerase chain reaction with specific primers. In such an example the codons of the cDNA can be chosen to favor the expression of said protein in the desired expression system. In other examples, nucleic acids encoding  
10 diterpene synthases, including any of the diTPS provided herein, can be cloned or isolated using any available methods known in the art for cloning and isolating nucleic acid molecules. Such methods include PCR amplification of nucleic acids and screening of libraries, including nucleic acid hybridization screening. In some examples, methods for amplification of nucleic acids can be used to isolate nucleic  
15 acid molecules encoding a diTPS polypeptide, including for example, polymerase chain reaction (PCR) methods. A nucleic acid containing material can be used as a starting material from which a diTPS-encoding nucleic acid molecule can be isolated. For example, DNA and mRNA preparations from fir (*Abies* sp.), including but not limited to *Abies balsamea*, can be used to obtain diterpene synthase genes.

20 Nucleic acid libraries also can be used as a source of starting material. Primers can be designed to amplify a diterpene synthase-encoding molecule, such as a diTPS-encoding molecule. For example, primers can be designed based on known nucleic acid sequences encoding a diterpene synthase, such as a class I/class II bifunctional diterpene synthase, such as those set forth in SEQ ID NOS:35-46. Nucleic acid  
25 molecules generated by amplification can be sequenced and confirmed to encode a diTPS polypeptide.

Additional nucleotide sequences can be joined to a diTPS-encoding nucleic acid molecule, including linker sequences containing restriction endonuclease sites for the purpose of cloning the synthetic gene into a vector, for example, a protein  
30 expression vector or a vector designed for the amplification of the core protein coding DNA sequences. Furthermore, additional nucleotide sequences specifying functional DNA elements can be operatively linked to a diTPS-encoding nucleic acid molecule.

- 56 -

Still further, nucleic acid encoding other moieties or domains also can be included so that the resulting synthase is a fusion protein. For example, nucleic acids encoding other enzymes, such as a GGPP synthase, or protein purification tags, such as His or Flag tags.

5                   **2. Vectors and Cells**

          The disclosure also relates, in part, to vectors containing such sequences, transformed cells, cell lines, and transgenic organisms. For recombinant expression of one or more of the diterpene synthase polypeptides provided herein, including diTPS polypeptides, the nucleic acid containing all or a portion of the nucleotide sequence  
10       encoding the synthase can be inserted into an appropriate expression vector, *i.e.*, a vector that contains the necessary elements for the transcription and translation of the inserted protein coding sequence. Depending upon the expression system used, the necessary transcriptional and translational signals also can be supplied by the native promoter for a diTPS gene, and/or their flanking regions. For example, vectors  
15       containing a polynucleotide sequence encoding a diTPS are provided herein. The vector can be obtained and introduced in a host cell by well-known recombinant DNA and genetic engineering techniques. In some examples, a vector can contain the gene encoding a GGPP synthase, such as the gene encoding the GGPP synthase crtE from *Xanthophyllomyces dendrorhous* (SEQ ID NO:59).

20           The disclosure also provides a prokaryotic or eukaryotic host cell which is modified by a polynucleotide or a vector as provided herein. The host cell can be prokaryotic, such as bacterial, or eukaryotic, such as fungal (e.g., yeast), plant, Archea, insect, amphibian or animal cell. The host cell can contain a diTPS vector, a synthetic diTPS gene, and/or diTPS nucleic acid. The host cell can be any cell that is  
25       capable of being transformed by the vector, synthetic gene, and/or nucleic acid. The host cell can also be any cell that is capable of expressing the diTPS polypeptide. The host cell can be incubated under conditions that allow expression of the diTPS polypeptide.

          Several of these organisms do not produce GGPP naturally. To be suitable to  
30       carry out the method of the invention, these organisms may need to be transformed with one or more sequences, such as a sequence encoding a GGPP synthase, that result in production of the precursor, GGPP. They can be so transformed either

- 57 -

before the modification with the nucleic acid described according to any of the above embodiments, or simultaneously with a nucleotide sequence encoding diTPS, or a vector containing a nucleotide sequence encoding diTPS. Alternatively, in particular examples, the cells are yeast, such as *Saccharomyces cerevisiae*, that express an  
5 acyclic pyrophosphate terpene precursor, such as GGPP. The cells are used to produce a diterpene synthase, such as a diTPS polypeptide, by growing the above-described cells under conditions whereby the encoded diTPS is expressed by the cell. In some instances, the expressed synthase is purified. In other instances, the expressed synthase, such as an *AbCAS* synthase, converts GGPP to one or more terpenes (*e.g.*  
10 *cis-abienol*) in the host cell.

Any method known to those of skill in the art for the insertion of DNA fragments into a vector can be used to construct expression vectors containing a chimeric gene containing appropriate transcriptional/translational control signals and protein coding sequences. These methods can include *in vitro* recombinant DNA and  
15 synthetic techniques and *in vivo* recombinants (genetic recombination). Expression of nucleic acid sequences encoding a diTPS polypeptide, or a fragment thereof, can be regulated by a second nucleic acid sequence so that the genes or fragments thereof are expressed in a host transformed with the recombinant DNA molecule(s). For example, expression of the proteins can be controlled by any promoter/enhancer known in the  
20 art. In a specific embodiment, the promoter is not native to the genes for a diTPS protein. Promoters that can be used include but are not limited to prokaryotic, yeast, mammalian and plant promoters. The type of promoter depends upon the expression system used, described in more detail below.

In a specific embodiment, a vector is used that contains a promoter operably  
25 linked to nucleic acids encoding a diTPS polypeptide, or a fragment thereof, thereof, one or more origins of replication, and optionally, one or more selectable markers (*e.g.*, an antibiotic resistance gene). Vectors and systems for expression of diTPS polypeptides are described, including, for example, the pET28b(+) vector.

### 3. Expression Systems

30 Diterpene synthases, including diTPS polypeptides provided herein, can be produced by any methods known in the art for protein production including *in vitro* and *in vivo* methods such as, for example, the introduction of nucleic acid molecules

- 58 -

encoding the diterpene synthase (*e.g. AbIso, AbLAS and AbCAS*) into a host cell or host plant for *in vivo* production or expression from nucleic acid molecules encoding the diterpene synthase (*e.g. AbIso, AbLAS and AbCAS*) *in vitro*. Diterpene synthases such as *AbIso, AbLAS and AbCAS* polypeptides can be expressed in any organism  
 5 suitable to produce the required amounts and forms of a synthase polypeptide.

Expression hosts include prokaryotic and eukaryotic organisms such as *E. coli*, yeast, plants, insect cells, mammalian cells, including human cell lines and transgenic animals. Expression hosts can differ in their protein production levels as well as the types of post-translational modifications that are present on the expressed proteins.

10 The choice of expression host can be made based on these and other factors, such as regulatory and safety considerations, production costs and the need and methods for purification.

Isolated higher eukaryotic cells, such for example cell culture, can also be used, instead of complete organisms, as hosts to carry out the method provided herein  
 15 *in vivo*. Suitable eukaryotic cells can be any non-human cell, but are generally plant cells. Representative examples of a plant host cell include for example plants that naturally produce high amounts of terpenes. The plant can be selected from the family of Pinaceae, Funariaceae, Solanaceae, Poaceae, Brassicaceae, Fabaceae, Malvaceae, Asteraceae or Lamiaceae. For example, the plant is selected from the genera *Picea*  
 20 (spruce), *Pinus* (pine), *Abies* (fir), *Physcomitrella*, *Funariaceae*, *Nicotiana*, *Solanum*, *Sorghum*, *Arabidopsis*, *Brassica* (rape), *Medicago* (alfalfa), *Gossypium* (cotton), *Artemisia*, *Salvia* and *Mentha*. Preferably, the plant belongs to the species of *Nicotiana tabacum*, *Nicotiana benthamiana* or *Physcomitrella patens*. Additional plants and plant cells include, for example, citrus, corn, rice, algae, and lemna. In  
 25 other examples, the eukaryotic cells are yeast cells. Representative examples of a yeast host cell include those from the *Saccharomyces* genus (*e.g. Saccharomyces cerevisiae*) and *Pichia* genus (*e.g. Pichia pastoris*). In some examples, insect cells such as *Drosophila* cells and *lepidopteran* cells are used for the expression of a diTPS provided herein. Eukaryotic cells for expression also include mammalian cells lines  
 30 such as Chinese hamster ovary (CHO) cells or baby hamster kidney (BHK) cells.

Eukaryotic expression hosts also include production in transgenic animals, for example, including production in serum, milk and eggs. There are several methods

- 59 -

known in the art for the creation of transgenic host organisms or cells such as plants, fungi, prokaryotes, or cultures of higher eukaryotic cells. Appropriate cloning and expression vectors for use with bacterial, fungal, yeast, plant and mammalian cellular hosts are described, for example, in Pouwels *et al.*, Cloning Vectors: A Laboratory  
5 Manual, 1985, Elsevier, New York and Sambrook et al., Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> edition, 1989, Cold Spring Harbor Laboratory Press. Cloning and expression vectors for higher plants and/or plant cells in particular are available to the skilled person. See for example Schardl *et al.* (1987) *Gene* 61: 1-11.

Methods for transforming host organisms or cells to harbor transgenic nucleic  
10 acids are familiar to the skilled person. For the creation of transgenic plants, for example, current methods include: electroporation of plant protoplasts, liposome-mediated transformation, agrobacterium-mediated transformation, polyethylene-glycol-mediated transformation, particle bombardment, microinjection of plant cells, and transformation using viruses.

15 Many expression vectors are available and known to those of skill in the art for the expression of a diterpene synthase, such as a diTPS provided herein. Exemplary of expression vectors are pET expression vectors, such as pET28b(+). The choice of expression vector is influenced by the choice of host expression system. Such selection is well within the level of skill of the skilled artisan. In general,  
20 expression vectors can include transcriptional promoters and optionally enhancers, translational signals, and transcriptional and translational termination signals. Expression vectors that are used for stable transformation typically have a selectable marker which allows selection and maintenance of the transformed cells. In some cases, an origin of replication can be used to amplify the copy number of the vectors  
25 in the cells.

Diterpene synthases, including diTPS polypeptides, also can be utilized or expressed as protein fusions. For example, a fusion can be generated to add additional functionality to a polypeptide. Examples of fusion proteins include, but are not limited to, fusions of a signal sequence, a tag such as for localization, *e.g.* a his<sub>6</sub> tag or a myc  
30 tag, or a tag for purification, for example, a GST fusion, GFP fusion or CBP fusion, and a sequence for directing protein secretion and/or membrane association. In other examples, diterpene synthases such as diTPS polypeptides provided herein can be

- 60 -

fused to GGPP synthase (see, e.g., Brodelius *et al.* (2002) *Eur. J. Biochem.* 269:3570-3579).

Methods of production of diterpene synthase polypeptides, including *AbIso*, *AbLAS* and *AbCAS* polypeptides, can include co-expression of an acyclic  
5 pyrophosphate terpene precursor, such as GGPP, in the host cell. In some instances, the host cell naturally expresses GGPP. Such a cell can be modified to express greater quantities of GGPP (see *e.g.* U.S. Pat. Nos. 6,531,303, 6,689,593, 7,838,279 and 7,842,497). In other instances, a host cell that does not naturally produce GGPP is modified genetically to produce GGPP.

10 **a. Prokaryotic cells**

Prokaryotes, especially *E. coli*, provide a system for producing large amounts of the diTPS polypeptides provided herein. Transformation of *E. coli* is a simple and rapid technique well known to those of skill in the art. Representative examples of a bacterial host cell include, but are not limited to, *E. coli* strains such as for example *E.*  
15 *coli* BL21DE3-C41 (Miroux and Walker (1996) *J Mol Biol* 260:289-298). Exemplary expression vectors for transformation of *E. coli* cells, include, for example, the pGEM expression vectors, the pQE expression vectors, and the pET expression vectors (see, U.S. Pat. No. 4,952,496; available from Novagen, Madison, WI; see, also literature published by Novagen describing the system). Such plasmids include pET11a, which  
20 contains the T7lac promoter, T7 terminator, the inducible *E. coli* lac operator, and the lac repressor gene; pET12a-c, which contains the T7 promoter, T7 terminator, and the *E. coli* ompT secretion signal; and pET15b and pET19b (Novagen, Madison, WI), which contain a His-Tag<sup>TM</sup> leader sequence for use in purification with a His column and a thrombin cleavage site that permits cleavage following purification over the  
25 column, the T7-lac promoter region and the T7 terminator; and pET28b (Novagen, Madison, WI), which contains a His-Tag<sup>TM</sup> leader sequence for use in purification with a His column and a thrombin cleavage site that permits cleavage following purification over the column, the T7-lac promoter region and the T7 terminator; and the pJET vectors (Thermo Scientific), such as the pJET1.2 vector which contains a  
30 lethal gene which is disrupted by ligation of a DNA insert into the cloning site and a T7 promoter for *in vitro* transcription.

- 61 -

Expression vectors for *E. coli* can contain inducible promoters that are useful for inducing high levels of protein expression and for expressing proteins that exhibit some toxicity to the host cells. Exemplary prokaryotic promoters include, for example, the  $\beta$ -lactamase promoter (Jay *et al.*, (1981) *Proc. Natl. Acad. Sci. USA* 78:5543) and the *tac* promoter (DeBoer *et al.*, (1983) *Proc. Natl. Acad. Sci. USA* 80:21-25); see also "Useful Proteins from Recombinant Bacteria": in *Scientific American* 242:79-94 (1980)). Examples of inducible promoters include the *lac* promoter, the *trp* promoter, the hybrid *tac* promoter, the T7 and SP6 RNA promoters and the temperature regulated  $\lambda P_L$  promoter.

10 Diterpene synthases, including diTPS polypeptides provided herein can be expressed in the cytoplasmic environment of *E. coli*. The cytoplasm is a reducing environment and for some molecules, this can result in the formation of insoluble inclusion bodies. Reducing agents such as dithiothreitol and  $\beta$ -mercaptoethanol and denaturants (*e.g.*, such as guanidine-HCl and urea) can be used to resolubilize the

15 proteins. An alternative approach is the expression diTPS polypeptides in the periplasmic space of bacteria which provides an oxidizing environment and chaperonin-like and disulfide isomerases leading to the production of soluble protein. Typically, a leader sequence is fused to the protein to be expressed which directs the protein to the periplasm. The leader is then removed by signal peptidases inside the

20 periplasm. Examples of periplasmic-targeting leader sequences include the *pelB* leader from the pectate lyase gene and the leader derived from the alkaline phosphatase gene. In some cases, periplasmic expression allows leakage of the expressed protein into the culture medium. The secretion of proteins allows quick and simple purification from the culture supernatant. Proteins that are not secreted can be

25 obtained from the periplasm by osmotic lysis. Similar to cytoplasmic expression, in some cases proteins can become insoluble and denaturants and reducing agents can be used to facilitate solubilization and refolding. Temperature of induction and growth also can influence expression levels and solubility. Typically, temperatures between 25° C and 37° C are used. Mutations also can be used to increase solubility of

30 expressed proteins. Typically, bacteria produce aglycosylated proteins.

- 62 -

**b. Yeast cells**

Yeast systems, such as, but not limited to, those from the *Saccharomyces* genus (*e.g. Saccharomyces cerevisiae*), *Schizosaccharomyces pombe*, *Yarrowia lipolytica*, *Kluyveromyces lactis*, and *Pichia pastoris* can be used to express the  
5 diterpene synthases, such as the diTPS polypeptides, provided herein. Yeast expression systems also can be used to produce diterpenes whose reactions are catalyzed by the synthases. Yeast can be transformed with episomal replicating vectors or by stable chromosomal integration by homologous recombination. In some examples, inducible promoters are used to regulate gene expression. Exemplary  
10 promoter sequences for expression of diTPS polypeptides in yeast include, among others, promoters for metallothionine, 3-phosphoglycerate kinase (Hitzeman *et al.* (1980) *J. Biol. Chem.* 255:2073), or other glycolytic enzymes (Hess *et al.* (1968) *J. Adv. Enzyme Reg.* 7:149; and Holland *et al.* (1978) *Biochem.* 17:4900), such as enolase, glyceraldehyde phosphate dehydrogenase, hexokinase, pyruvate  
15 decarboxylase, phosphofructokinase, glucose phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other suitable vectors and promoters for use in yeast expression are further described in Hitzeman, EPA-73,657 or in Fleer *et al.* (1991) *Gene*, 107:285-195; and  
20 van den Berg *et al.* (1990) *Bio/Technology*, 8:135-139. Another alternative includes, but is not limited to, the glucose-repressible ADH2 promoter described by Russell *et al.* (*J. Biol. Chem.* 258:2674, 1982) and Beier *et al.* (*Nature* 300:724, 1982), or a modified ADH1 promoter. Shuttle vectors replicable in yeast and *E. coli* can be constructed by, for example, inserting DNA sequences from pBR322 for selection and  
25 replication in *E. coli* (Amp<sup>r</sup> gene and origin of replication) into a yeast vector.

Yeast expression vectors can include a selectable marker such as LEU2, TRP1, HIS3, and URA3 for selection and maintenance of the transformed DNA. Proteins expressed in yeast are often soluble and co-expression with chaperonins, such as Bip and protein disulfide isomerase, can improve expression levels and  
30 solubility. Additionally, proteins expressed in yeast can be directed for secretion using secretion signal peptide fusions such as the yeast mating type alpha-factor secretion signal from *Saccharomyces cerevisiae* and fusions with yeast cell surface proteins

- 63 -

such as the Aga2p mating adhesion receptor or the *Arxula adenivorans* glucoamylase. A protease cleavage site (*e.g.*, the Kex-2 protease) can be engineered to remove the fused sequences from the polypeptides as they exit the secretion pathway.

Yeast naturally express the required proteins, including GGPP synthase (BST1; which can produce GGPP) for the mevalonate-dependent isoprenoid biosynthetic pathway. Thus, expression of the diterpene synthases, including diTPS polypeptides provided herein, in yeast cells can result in the production of diterpenes, such as cis-abienol from GGPP. Exemplary yeast cells for the expression of terpene synthases, including diTPS polypeptides, include yeast modified to express increased levels of FPP and/or GGPP. For example, yeast cells can be modified to produce less squalene synthase or less active squalene synthase (*e.g.* *erg9* mutants; see *e.g.* U.S. Patent Nos. 6,531,303 and 6,689,593). This results in accumulation of FPP in the host cell at higher levels compared to wild type yeast cells, which in turn can result in increased yields of GGPP and diterpenes (*e.g.* cis-abienol, isopimaradiene, abietadiene, levopimaradiene, palustradiene and/or neoabietadiene). In another example, yeast cells can be modified to produce more GGPP synthase by introduction of a GGPP synthase gene, such as BTS1 from *S. cerevisiae*, crtE from *Erwinia uredovora*, crtE from *Xanthophyllomyces dendrorhous*, al-3 from *Neuspora crassa* or ggs from *Giverella fujiiuroi* (see U.S. Pat. No. 7,842,497). In some examples, the native GGPP gene in such yeast can be deleted. Other modifications that enable increased production of GGPP in yeast include, for example, but are not limited to, modifications that increase production of acetyl CoA, inactivate genes that encode enzymes that use FPP and GPP as substrate and overexpress of HMG-CoA reductases, as described in U.S. Pat. No. 7,842,497. Exemplary modified yeast cells include, but are not limited to, modified *Saccharomyces cerevisiae* strains CALI5-1 (*ura3, leu2, his3, trp1, Δ erg9::HIS3, HMG2cat/TRP1::rDNA, dpp1, sue*), ALX7-95 (*ura3, his3, trp1, Δ erg9::HIS3, HMG2cat/TRP1::rDNA, dpp1 sue*), ALX11-30 (*ura3, trp1, erg9<sup>def</sup>25, HMG2cat/TRP1::rDNA, dpp1, sue*), which are known and described in one or more of U.S. Pat. Nos. 6,531,303, 6,689,593, 7,838,279, 7,842,497, and published U.S. Pat. Application Serial Nos. 20040249219 and 20110189717.

- 64 -

**c. Plants and plant cells**

Transgenic plant cells and plants can be used for the expression of diterpene synthases, including diTPS polypeptides provided herein. Expression constructs are typically transferred to plants using direct DNA transfer such as microprojectile bombardment and PEG-mediated transfer into protoplasts, and with agrobacterium-mediated transformation. Expression vectors can include promoter and enhancer sequences, transcriptional termination elements, and translational control elements. Expression vectors and transformation techniques are usually divided between dicot hosts, such as *Arabidopsis* and tobacco, and monocot hosts, such as corn and rice.

Examples of plant promoters used for expression include the cauliflower mosaic virus promoter, the nopaline synthase promoter, the ribose biphosphate carboxylase promoter and the ubiquitin and UBQ3 promoters. Selectable markers such as hygromycin, phosphomannose isomerase and neomycin phosphotransferase are often used to facilitate selection and maintenance of transformed cells. Transformed plant cells can be maintained in culture as cells, aggregates (callus tissue) or regenerated into whole plants. Transgenic plant cells also can include algae engineered to produce proteins (*see, for example, Mayfield et al. (2003) Proc Natl Acad Sci USA 100:438-442*). Transformed plants include, for example, plants selected from the genera *Picea* (spruce), *Pinus* (pine), *Abies* (fir), *Physcomitrella*, *Funariaceae*, *Nicotiana*, *Solanum*, *Sorghum*, *Arabidopsis*, *Medicago* (alfalfa), *Gossypium* (cotton), *Brassica* (rape), *Artemisia*, *Salvia* and *Mentha*. In some examples, the plant belongs to the species of *Nicotiana tabacum*, *Nicotiana benthamiana* or *Physcomitrella patens*, and is transformed with vectors that overexpress a diTPS and optionally a geranylgeranyl diphosphate synthase, such as described in U.S. Pat. Pub. No. 20090123984 and U.S. Pat. No. 7,906,710.

**d. Insects and insect cells**

Insects and insect cells, particularly a baculovirus expression system, can be used for expressing diterpene synthases, including diTPS polypeptides provided herein (*see, for example, Muneta et al. (2003) J. Vet. Med. Sci. 65(2):219-223*). Insect cells and insect larvae, including expression in the haemolymph, express high levels of protein and are capable of most of the post-translational modifications used by higher eukaryotes. Baculoviruses have a restrictive host range which improves the

- 65 -

safety and reduces regulatory concerns of eukaryotic expression. Typically, expression vectors use a promoter such as the polyhedrin promoter of baculovirus for high level expression. Commonly used baculovirus systems include baculoviruses such as *Autographa californica* nuclear polyhedrosis virus (AcNPV), and the *Bombyx mori* nuclear polyhedrosis virus (BmNPV) and an insect cell line such as Sf9 derived from *Spodoptera frugiperda*, *Pseudaletia unipuncta* (A7S) and *Danaus plexippus* (DpN1). For high level expression, the nucleotide sequence of the molecule to be expressed is fused immediately downstream of the polyhedrin initiation codon of the virus. Mammalian secretion signals are accurately processed in insect cells and can be used to secrete the expressed protein into the culture medium. In addition, the cell lines *Pseudaletia unipuncta* (A7S) and *Danaus plexippus* (DpN1) produce proteins with glycosylation patterns similar to mammalian cell systems.

An alternative expression system in insect cells is the use of stably transformed cells. Cell lines such as the Schneider 2 (S2) and Kc cells (*Drosophila melanogaster*) and C7 cells (*Aedes albopictus*) can be used for expression. The *Drosophila* metallothionein promoter can be used to induce high levels of expression in the presence of heavy metal induction with cadmium or copper. Expression vectors are typically maintained by the use of selectable markers such as neomycin and hygromycin.

#### **e. Mammalian expression**

Mammalian expression systems can be used to express diterpene synthases, including diTPS polypeptides provided herein and also can be used to produce diterpenes whose reactions are catalyzed by the synthases. Expression constructs can be transferred to mammalian cells by viral infection such as adenovirus or by direct DNA transfer such as liposomes, calcium phosphate, DEAE-dextran and by physical means such as electroporation and microinjection. Expression vectors for mammalian cells typically include an mRNA cap site, a TATA box, a translational initiation sequence (Kozak consensus sequence) and polyadenylation elements. Such vectors often include transcriptional promoter-enhancers for high level expression, for example the SV40 promoter-enhancer, the human cytomegalovirus (CMV) promoter, and the long terminal repeat of Rous sarcoma virus (RSV). These promoter-enhancers are active in many cell types. Tissue and cell-type promoters and enhancer regions

- 66 -

also can be used for expression. Exemplary promoter/enhancer regions include, but are not limited to, those from genes such as elastase I, insulin, immunoglobulin, mouse mammary tumor virus, albumin, alpha-fetoprotein, alpha 1-antitrypsin, beta-globin, myelin basic protein, myosin light chain-2 and gonadotropic releasing hormone gene control. Selectable markers can be used to select for and maintain cells with the expression construct. Examples of selectable marker genes include, but are not limited to, hygromycin B phosphotransferase, adenosine deaminase, xanthine-guanine phosphoribosyl transferase, aminoglycoside phosphotransferase, dihydrofolate reductase and thymidine kinase. Fusion with cell surface signaling molecules such as TCR- $\zeta$  and Fc $\epsilon$ RI- $\gamma$  can direct expression of the proteins in an active state on the cell surface.

Many cell lines are available for mammalian expression including mouse, rat human, monkey, and chicken and hamster cells. Exemplary cell lines include, but are not limited to, BHK (*i.e.* BHK-21 cells), 293-F, CHO, CHO Express (CHOX; Excellgene), Balb/3T3, HeLa, MT2, mouse NS0 (non-secreting) and other myeloma cell lines, hybridoma and heterohybridoma cell lines, lymphocytes, fibroblasts, Sp2/0, COS, NIH3T3, HEK293, 293S, 293T, 2B8, and HKB cells. Cell lines also are available adapted to serum-free media which facilitates purification of secreted proteins from the cell culture media. One such example is the serum free EBNA-1 cell line (Pham *et al.* (2003) *Biotechnol. Bioeng.* 84:332-342).

#### 4. Purification

Also provided is a method of producing the diTPS polypeptide. The diTPS polypeptide can be purified using standard chromatographic techniques.

The polypeptide to be used when the method is carried out *in vitro* can be obtained by extraction from any organism expressing it, using standard protein or enzyme extraction technologies. If the host organism is a unicellular organism or cell releasing the polypeptide of the invention into the culture medium, the polypeptide can simply be collected from the culture medium, for example by centrifugation, optionally followed by washing steps and re-suspension in suitable buffer solutions. If the organism or cell accumulates the polypeptide within its cells, the polypeptide can be obtained by disruption or lysis of the cells and further extraction of the polypeptide from the cell lysate.

- 67 -

Methods for purification of diterpene synthases, such as diTPS polypeptides, from host cells depend on the chosen host cells and expression systems. For secreted molecules, proteins are generally purified from the culture media after removing the cells. For intracellular expression, cells can be lysed and the proteins purified from the extract. When transgenic organisms such as transgenic plants and animals are used for expression, tissues or organs can be used as starting material to make a lysed cell extract. Additionally, transgenic animal production can include the production of polypeptides in milk or eggs, which can be collected, and if necessary the proteins can be extracted and further purified using standard methods in the art.

Diterpene synthases, including diTPS polypeptides provided herein, can be purified using standard protein purification techniques known in the art including but not limited to, SDS-PAGE, size fraction and size exclusion chromatography, ammonium sulfate precipitation, chelate chromatography and ionic exchange chromatography. Expression constructs also can be engineered to add an affinity tag such as a myc epitope, GST fusion or His<sub>6</sub> and affinity purified with myc antibody, glutathione resin, and Ni-resin, respectively, to a protein. Purity can be assessed by any method known in the art including gel electrophoresis and staining and spectrophotometric techniques. The polypeptides, either in an isolated form or together with other proteins, for example in a crude protein extract obtained from cultured cells or microorganisms, can then be suspended in a buffer solution at optimal pH. If adequate, salts, DTT, BSA and other kinds of enzymatic co-factors, can be added in order to optimize enzyme activity.

### 5. Fusion Proteins

Fusion proteins containing a diterpene synthase, including diTPS polypeptides, and one or more other polypeptides also are provided. Linkage of a diterpene synthase polypeptide with another polypeptide can be effected directly or indirectly via a linker. In one example, linkage can be by chemical linkage, such as via heterobifunctional agents or thiol linkages or other such linkages. Fusion also can be effected by recombinant means. Fusion of a diterpene synthase, such as a diTPS polypeptide, e.g., *AbIso*, *AbLAS* and *AbCAS*, to another polypeptide can be to the N- or C- terminus of the diTPS polypeptide.

- 68 -

A fusion protein can be produced by standard recombinant techniques. For example, DNA fragments coding for the different polypeptide sequences can be ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, *e.g.*, Ausubel *et al.* (eds.) Current Protocols in Molecular Biology, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). For example, an *AbCAS* polypeptide-encoding nucleic acid can be cloned into such an expression vector such that nucleic acid encoding *AbCAS* is linked in-frame to a polypeptide encoding a protein purification tag, such as a His tag. In another example, a nucleic acid molecule encoding a diTPS polypeptide can be linked in-frame to a polypeptide encoding a GGPP synthase. The diTPS polypeptide and additional polypeptide can be linked directly, without a linker, or alternatively, linked indirectly in-frame with a linker.

## **G. EXAMPLES**

The following examples are included for illustrative purposes only and are not intended to limit the scope of the invention.

25

### **Example 1**

#### **Terpenoid Metabolic Profiling of Balsam Fir**

A diterpenoid metabolic profile was established for wood and xylem, bark and phloem, and needles of balsam fir sapling trees to determine if it is a suitable tissue source for transcriptome mining of synthase genes involved in diterpenoid biosynthesis. Mono- and sesquiterpenoids were also measured.

30

#### **A. Methods**

##### **1. Plant Material**

- 69 -

Two-year old saplings of *Abies balsamea* var. *phanerolepsis* (L.) Mill. were purchased from Arbutus Grove Nursery Ltd. (North Saanich, BC, Canada) and maintained in a greenhouse as previously described in Miller *et al.* (*Plant Physiol.* 137:369-382 (2005)). Needles, xylem/wood and phloem/bark were harvested from the upper interwhorls and used fresh or flash frozen in liquid N<sub>2</sub> for metabolite and RNA extraction, respectively.

## 2. Diterpene Standards

Authentic diterpene resin acid standards were purchased from Orchid Celmark (New Westminster, BC, Canada). The corresponding diterpene olefins were synthesized from the acids at Best West Labs Inc. (Salt Lake City, UT, USA) as previously described (Ro *et al.* (2005) *Proc. Natl. Acad. Sci. USA* 102, 8060-8065).

## 3. Metabolite Analysis

Terpenoids were extracted from samples of 2 g of pulverized tissue with 1.5 mL of diethyl ether for 16 h at room temperature. Water was removed by addition of anhydrous Na<sub>2</sub>SO<sub>4</sub>, and extracts were passed through 0.22 µm GHP membrane filters (PALL Corporation, De Miniac, QC, Canada). GC-MS analysis was performed on an Agilent 6890N GC (Agilent Technologies Inc, Mississauga, ON, Canada), 7683B series autosampler, and a 5975 Inert XL MS Detector at 70 eV and 1 ml min<sup>-1</sup> He as carrier gas using a SGE Solgel-Wax column (polyethylene glycol, 30 m, 250 µm i.d., 0.25 µm film) with the following GC temperature program: 40° C for 5 min, 3° C min<sup>-1</sup> to 80° C, 8° C min<sup>-1</sup> to 250° C, 10° C min<sup>-1</sup> to 270° C, hold 5 min, pulsed splitless injector held at 250° C.

Diterpene resin acids were extracted from samples of approximately 150 mg tissue with 1.5 mL methyl *tert*-butyl ether following the method described in Lewinsohn *et al.* (*Plant Physiol.* 101:1021-1028 (1993)) and derivatized with 2 M trimethylsilyl diazomethane (Sigma, Oakville, ON, Canada). GC-MS analysis was performed using an Alltech AT-1000 column (polyethylene glycol-acid modified, 30 m, 250 µm i.d., 0.25 µm film) and GC specifications as follows: 150° C initial temperature, 1.5° C min<sup>-1</sup> to 220° C, 20° C min<sup>-1</sup> to 240° C.

Compound identification was achieved by comparison of mass spectra to those of authentic standards and reference mass spectral databases of the National Institute

- 70 -

of Standards and Technology (NIST) MS library searches (Wiley W9N08).

Quantifications were based on three independent biological replicates.

## B. Results

While only trace amounts of terpenoids were detected in wood tissue,  
 5 bark/phloem and needle samples had similar total amounts of terpenoids, made up  
 mostly of diterpenoids, with lesser amounts of monoterpenoids and relatively minor  
 amounts of sesquiterpenoids (see Table 3 below). Both the bark/phloem and the  
 needle samples contained the diterpene resin acids, abietic acid, isopimaric acid,  
 dehydroabietic acid and palustric acid, as major diterpenoids. These diterpene acids  
 10 are all commonly abundant in conifers. *Cis*-abienol was found almost exclusively in  
 the bark/phloem tissue, where it was the most abundant terpenoid metabolite,  
 accounting for more than 25% of the total amount of diterpenoids of balsam fir stem  
 tissue.

| <b>Table 3. Terpenoid metabolite profiles of <i>Abies balsamea</i> bark/phloem and needle tissues.</b> |                             |                         |
|--------------------------------------------------------------------------------------------------------|-----------------------------|-------------------------|
| <b>Terpenoid Metabolite</b>                                                                            | <b>µg g<sup>-1</sup> DW</b> |                         |
| <b><i>Monoterpenes</i><sup>1</sup></b>                                                                 | <b>Bark/Phloem</b>          | <b>Needles</b>          |
| α-Pinene <sup>1</sup>                                                                                  | 126 ± 1                     | 72 ± 19                 |
| Camphene <sup>1</sup>                                                                                  | 18 ± 15                     | 65 ± 12                 |
| β-Pinene <sup>1</sup>                                                                                  | 171 ± 41                    | 100 ± 36                |
| 3-Carene <sup>1</sup>                                                                                  | 111 ± 36                    | 188 ± 23                |
| Myrcene <sup>1</sup>                                                                                   | 11 ± 1                      | 10 ± 0                  |
| Limonene                                                                                               | 134 ± 17                    | 25 ± 5                  |
| β-Phellandrene <sup>1</sup>                                                                            | 67 ± 13                     | 20 ± 2                  |
| Terpinolene <sup>1</sup>                                                                               | 15 ± 7                      | 15 ± 2                  |
| Borneol <sup>1</sup>                                                                                   | 2 ± 1                       | 17 ± 7                  |
| Bornyl acetate <sup>1</sup>                                                                            | 130 ± 121                   | 305 ± 29                |
| <b>Total</b>                                                                                           | <b>785 ± 253</b>            | <b>817 ± 135</b>        |
| <b><i>Sesquiterpenes</i><sup>1</sup></b>                                                               |                             |                         |
| Longipinene <sup>1</sup>                                                                               | 11 ± 8                      | 5 ± 2                   |
| Longifolene <sup>1</sup>                                                                               | 51 ± 36                     | 2 ± 1                   |
| α-Humulene <sup>1</sup>                                                                                | 5 ± 3                       | 12 ± 11                 |
| β-Bisabolene <sup>1</sup>                                                                              | 26 ± 13                     | 34 ± 19                 |
| <i>trans</i> -Caryophyllene <sup>1</sup>                                                               | 11 ± 7                      | 22 ± 20                 |
| <b>Total</b>                                                                                           | <b>104 ± 67</b>             | <b>75 ± 53</b>          |
| <b><i>Diterpenes</i><sup>1,2</sup></b>                                                                 |                             |                         |
| Isopimaradiene <sup>1</sup>                                                                            | 16 ± 1                      | traces                  |
| (epi)-Manoyl oxide <sup>1</sup>                                                                        | 33 ± 3                      | traces                  |
| Abietadiene <sup>1</sup>                                                                               | 15 ± 5                      | 8 ± 1                   |
| <b><i>cis</i>-abienol<sup>2</sup></b>                                                                  | <b>3287 ± 25</b>            | <b>n.d.<sup>*</sup></b> |

- 71 -

|                                    |            |            |
|------------------------------------|------------|------------|
| Abietadieneal <sup>1</sup>         | 35 ± 20    | 9 ± 2      |
| Pimaric acid <sup>2</sup>          | traces     | n.d.*      |
| Isopimaric acid <sup>2</sup>       | 1751 ± 422 | 259 ± 244  |
| Sandaracopimaric acid <sup>2</sup> | 437 ± 298  | 160 ± 60   |
| Palustric acid <sup>2</sup>        | 1301 ± 1   | 640 ± 243  |
| Levopimaric acid <sup>2</sup>      | 682 ± 56   | 1364 ± 704 |
| Abietic acid <sup>2</sup>          | 2037 ± 830 | 717 ± 220  |
| Dehydroabietic acid <sup>2</sup>   | 1262 ± 742 | 1617 ± 199 |

<sup>1,2</sup>Terpenoid metabolites extracted with diethyl ether<sup>1</sup> or methyl tert-butyl ether<sup>2</sup>;

\*n.d., not detected

## Example 2

### 5 Isolation and Sequencing of Diterpene Synthases (diTPS)

Bark/phloem of balsam fir sapling stems was used as source for the preparation of a non-normalized cDNA library and subsequent transcriptome sequencing for identification of diterpene synthases (diTPS).

#### A. Identification of Candidate diTPS

10 Total RNA was isolated from samples of 150 mg bark tissue according to the method of Kolosova *et al.* (*BioTechniques* 36:821-824 (2004)) and mRNA was purified on Dynabeads (Invitrogen, Burlington, ON, Canada). RNA integrity and amounts were determined on a Bioanalyzer 2100 using an RNA Pico Chip (Agilent).

Construction of a non-normalized bark tissue cDNA library and subsequent  
15 454 transcriptome sequencing was conducted at the McGill University and Génome Québec Innovations Centre (Montreal, QC, Canada). The cDNA library was constructed from 200 ng of fragmented mRNA using the cDNA Rapid Library Preparation kit, GS FLX Titanium series (Roche Diagnostics, Laval, QC, Canada) following manufacturer's protocols. Yield of cDNA and the size range of cDNA  
20 fragments were assessed using a Bioanalyzer 2100 Pico Chip (Agilent), and 200 ng of the cDNA library were subjected to a half-plate reaction of 454 pyrosequencing using the Roche GS FLX Titanium technology. A half-plate reaction of Roche 454 sequencing generated a total of 797,060 sequence reads with an average GC content of 46% and an average read length of 359 bp.

25 After adapter trimming, the remaining high quality reads were subjected to a *de novo* assembly in GS *De novo* Assembler 2.5p1 with a size exclusion of 45 bp.

The assembly contained 85% of all input reads, yielding 14,699 isogroups from 17,122 isotigs of average size of 1,114 nucleotides.

A subset of candidate isotigs were identified by comparing the assembled sequences against 146 known plant TPSs using a BLASTx search (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410). This resulted in a subset of isotigs that represented candidate genes for nine different putative mono- and sesqui TPSs and five putative diTPSs (see Table 4 below).

| Table 4: Summary of <i>in silico</i> annotation of putative TPS assemblies. |       |                               |                      |                     |                |
|-----------------------------------------------------------------------------|-------|-------------------------------|----------------------|---------------------|----------------|
| Isotig                                                                      | Reads | BlastX hit                    | Species              | E-value             | Identity [%] * |
| 04065 (diTPS4)                                                              | 170   | Isopimaradiene synthase       | <i>P. abies</i>      | 0                   | 70             |
| 09468                                                                       | 170   | δ-Selinene synthase           | <i>A. grandis</i>    | 1.8 <sup>-155</sup> | 92             |
| 00922 (diTPS3)                                                              | 167   | Abietadiene synthase          | <i>A. grandis</i>    | 0                   | 94             |
| 01412                                                                       | 165   | Myrcene synthase              | <i>P. abies</i>      | 0                   | 70             |
| 00919 (diTPS1)                                                              | 163   | Abietadiene synthase          | <i>A. grandis</i>    | 0                   | 97             |
| 00921 (diTPS2)                                                              | 162   | Abietadiene synthase          | <i>A. grandis</i>    | 0                   | 92             |
| 04459                                                                       | 160   | β-Phellandrene synthase       | <i>A. grandis</i>    | 0                   | 92             |
| 04736                                                                       | 157   | α/β-Pinene synthase           | <i>A. grandis</i>    | 0                   | 97             |
| 04513                                                                       | 155   | Sabinene synthase             | <i>P. sitchensis</i> | 0                   | 76             |
| 09269                                                                       | 154   | 4 <i>S</i> -Limonene synthase | <i>A. grandis</i>    | 9.8 <sup>-161</sup> | 93             |
| 04902                                                                       | 153   | γ-Humulene synthase           | <i>A. grandis</i>    | 0                   | 92             |
| 07269                                                                       | 145   | <i>ent</i> -Kaurene synthase  | <i>P. sitchensis</i> | 0                   | 91             |
| 11523                                                                       | 141   | Linalool synthase             | <i>P. abies</i>      | 1.3 <sup>-113</sup> | 80             |
| 14736                                                                       | 128   | Limonene synthase             | <i>P. sitchensis</i> | 2.2 <sup>-90</sup>  | 91             |

\*Identity is given as amino acid sequence identity between isotig and Blast hit

**B. Isolation of Full-length cDNAs of bifunctional class I/II**

**gymnosperm diTPSs**

Of the diTPSs, four candidate isotigs (*AbdiTPS1*, *AbdiTPS2*, *AbdiTPS3*, and *AbdiTPS4*) resembled bifunctional class I/II gymnosperm diTPSs of the TPS-d group (Martin *et al.* (2004) *Plant Physiol* 135:1908-1927; Keeling *et al.* (2011) *BMC Plant Biol* 11:43; Chen *et al.* (2011) *The Plant Journal* 66:212-229) containing the characteristic DxDD, DDxxD and NSE/DTE motifs (SEQ ID NOS:12-14, respectively) (see Peters *et al.* (2003) *Biochemistry* 42:2700-2707; Peters & Croteau (2002) *Biochemistry* 41:1836-1842; Zhou & Peters (2009) *Phytochemistry* 70:366-

- 73 -

369). Reads corresponding to these four isotigs were reassembled in PHRAP (de la Bastide & McCombie (2007) *Curr Protoc Bioinformatics* Chapter 11, Unit 11.4), from which three unique partial diTPS cDNA sequences (AbdiTPS1, AbdiTPS2, AbdiTPS3) and one FLcDNA sequence (AbdiTPS4) were obtained.

5 A full length (FL) cDNA clone for *AbdiTPS4* was amplified based on the *in silico* assembled full-length sequence. For the other synthases, synthesis of cDNA as template for PCR amplification of candidate genes was carried out with random hexamer oligonucleotides using the SMARTer cDNA RACE amplification kit (Clontech, Mountain View, CA, USA). Completion of the 3'-sequences of  
10 *AbdiTPS1* and *AbdiTPS2* was achieved by rapid amplification of cDNA ends (3'-RACE) using the SMARTer cDNA RACE cDNA amplification kit (Clontech) and Phusion DNA-polymerase (New England Biolabs, Pickering, ON, Canada) with primers set forth in Table 5. The following PCR program was used: Initial denaturation at 98° C for 30 sec, 5 cycles of denaturation at 98° C for 20 sec and  
15 extension at 72° C for 90 sec, 5 cycles of denaturation at 98° C for 20 sec, annealing at 70° C for 20 sec and extension at 72° C for 90 sec, 27 cycles of denaturation at 98° C for 20 sec, annealing at 65° C for 20 sec and extension at 72° C for 90 sec, and a final extension at 72° C for 3 min. The obtained amplicons were gel-purified, and ligated into pJET (SEQ ID NO:32) using the CloneJET kit (Clontech).

| Table 5. Primers for RACE cDNA amplification |                                |           |
|----------------------------------------------|--------------------------------|-----------|
| Primer                                       | Sequence                       | SEQ ID NO |
| AbdiTPS1-3'RACE                              | CTGAGAGAGGTCAAGGTGAGGAGGCTTCTG | 18        |
| AbdiTPS2-3'RACE                              | CAGAGAGAGGTCAAGGTGAGGTGGCTTCTG | 19        |

20

Initial sequences of *AbdiTPS1*, *AbdiTPS2*, *AbdiTPS3*, and *AbdiTPS4* were obtained based on the initial high-throughput (lower accuracy) transcriptome reads and are set forth in Figure 7. Higher accuracy Sanger sequencing of the cloned cDNAs was performed on the *AbdiTPS1* and *AbdiTPS2* to confirm the sequences,  
25 which demonstrated some differences in the N-terminal sequences of *AbdiTPS1* and *AbdiTPS2* compared to that obtained from transcriptome reads. SEQ ID NOS corresponding to the protein and nucleic acid sequences of the isolated balsam fir diterpene synthases are set forth in Table 6 below. *AbdiTPS1* and *AbdiTPS4* are full

- 74 -

length sequences. *AbdiTPS2* is a near full-length sequencing lacking 40-50 basepair residues of the plastidial transit peptide and *AbdiTPS3* is a partial sequence.

| <b>Table 6. <i>AbdiTPS</i> Diterpene Synthases</b> |                              |                                   |
|----------------------------------------------------|------------------------------|-----------------------------------|
| <b>Diterpene Synthase</b>                          | <b>Protein<br/>SEQ ID NO</b> | <b>Nucleic Acid<br/>SEQ ID NO</b> |
| <i>AbdiTPS1</i>                                    | 1                            | 2                                 |
| <i>AbdiTPS2</i>                                    | 3                            | 4                                 |
| <i>AbdiTPS3</i>                                    | 5                            | 6                                 |
| <i>AbdiTPS4</i>                                    | 7                            | 8                                 |

5

### Example 3

#### Functional Characterization of Balsam Fir diTPSs and Discovery of a *cis*-Abienol Synthase

##### A. Cloning and Expression of Pseudomature Enzyme

For functional characterization of *AbdiTPS1*, *AbdiTPS2* and *AbdiTPS4*,  
 10 pseudomature versions of these diTPSs starting at the conserved N/KRx6W motif (SEQ ID NO:15) were cloned lacking the putative plastidial transit peptide so that expression in *Escherichia coli* (*E. coli*) could be achieved (Keeling *et al.* (2008) *Proc Natl Acad Sci USA* 105:1085-1090). To generate pseudomature versions, cDNAs described in Example 2.B were amplified using primers set forth in Table 7 to  
 15 generate products lacking the plastidial transit peptide. In addition, *AbdiTPS4* with a less conserved 5' sequence was cloned as a full length gene. For the amplification, the following PCR program was used: initial denaturation at 98° C for 30 sec followed by 30 cycles of denaturation at 98° C for 10 sec, annealing at 65-68° C for 30 sec and extension at 72° C for 80 sec, and final extension at 72° C for 10 min.  
 20 Amplified products were cloned into pJET (SEQ ID NO:32), and subcloned into the *NheI*/*Sall* and *NotI* restriction sites of pET28b(+) (SEQ ID NO:33; EMD Biosciences, San Diego, CA, USA). Using this expression vector, a 40 amino acid linker (MGRSHHHHHHSSGLVPRGSHMASMTGGQQMGRDPNSSSV; SEQ ID NO:23) was added to the 5' end of the gene-specific sequence, including the 6xHis-  
 25 tag, a T7-tag and a thrombin cleavage site. Hence, for the N-terminally tagged expression constructions, a methionine at which translation is initiated was located upstream of the His-tag. All cDNA constructs were sequence verified at the Nucleic

- 75 -

Acid Protein Service Unit (NAPS; University of British Columbia, BC, Canada) prior to expression in *E. coli*.

Recombinant proteins were expressed in *E. coli* BL21DE3-C41 cells, Ni<sup>+</sup> affinity purified as described elsewhere (Keeling *et al.* (2008) *Proc Natl Acad Sci USA* 105:1085-1090), and desalted against 20 mM HEPES (pH 7.2), 150 mM NaCl, 10% glycerol, 5 mM DTT using PD MiniTrap G-25 columns (GE Healthcare, Piscataway, NJ, USA), resulting in soluble proteins of the expected molecular weight of 90 to 91 kDa. SEQ ID NOS corresponding to the protein and encoding nucleic acid sequences of the expressed pseudomature form of *AbdiTPS1*, *AbdiTPS2* and *AbdiTPS4* are set forth in Table 8 below.

| Table 7. Primers for cDNA amplification |                                   |           |
|-----------------------------------------|-----------------------------------|-----------|
| Primer                                  | Sequence                          | SEQ ID NO |
| AbdiTPS4-For-FL-SalI                    | TAGTCGACATGGCCCTGCCTGTCTATTC      | 20        |
| AbdiTPS4-ForΔ85-SalI                    | TAGTCGACATGCGAGAATTCCTCCTTCATTTTG | 21        |
| AbdiTPS4-Rev-NotI                       | TAGCGGCCGCTTAGGTAGCCGGCTCGAAG     | 22        |
| AbdiTPS2-ForΔ68-NheI                    | TATGTCGACAAACGAGAATTCCTCCAGGA     | 24        |
| AbdiTPS2-Rev-NotI                       | TGGGCGGCCGCTTACGCAATCGGTATGAAGAG  | 25        |
| <i>AbdiTPS1</i> -For-Δ83-SalI           | TATGTCGACAAACGAGAATTCCTCCAGGA     | 26        |
| AbdiTPS1-Rev-NotI                       | TGGGCGGCCGCTTAGGCAACTGGTTGGAAGAG  | 27        |

| Table 8. Pseudomature Versions of diTPSs |                      |                           |
|------------------------------------------|----------------------|---------------------------|
| Diterpene Synthase                       | Protein<br>SEQ ID NO | Nucleic Acid<br>SEQ ID NO |
| <i>AbdiTPS1</i>                          | 47                   | 51                        |
| <i>AbdiTPS2</i>                          | 48                   | 52                        |
| <i>AbdiTPS4</i>                          | 55                   | 56                        |

## B. Characterization of Enzyme Activity

Enzyme assays with geranylgeranyl pyrophosphate (GGPP) as substrate followed by GC-MS or LC-MS analysis of reaction products against controls and authentic standards was used to identify functions for *AbdiTPS1*, *AbdiTPS2* and *AbdiTPS4*. Enzyme assays were carried out as described before (Keeling *et al.* (2008))

- 76 -

*Proc Natl Acad Sci USA* 105:1085-1090). Assays were carried out in 50 mM HEPES (pH 7.2), 10  $\mu$ M MgCl<sub>2</sub>, 5% glycerol, 5 mM DTT, using 100  $\mu$ g of purified protein (100  $\mu$ g each for coupled assays) and 15  $\mu$ M of (E,E,E)-GGPP (Sigma) with incubation for 1 h at 30° C. After extraction of reaction products with 500  $\mu$ L pentane, GC-MS analysis was conducted on an Agilent 7890A GC, 7683B series autosampler, and a 7000A Triple Quad MS Detector at 70 eV and 1.2 ml min<sup>-1</sup> He flow using a HP5ms column (5% phenyl methyl siloxane, 30 m, 250  $\mu$ m i.d., 0.25  $\mu$ m film) and the following GC temperature program: 40° C for 2 min, 20° C min<sup>-1</sup> to 300° C, hold 2 min; pulsed splitless injection (injector held at 250° C).

For the detection of diphosphate intermediates, reaction products were dephosphorylated prior to extraction by incubation with 10 U of calf intestinal alkaline phosphatase (Invitrogen) for 16 h at 37° C and analyzed on a Solgel-Wax column as described above in Example 1 with modified GC parameters: 40° C for 2 min, 25° C min<sup>-1</sup> to 250° C, hold 5 min, pulsed splitless injector held at 250° C.

Analysis of reaction products via LC-MS was performed on an Agilent 1100 Series LC/MSD Trap XCT Plus MS with atmospheric pressure chemical ionization (APCI) in positive mode on an Agilent Zorbax RX-Sil silica column (4.6 mm ID x 150 mm x 5  $\mu$ m) as previously reported (Keeling *et al.* (2011) *J. Biol. Chem.* 286, 21145-21153).

Authentic standards of abietadiene, isopimaradiene, levopimaradiene, neoabietadiene and palustradiene were used as positive controls. Empty vector was used as a negative control. As shown in Figure 2, the authentic standard of abietadiene contained an unknown contamination and the authentic standard of levopimaradiene contained 30% other diterpene isomers, which could not be removed.

Enzymatic activity assays were confirmed with three independent experiments. The product profile results are set forth in Figure 2, and summarized as follows:

#### 1. *AbdiTPS1*

The product profile of *AbdiTPS1* was identified by GC-MS to be made up of four peaks corresponding to abietadiene, levopimaradiene, neoabietadiene and palustradiene in a ratio of approximately 4:3:2:1 (see Figure 2). This profile closely matches that previously identified for grand fir *Abies grandis* abietadiene synthase

- 77 -

(*AgAS*) (NCBI accession no. AAK83563, SEQ ID NO:40) (Peters *et al.* (2002) *Biochemistry* 39:15592-15602). As shown in Example 5 and Figure 7, *AgAS* is a diTPS that is 99% identical with *AbdiTPS1* on the amino acid level. According to these patterns, *AbdiTPS1* can be classified as an LAS-type bifunctional class I/II diTPS (referred to herein also as *AbLAS*), which appears to be orthologous with *AgAS* (SEQ ID NO:40).

Following the recent discovery of epimers of a tertiary alcohol, 13-hydroxy-8(14)-abietene, as the initial enzyme products of Norway spruce *Picea abies* levopimaradiene/abietadiene synthase (*PaLAS*) (SEQ ID NO:37; Keeling *et al.* (2011) *J. Biol. Chem.* 286:21145-21153), the product of *AbdiTPS1* (*AbLAS*) by LC-MS was also investigated. As shown in Figure 3, the detected mass fragment of  $m/z$  273 in the LC-MS analysis corresponds to the dehydration product of 13-hydroxy-8(14)-abietene, and is indicative of dehydration of the unstable diterpene alcohol compounds in the APCI interface, as detailed in Keeling *et al.* (Keeling *et al.* (2011) *J. Biol. Chem.* 286:21145-21153). Comparison with abietadiene standard showed separation of olefin compound and the polar hydroxylated diterpenes.

## 2. *AbdiTPS2*

The product profile of *AbdiTPS2* was identified by GC-MS to be made up of a single peak demonstrating that *AbdiTPS2* formed exclusively isopimaradiene (see Figure 2). This product profile is similar to the previously characterized single-product isopimaradiene synthases from Norway spruce (*PaIso*; SEQ ID NO:35) and Sitka spruce (*PsIso*; SEQ ID NO:36) (Martin *et al.* (2004) *Plant Physiol* 135:1908-1927; Keeling *et al.* (2008) *Proc Natl Acad Sci USA* 105:1085-1090; Keeling *et al.* (2011) *BMC Plant Biol* 11:43). *AbdiTPS2* was thereby identified as an Iso-type bifunctional class I/II diTPS (referred to herein also as *AbIso*).

## 3. *AbdiTPS4*

The product profile of *AbdiTPS4* was identified by GC-MS to be a unique single peak product profile (see Figure 2 and Figure 3). The product corresponded to *cis*-abienol according to retention time consistent with the compound extracted from plant tissue and shared characteristic mass fragments (e.g.,  $m/z$  290 [M<sup>+</sup>],  $m/z$  272 [M<sup>+</sup>-H<sub>2</sub>O], and  $m/z$  134) in comparison with reference mass spectra (Vlad *et al.* (1974) *Khimiya Prirodnikh Soedinenii* 1:30-35) and obtained from the National

- 78 -

Institute of Standards and Technology MS library searches (Wiley W9N08). An authentic *cis*-abienol standard is not commercially available. When the product of *AbdiTPS4* was analyzed by GC-MS, under a number of different conditions, poor resolution of the shape of the *cis*-abienol peak (see Figure 2) was observed, relative to the peak shape of diterpene olefins, and the occurrence of trace amounts of two additional compounds. This profile is likely due to degradation of *cis*-abienol during GC-MS as previously reported (Severson *et al.* (1984) *Journal of Agricultural and Food Chemistry* 32:566-570; Ding *et al.* (2007) *Chromatographia* 66:529-532; Carman & Duffield (1993) *Aust. J. Chem.* 46:1105-1114). LC-MS analysis confirmed *cis*-abienol as a single product of *AbdiTPS4*, with *m/z* 273 in the LC MS analysis corresponding to the predicted dehydration product of the diterpene alcohol (see Figure 3). *AbdiTPS4* was thereby identified as a bifunctional class I/II *cis*-abienol synthase (referred to herein also as *AbCAS*). *AbCAS* represents a new type of conifer diTPS, producing a bicyclic, tertiary diterpene alcohol, as opposed to tricyclic products of the LAS- and Iso-type diTPSs.

Additional structural and stereochemical analysis by proton and carbon NMR and comparison to previously reported analyses (Ding *et al.* (2007) *Chromatographia*, 66:529-532; Carman *et al.* (1993) *Aust. J. Chem.*, 46:1105-1114) confirmed the identity of the *AbdiTPS4* product as *cis*-abienol. For Nuclear magnetic resonance (NMR) analysis, *Cis*-abienol was prepared from a pool of ten individual enzyme assays as described above using 20  $\mu$ M GGPP and an incubation time of 2 h to maximize product formation. To enhance product purity, pentane was purified on alumina prior to use and assays were performed in buffer pre-extracted with alumina-purified pentane. Nuclear magnetic resonance (NMR) spectra were recorded on a Bruker Avance 600 equipped with a QNP or TCI cryoprobe (600 MHz) using deuteriochloroform ( $\text{CDCl}_3$ ) as the solvent, which was neutralized by distillation and filtration through potassium carbonate prior to preparation of the sample. Signal positions ( $\delta$ ) were calculated in parts per million (ppm) as compared to tetramethylsilane ( $\delta$  0) and were measured relative to the signal of the solvent ( $\text{CDCl}_3$ :  $\delta$  7.26,  $^1\text{H}$  NMR;  $\delta$  77.0,  $^{13}\text{C}$  NMR).

#### Example 4

- 79 -

### Phylogenetic Analysis of Bifunctional class I/II Gymnosperm diTPSs

To perform phylogenetic analysis, the encoded amino acid sequences of *AbdiTPS1*, *AbdiTPS2*, *AbdiTPS3* and *AbdiTPS4* were compared to the amino acid sequence of other synthases. Multiple protein sequence alignments were performed using the CLC bio Main Workbench 5.7.1 (CLC bio, Århus, Denmark). Phylogenetic analyses were conducted on the basis of the maximum likelihood algorithm using PhyML 3.0 (Guindon *et al.* (2010) *Syst. Biol* 59:307-321) with four rate substitution categories, LG substitution model, BIONJ starting tree and 100 bootstrap repetitions, and displayed as phylogram using treeview32 1.6.6, by rooting with the outgroup *Physcomitrella patens* *ent*-copalyl diphosphate/*ent*-kaurene synthase (*PpCPS/KS*; NCBI accession no. BAF61135; SEQ ID NO:42) and diterpene synthases *TbrTS*, *Taxus brevifolia* taxadiene synthase (NCBI accession no. AAC49310; SEQ ID NO:43); *TcuTS*, *T. cuspidata* taxadiene synthase (NCBI accession no. ABW82997; SEQ ID NO:44); *TxmTS*, *T. x media* taxadiene synthase (NCBI accession no. AAS18603; SEQ ID NO:45); *PaIso*, *Picea abies* isopimaradiene synthase (NCBI accession no. AAS47690; SEQ ID NO:35); *PsIso*, *Picea sitchensis* isopimaradiene synthase (NCBI accession no. ADZ45512; SEQ ID NO:36); *PaLAS*, *P. abies* (NCBI accession no. AAS47691; SEQ ID NO:37); *PsLAS*, *P. sitchensis* levopimaradiene/abietadiene synthase (NCBI accession no. ADZ45517; SEQ ID NO:38); *PtLAS*, *Pinus taeda* levopimaradiene synthase (NCBI accession no. AY779541; SEQ ID NO:39); *AgAS*, *Abies grandis* abietadiene synthase (NCBI accession no. AAK83563; SEQ ID NO:40); and *GbLS*, *Ginkgo biloba* levopimaradiene synthase (NCBI accession no. AAL09965; SEQ ID NO:41).

The encoded proteins of *AbdiTPS1*, *AbdiTPS2*, and *AbdiTPS3* showed highest similarity with grand fir (*Abies grandis*) abietadiene synthase (*AgAS*; SEQ ID NO:40) (Stofer Vogel *et al.* (1996) *J. Biol. Chem* 271:23262-23268; Peters *et al.* (2000) *Biochemistry* 39:15592-15602), with more than 90% protein sequence identity. In contrast, the deduced protein sequence *AbdiTPS4* showed highest similarity with Norway spruce (*Picea abies*) isopimaradiene synthase (*PaIso*; SEQ ID NO:35) (Martin *et al.* (2004) *Plant Physiol* 135:1908-1927), although on a substantially lower level of sequence identity of only 75%, highlighting *AbdiTPS4* as a unique candidate diTPS sequence.

- 80 -

The results of the phylogentic analysis are shown in Figure 4. The results show that other conifers contain diTPSs of the levopimaradiene synthase (LAS) and isopimaradiene synthase (Iso) types. LAS- or Iso type diTPSs have now been identified in three different genera of the pine family, namely in true firs (*Abies*),  
 5 spruce (*Picea*), and pine (*Pinus*), which allows for analysis of gene orthology within this family. It appears that gene duplications and neo-functionalization leading to paralogous pairs of LAS and Iso genes occurred independently in *Abies* and *Picea*, after the separation of these genera. Within the spruce genus, the dichotomy of LAS- and Iso- genes happened apparently prior to the speciation of Norway spruce and  
 10 Sitka spruce.

Within the group of conifer class I/II diTPSs, the *AbdiTPS4* (*AbCAS*) gene is separate from, and appears basal to, the clade of levopimaradiene synthase (LAS) and isopimaradiene synthase (Iso) genes from firs, spruces, and pine. The bifunctional class I/II *AbCAS* described here has only been cloned from balsam fir, matching the  
 15 major diterpenoid produced by this species as described in Example 1 (see also Table 3).

### Example 5

#### Analysis of the Reaction Sequence of the Bifunctional Class I/II *AbCAS* by Site

#### 20 Directed Mutagenesis

A mechanism for the formation of *cis*-abienol requires hydroxylation via water quenching of a labda-13-en-8-yl carbocation at C-8 (see Figure 1A). In principle, hydroxylation can occur during reactions at the class II active site or at the class I active site. To delineate which of the two active sites of bifunctional *AbCAS*  
 25 (*AbdiTPS4*; SEQ ID NO:7) catalyzed the formation of the tertiary alcohol, a set of alanine substitutions of the DxDD (class II active site) (SEQ ID NO:12) and DDxxD (class I active site) (SEQ ID NO:13) motifs to obtain monofunctional *AbCAS* variants were generated.

#### 30 A. Generation of Site Directed Protein Variants of *AbdiTPS4* (*AbCAS*)

Amino acid mutations were generated and tested in the pseudomature form of SEQ ID NO:7 as described in Example 3 (i.e. set forth in SEQ ID NO:55 and encoded

- 81 -

by the sequence of nucleotides set forth in SEQ ID NO:56). Alanine substitutions of Asp402, Asp404 or Asp621 of AbdiTPS4 with reference to positions set forth in SEQ ID NO:58 (corresponding to residues Asp403, Asp405 or Asp 622 with reference to SEQ ID NO:7) were generated by QuikChange site directed mutagenesis (Stratagene, Mississauga, ON, Canada) of template cDNA corresponding to the pseudomature form (SEQ ID NO:56) in the vector pET28b(+) (SEQ ID NO:33) using primers as set forth in Table 9 below. Specifically, monofunctional *AbCAS* variants were generated that contained either a non-functional class II (*AbCAS*:D402A/D404A) or a non-functional class I (*AbCAS*:D621A) active site. For the mutagenesis reaction, the following PCR program was used: Initial denaturation at 98° C for 90 sec followed by 29 cycles of denaturation at 98° C for 50 sec, annealing at 60° C for 30 sec and extension at 72° C for 4 min, and final extension at 72° C for 10 min.

| Table 9. Primers for cDNA amplification |                                             |           |
|-----------------------------------------|---------------------------------------------|-----------|
| Primer                                  | Sequence                                    | SEQ ID NO |
| D402A-D404A-sense                       | GACAGCCCTATAGCCGCTATTGCTGATACAGCCA<br>TGGGC | 28        |
| D402A-D404A-antisense                   | GCCCATGGCTGTATCAGCAATAGCGGCTATAGGG<br>CTGTC | 29        |
| D621A-sense                             | TCTCCTCGTCATTTTAGCCGACCTTTACGATGGG          | 30        |
| D621A-antisense                         | CCCATCGTAAAGGTCGGCTAAATGACGAGGAGA           | 31        |

### B. Activity Analysis of Protein Variants

Activity analysis of protein variants of *AbdiTPS4* (*AbCAS*) was conducted with 20  $\mu$ M GGPP as substrate and dephosphorylation of the reaction products prior to GC-MS analysis. The results are set forth in Figure 5. The results showed that *AbCAS*:D621A converted GGPP (peak 9 in Figure 5) into trace amounts of CPP (peak 10 in Figure 5) and a product with a significantly longer retention time, indicative of a hydroxylated diphosphate (peak 13 in Figure 5). This compound was identified, upon cleavage of the diphosphate group, as labda-13-en-8,15-diol based on characteristic mass fragments ( $m/z$  290 [M+],  $m/z$  275,  $m/z$  257,  $m/z$  192 and  $m/z$  177) as compared to the mass spectrum of the authentic compound (Falara *et al.* (2010) *Plant Physiol.* 154:301-310). Epi-manoyl oxide and manoyl oxide (peaks 11 and 12 in Figure 5) were apparently not products of *AbCAS*, but rather result from ether formation of the dephosphorylated labda-13-en-8-ol diphosphate under GC-MS

- 82 -

conditions, as they were not seen in the wild type enzyme product profile and did not occur under different GC conditions. Results from enzyme assays with *AbCAS*:D621A showed that the class II active site of *AbCAS* catalyzes protonation-initiated formation of bicyclic labda-13-en-8-ol diphosphate via water quenching of the parental carbocation.

Alanine substitution of Asp402 and Asp404 (*AbCAS*:D402A/D404A) in the class II active site resulted in complete loss of enzymatic activity with GGPP as substrate. To determine if the activity could be restored, a coupled assay was performed combining the *AbCAS*:D621A variant with a point mutant of *PaLAS* (*PaLAS*:D611A; SEQ ID NO:34) containing a non-functional class I active site that accumulates (9S,10S)-CPP (peak 10 in Figure 5). Activity of *AbCAS*:D402A/D404A could not be restored by combining this mutant with a protein variant of *PaLAS* (*PaLAS*:D611A). This result shows that, unlike the LAS- and Iso-type enzymes, the class I active site of *AbCAS* is not active with (9S,10S)-CPP as an intermediate substrate.

Coupled assays also were performed using the two monofunctional *AbCAS* class I and class II protein variants, *AbCAS*:D402A/D404A and *AbCAS*:D621A. Formation of *cis*-abienol (peak 8 in Figure 5) from GGPP was restored in coupled assays, confirming that the non-mutated active sites remained functional in the two mutated proteins.

The mutational analysis and complementation assays show that the class I active site of *AbCAS* is selective for labda-13-en-8-ol diphosphate as a substrate, but does not convert (9S,10S)-CPP, which is the class I active site substrate of LAS-, and Iso type diTPSs (see Figure 5). Unlike LAS- and Iso-enzymes, the class I active site of *AbCAS* catalyzes the ionization of the diphosphate group without cyclization of a C-ring (see Figure 1). Together, the mutational analysis of class I and class II active sites of *AbCAS* demonstrated a reaction sequence of the bifunctional class I/II enzyme in which both the bicyclization and water capture occur in the class II active site, resulting in an intermediary labda-13-en-8-ol diphosphate, which undergoes cleavage of the diphosphate group and final deprotonation at the class I active site (see Figure 1).

### Example 6

- 83 -

### Computational Structure Analysis

Homology models of *AbLAS* (*AbTPS1*), *AbIso* (*AbTPS2*) and *AbCAS* (*AbTPS4*) were built using the CPHmodels 3.0 server (Nielsen *et al.* (2010) *Nucleic Acids Res.* 38:W576-581) based on the tertiary structure of *Taxus brevifolia* taxadiene synthase (*TbTXS*; PDB-ID 3p5pA, chain A; SEQ ID NO:46) (Köksal *et al.* (2011) *Nature* 469:116-120). The homology models were certified as high quality exceeding 91% residues assigned to most favored regions in Ramachandran plot statistics using PROCHECK (Laskowski, *et al.* (1993) *J Applied Crystallography* 26:283-291). Lack of structural errors in the models was validated using the ProSA-web server (Wiederstein & Sippl, (2007) *Nucleic Acids Res.* 35:W407-410). Pairwise comparison of these modeled structures with the DaliLite server (Holm & Park, (2000) *Bioinformatics* 16:566-567) demonstrated a high structural similarity of *AbLAS*, *AbIso*, *AbCAS* and *TbTXS* with root mean square deviations of  $\leq 1 \text{ \AA}$  for the  $\beta\gamma$  domain and the  $\alpha$  domain, respectively, thus allowing for a structural comparison of the active sites.

Using Molegro Virtual Docker 2010.4.0.0 (Thomsen & Christensen, (2006) *J. Med. Chem.* 49:3315-3321), a semi-automated docking approach was applied to place labda-13-en-8-ol diphosphate in the class I and class II active site of *AbCAS* (see Figure 6). For this purpose, proposed active site cavities were identified using a probe radius of 1.0  $\text{\AA}$  and a grid resolution of 0.6  $\text{\AA}$  with further manual optimization. Substrate docking was then restricted to the cavities, inclusive of the required  $\text{Mg}^{2+}$ -complex. Energy minimized PDB coordinates of the ligands were generated using the PRODRG server (Schüttelkopf & van Aalten, (2004) *Acta Crystallogr. D Biol. Crystallogr.* 60:1355-1363).

Structural modeling and molecular docking of labda-13-en-8-ol diphosphate in the class I and class II active sites of *AbCAS* (Figure 6) revealed only one unique amino acid in the class II active site, namely Asp348 with reference to numbering set forth in SEQ ID NO:58 (corresponding to Asp349 as set forth in SEQ ID NO:7), in proximity of the docked intermediate. Its locations shows that Asp348 is likely to control this particular hydroxylation reaction. Located at the posterior of the active site opposite of the DIDD motif (SEQ ID NO:16) and Trp358, which have previously been reported to contribute to the *AgAS*-catalyzed class II reaction (Peters & Croteau

- 84 -

(2002) *Biochemistry* 41:1836-1842), the negatively charged side chain of Asp348 is positioned to stabilize the positive charge at C-8 for water quenching to occur in the formation of labda-13-en-8-ol diphosphate (Figure 6B). This local negative charge is not present in any LAS- or Iso-type enzymes, which contain a conserved histidine in this position (see Figure 7).

Several unique residues were found in a radius of 7Å around the hydroxy group of labda-13-en-8 ol diphosphate docked within the class I active site of *AbCAS* that are likely involved in the *AbCAS* reaction (see Figures 6C and 6D). Among these, Leu617, Phe696 and Gly723 with reference to numbering set forth in SEQ ID NO:58 (corresponding to residues Leu618, Phe697 and Gly 724 as set forth in SEQ ID NO:7) are of particular interest as they contribute to contour of the active site cavity. Leu617 and Phe696 are located in the class I active site cavity upstream of the DDxxD motif (SEQ ID NO:13) on helix D, creating an expansion of the hydrophobic pocket relative to *AbLAS* and *AbIso*. Gly723 and Val724 account for a change in the hinge region between helix G1 and G2. These residues are likely to contribute to the release of a bicyclic product rather than facilitating a secondary cyclization. Interestingly, residues corresponding to Leu617, Phe696 and Gly723 have previously been shown to be critical for the catalytic plasticity of conifer diTPSs (Keeling *et al.* (2008) *Proc Natl Acad Sci USA* 105:1085-1090; Wilderman & Peters (2007) *J. Am. Chem. Soc.* 129:15736-15737; Peters & Croteau (2002) *Proc. Natl. Acad. Sci. USA* 99:580-584; Leonard *et al.* (2010) *Proc. Natl. Acad. Sci. USA* 107:13654-13659), and thus appear to represent positions for the functional evolution of these enzymes.

### Example 7

#### Production of Ambroxide

In this example, *cis*-abienol is produced from geranylgeranyl diphosphate (GGPP) in the presence of an *AbCAS* (*AbTSP4*) as described herein. Then, (-)-ambroxide is produced from the produced *cis*-abienol as previously described in the art (see, e.g., Barrero *et al.* (1993) *Tetrahedron* 49(45): 10405-10412; Barrero *et al.* (1994) *Tetrahedron* 50:6653-6662; Barrero *et al.*, (2004) *Synthetic Communications* 34(19):3631-3643; and U.S. Pat. No. 5,525,728).

#### A. Scheme 1

- 85 -

As shown in Figure 8A, Scheme I, *cis*-abienol is subjected to ozonolysis by reaction with ozone in methylene chloride at -78° C followed by reduction with lithium aluminum hydride affords the diol (9). The diol (9) is purified by washing the organic phase with 10% HCl, saturated NaHCO<sub>3</sub> and water, followed by drying with anhydrous Na<sub>2</sub>SO<sub>4</sub> and evaporation of the organic solvent. The diol (9) is cyclized in the presence of tosylchloride and the weak base pyridine at room temperature for 1 hour resulting in (-)-ambroxide. (-)-Ambroxide is purified by extraction with organic solvent.

**B. Scheme 2**

The production of (-)-ambroxide from a starting compound of geranylgeranyl diphosphate (GGPP) is accomplished as described in Figure 8B, Scheme II. GGPP is converted to *cis*-abienol by the diterpene synthase *AbCas* (SEQ ID NO:7). *Cis*-abienol is isolated by extraction with organic solvent. *Cis*-abienol is subjected to ozonolysis by reaction with ozone in methylene chloride at -78° C followed by reduction with lithium aluminum hydride affords the diol (9). The diol (1) is cyclized in the presence of tosylchloride and the weak base pyridine at room temperature for 1 hour resulting in (-)-ambroxide. (-)-Ambroxide is purified by extraction with organic solvent.

Since modifications will be apparent to those of skill in this art, it is intended that this invention be limited only by the scope of the appended claims.

- 86 -

# CLAIMS:

1. An isolated nucleic acid molecule, comprising a sequence of nucleotides encoding a bifunctional class I/II cis-abienol synthase (CAS) polypeptide or an active fragment, wherein:
  - 5 the encoded polypeptide or active fragment comprises a class II active site comprising a DxDD motif and a class I active site comprising a DDxxD motif; and the encoded polypeptide or active fragment catalyzes the formation of cis-abienol from geranylgeranyl diphosphate (GGPP).
2. An isolated nucleic acid molecule of claim 1, comprising the sequence
  - 10 of nucleotides encoding a cis-abienol synthase (CAS) polypeptide set forth in SEQ ID NO:7, an active fragment thereof, or a sequence that exhibits at least 75% sequence identity to SEQ ID NO:7, wherein the encoded polypeptide or active fragment catalyzes the formation of cis-abienol from geranylgeranyl diphosphate (GGPP).
3. The isolated nucleic acid molecule of claim 1 or claim 2, wherein the
  - 15 sequence of nucleotides encodes a CAS polypeptide that exhibits at least 80%, 85%, 90%, 95% or 98% sequence identity to SEQ ID NO:7.
4. The isolated nucleic acid molecule of any of claims 1-3, wherein the sequence of nucleotides encodes a polypeptide that comprises the sequence of amino acids set forth in SEQ ID NO:7 or an active fragment thereof.
- 20 5. The isolated nucleic acid molecule of any of claims 1-3, wherein the sequence of nucleotides encodes a polypeptide that consists of the sequence of amino acids set forth in SEQ ID NO:7 or an active fragment thereof.
6. The isolated nucleic acid molecule of any of claims 1-5, wherein the active fragment is a pseudomature form.
- 25 7. The isolated nucleic acid molecule of any of claims 1-6, comprising the sequence of nucleotides set forth in SEQ ID NO:8 or a sequence of nucleotides that exhibits at least 75% sequence identity to SEQ ID NO:8 or the complement thereof, wherein the sequence of nucleotides encodes a polypeptide or active fragment that catalyzes the formation of cis-abienol from geranylgeranyl diphosphate (GGPP).
- 30 8. The isolated nucleic acid molecule of claim 7, comprising a sequence of nucleotides that exhibits at least 80%, 85%, 90%, 95% or 98% sequence identity to SEQ ID NO:8 or the complement thereof.

- 87 -

9. The isolated nucleic acid molecule of any of claims 1-8, comprising the sequence of nucleotides set forth in SEQ ID NO:8, an active fragment thereof or the complement thereof.

10. The isolated nucleic acid molecule of any of claims 1-8, consisting of the sequence of nucleotides set forth in SEQ ID NO:8, an active fragment thereof or the complement thereof.

11. A *cis*-abienol synthase (CAS) polypeptide encoded by the nucleic acid molecule of any of claims 1-10.

12. A vector, comprising the nucleic acid molecule of any of claims 1-10.

10 13. A host cell, comprising the vector of claim 12.

14. The host cell of claim 13 that is a prokaryotic or eukaryotic host cell.

15. The host cell of claim 13 or claim 14, wherein the host cell is selected from among a bacterial, fungal, plant, insect, amphibian or animal cell.

16. The host cell of claim 15 that is an *E. coli* cell.

15 17. The host cell of claim 15 that is a yeast cell.

18. The host cell of any of claims 13-17 that produces GGPP.

19. The host cell of any of claims 13-18 that is transformed with one or more sequences that result in production of GGPP.

20. A method for producing a bicyclic tertiary diterpene alcohol, comprising:

i) contacting (E,E,E)-geranylgeranyl diphosphate (GGPP) with a bifunctional class I/II synthase polypeptide having a diterpene synthase (diTPS) activity under conditions effective to produce bicyclic tertiary diterpene alcohols, wherein the bifunctional class I/II synthase is a conifer synthase;

25 ii) optionally, isolating the diterpenoids produced in step i).

21. A method of producing *cis*-abienol, comprising

i) contacting (E,E,E)-geranylgeranyl diphosphate (GGPP) with a bifunctional class I/II synthase polypeptide having a diterpene synthase (diTPS) activity under conditions effective to produce *cis*-abienol;

30 ii) optionally, isolating the *cis*-abienol produced in step i).

- 88 -

22. The method of claim 20 or claim 21, wherein the bifunctional class I/II synthase comprises a class II active site comprising a DxDD motif and a class I active site comprising a DDxxD motif.

23. The method of any of claims 20-22, wherein the bifunctional class I/II  
5 synthase is a *cis*-abienol synthase (CAS) polypeptide or active fragment that catalyzes the formation of *cis*-abienol from geranylgeranyl diphosphate (GGPP).

24. The method of claim 23, wherein the CAS polypeptide comprises the sequence of amino acids set forth in SEQ ID NO:7, an active fragment thereof, or a sequence that exhibits at least 75% sequence identity to SEQ ID NO:7, wherein the  
10 encoded polypeptide or active fragment catalyzes the formation of *cis*-abienol from geranylgeranyl diphosphate (GGPP).

25. The method of claim 24, wherein sequence of amino acids exhibits at least 80%, 85%, 90%, 95% or 98% sequence identity to SEQ ID NO:7.

26. The method of any of claims 23-25, wherein the CAS polypeptide  
15 comprises the sequence of amino acids set forth in SEQ ID NO:7 or an active fragment thereof.

27. The method of any of claim 23-25, wherein the CAS polypeptide consists of the sequence of amino acids set forth in SEQ ID NO:7 or an active fragment thereof.

28. The method of any of claims 23-27, wherein the active fragment is a  
20 pseudomature form.

29. The method of any of claims 23-28, wherein the CAS polypeptide is encoded by a sequence of nucleotides set forth in SEQ ID NO:8 or a sequence of nucleotides that exhibits at least 75% sequence identity to SEQ ID NO:8 or the  
25 complement thereof, wherein the sequence of nucleotides encodes a polypeptide or active fragment that catalyzes the formation of *cis*-abienol from geranylgeranyl diphosphate (GGPP).

30. The method of claim 29, wherein the sequence of nucleotides exhibits at least 80%, 85%, 90%, 95% or 98% sequence identity to SEQ ID NO:8 or the  
30 complement thereof.

- 89 -

31. The method of any of claims 23-30, wherein the CAS polypeptide is encoded by a sequence of nucleotides comprising the sequence set forth in SEQ ID NO:8 or an active fragment thereof.

32. The method of any of claims 23-30, wherein the CAS polypeptide is  
5 encoded by a sequence of nucleotides consisting of the sequence set forth in SEQ ID NO:8 or an active fragment thereof.

33. The method of any of claims 20-32, wherein the step of contacting E,E,E)-geranylgeranyl diphosphate (GGPP) with a bifunctional class I/II synthase polypeptide is effected *in vitro* or *in vivo*.

10 34. The method of any of claims 20-33, comprising isolating the the *cis*-abienol.

35. The method of any of claims 20-34, wherein the *cis*-abienol is used to produce (-)-ambroxide.

36. An isolated nucleic acid molecule, comprising the sequence of  
15 nucleotides encoding a levopimaradiene/abietadiene synthase (LAS) set forth in SEQ ID NO: 1, an active fragment thereof, or a sequence that exhibits greater than 99% sequence identity to SEQ ID NO:1, wherein the encoded polypeptide or active fragment catalyzes the formation of epimers of 13-hydroxy-8(14)-abietene from geranylgeranyl diphosphate (GGPP).

20 37. The isolated nucleic acid molecule of claim 36, wherein the sequence of nucleotides encodes a polypeptide that comprises the sequence of amino acids set forth in SEQ ID NO: 1 or an active fragment thereof.

38. The isolated nucleic acid molecule of claim 36, wherein the sequence of nucleotides encodes a polypeptide that consists of the sequence of amino acids set  
25 forth in SEQ ID NO:1 or an active fragment thereof.

39. The isolated nucleic acid molecule of any of claims 36-38, wherein the epimers are selected from among abietadine, levopimaradiene, neoabietadiene and plaustradiene.

40. The isolated nucleic acid molecule of any of claims 36-39, wherein the  
30 active fragment is a pseudomature form.

41. The isolated nucleic acid molecule of any of claims 36-40, comprising the sequence of nucleotides set forth in SEQ ID NO:2 or a sequence of nucleotides

- 90 -

that exhibits greater than 99% sequence identity to SEQ ID NO:2 or the complement thereof, wherein the sequence of nucleotides encodes a polypeptide or active fragment that catalyzes the formation of epimers of 13-hydroxy-8(14)-abietene from geranylgeranyl diphosphate (GGPP).

5           42.     The isolated nucleic acid molecule of claim 41, comprising the sequence of nucleotides set forth in SEQ ID NO:2, an active fragment thereof or the complement thereof.

            43.     The isolated nucleic acid molecule of claim 41, consisting of the sequence of nucleotides set forth in SEQ ID NO:1, an active fragment thereof or the  
10   complement thereof.

            44.     A levopimaradiene/abietadiene synthase (LAS) polypeptide encoded by the nucleic acid molecule of any of claims 36-43.

            45.     A vector, comprising the nucleic acid molecule of any of claims 36-43.

            46.     A host cell, comprising the vector of claim 45.

15           47.     The host cell of claim 46 that is a prokaryotic or eukaryotic host cell.

            48.     The host cell of claim 46 or claim 47, wherein the host cell is selected from among a bacterial, fungal, plant, insect, amphibian or animal cell.

            49.     The host cell of any of claims 46-48 that is an *E. coli* cell.

            50.     The host cell of any of claims 46-48 that is a yeast cell.

20           51.     The host cell of any of claims 46-50 that produces GGPP.

            52.     The host cell of any of claims 46-51 that is transformed with one or more sequences that result in production of GGPP.

            53.     A method of producing an epimer of 13-hydroxy-8(14)-abietene, comprising

25           i) contacting (E,E,E)-geranylgeranyl diphosphate (GGPP) with a levopimaradiene/abietadiene synthase (LAS) polypeptide encoded by the nucleic acid molecule of any of claims 36-43 under conditions effective to produce an epimer of 13-hydroxy-8(14)-abietene;

            ii) optionally, isolating the epimer of 13-hydroxy-8(14)-abietene produced in step  
30   i).

            54.     The method of claim 53, wherein the epimer is selected from among abietadine, levopimaradiene, neoabietadiene and plaustradiene.

- 91 -

55. The method of claim 53 or claim 54, wherein the step of contacting E,E,E)-geranylgeranyl diphosphate (GGPP) with a levopimaradiene/abietadiene synthase (LAS) polypeptide is effected *in vitro* or *in vivo*.

56. The method of any of claims 53-55, comprising isolating one or more  
5 of the diterpenes selected from among abietadiene, levopimaradiene, neoabietadiene and plaustadiene.

57. An isolated nucleic acid molecule, comprising the sequence of nucleotides encoding a isopimaradiene synthase (Iso) polypeptide set forth in SEQ ID NO: 3, an active fragment thereof, or a sequence that exhibits at least 95% sequence  
10 identity to SEQ ID NO:3, wherein the encoded polypeptide or active fragment catalyzes the formation of isopimaradiene from geranylgeranyl diphosphate (GGPP).

58. The isolated nucleic acid molecule of claim 57, wherein the sequence of nucleotides encodes a polypeptide that exhibits at least 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:3.

59. The isolated nucleic acid molecule of claim 57 or claim 58, wherein  
15 the sequence of nucleotides encodes a polypeptide that comprises the sequence of amino acids set forth in SEQ ID NO: 3 or an active fragment thereof.

60. The isolated nucleic acid molecule of claim 57 or claim 58, wherein  
20 the sequence of nucleotides encodes a polypeptide that consists of the sequence of amino acids set forth in SEQ ID NO:3 or an active fragment thereof.

61. The isolated nucleic acid molecule of any of claims 57-60, wherein the active fragment is a pseudomature form.

62. The isolated nucleic acid molecule of any of claims 57-61, comprising  
25 the sequence of nucleotides set forth in SEQ ID NO:4 or a sequence of nucleotides that exhibits at least 95% sequence identity to SEQ ID NO:4 or the complement thereof, wherein the sequence of nucleotides encodes a polypeptide or active fragment that catalyzes the formation of isopimaradiene from geranylgeranyl diphosphate (GGPP).

63. The isolated nucleic acid molecule of claim 62, comprising a sequence  
30 of nucleotides that exhibits at least 96%, 97%, 98%, 99% or more sequence identity to SEQ ID NO:4.

- 92 -

64. The isolated nucleic acid molecule of claim 62 or claim 63, comprising the sequence of nucleotides set forth in SEQ ID NO:4, an active fragment thereof or the complement thereof.

65. The isolated nucleic acid molecule of claim 62 or claim 63, consisting  
5 of the sequence of nucleotides set forth in SEQ ID NO:4, an active fragment thereof or the complement thereof.

66. A isopimaradiene synthase (Iso) polypeptide encoded by the nucleic acid molecule of any of claims 57-65.

67. A vector, comprising the nucleic acid molecule of any of claims 57-65.

10 68. A host cell, comprising the vector of claim 67.

69. The host cell of claim 68 that is a prokaryotic or eukaryotic host cell.

70. The host cell of claim 68 or claim 69, wherein the host cell is selected from among a bacterial, fungal, plant, insect, amphibian or animal cell.

71. The host cell of any of claims 68-70 that is an *E. coli* cell.

15 72. The host cell of any of claims 68-70 that is a yeast cell.

73. The host cell of any of claims 68-72 that produces GGPP.

74. The host cell of any of claims 68-73 that is transformed with one or more sequences that result in production of GGPP.

75. A method of producing isopimaradiene, comprising

20 i) contacting (E,E,E)-geranylgeranyl diphosphate (GGPP) with a isopimaradiene synthase (Iso) polypeptide encoded by the nucleic acid molecule of any of claims 57-65 under conditions effective to produce isopimaradiene;

ii) optionally, isolating the isopimaradiene produced in step i).

76. The method of claim 75, wherein the step of contacting E,E,E)-  
25 geranylgeranyl diphosphate (GGPP) with a isopimaradiene synthase (Iso) polypeptide is effected *in vitro* or *in vivo*.

77. The method of claim 75 or claim 76, comprising isolating isopimaradiene.

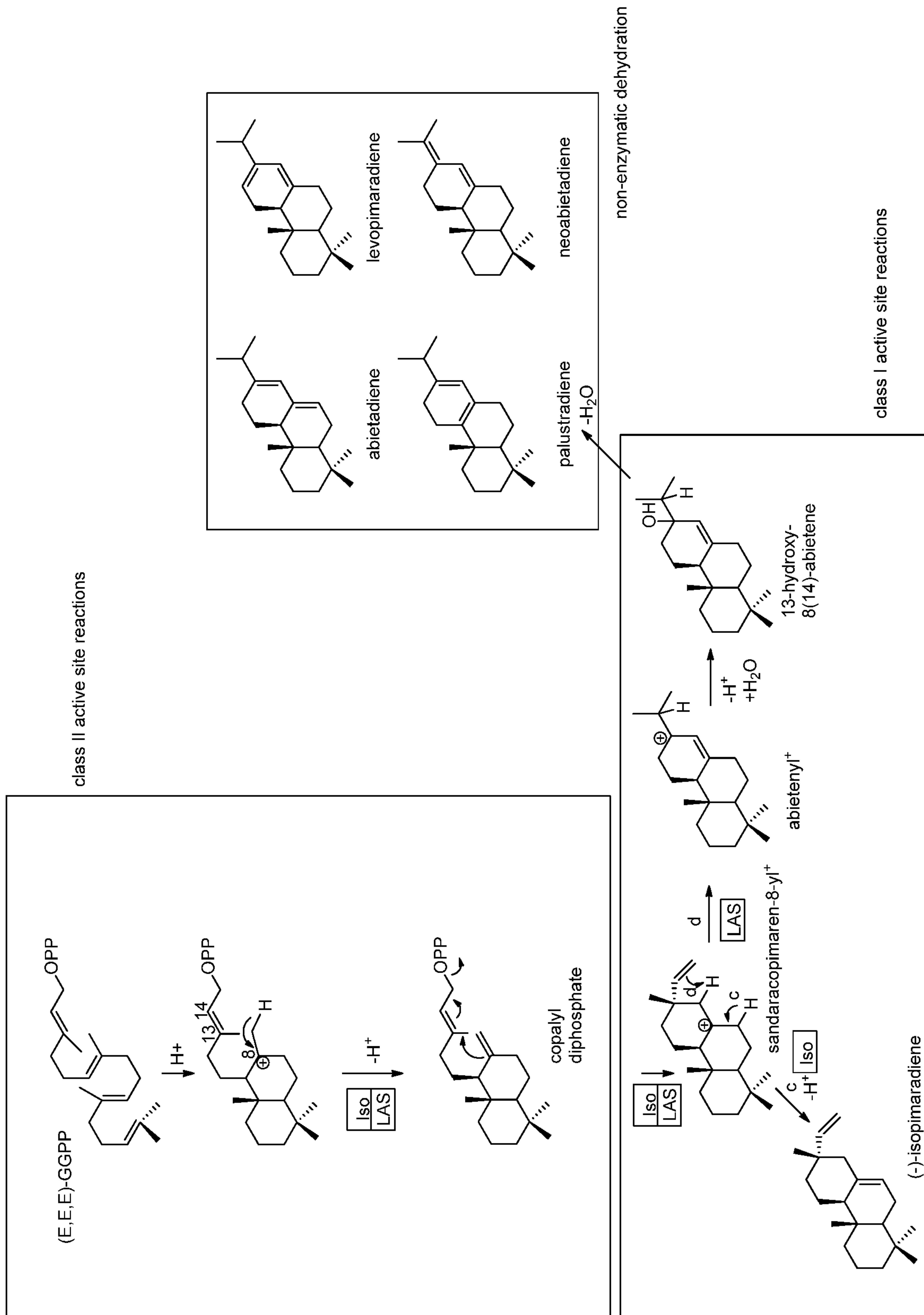


FIGURE 1A

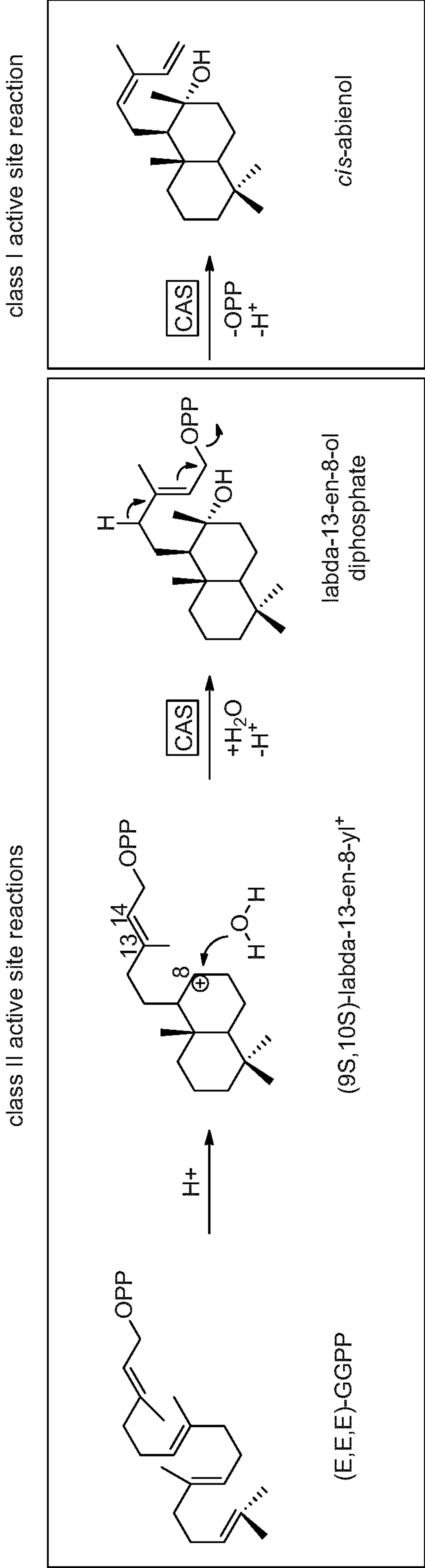


FIGURE 1B

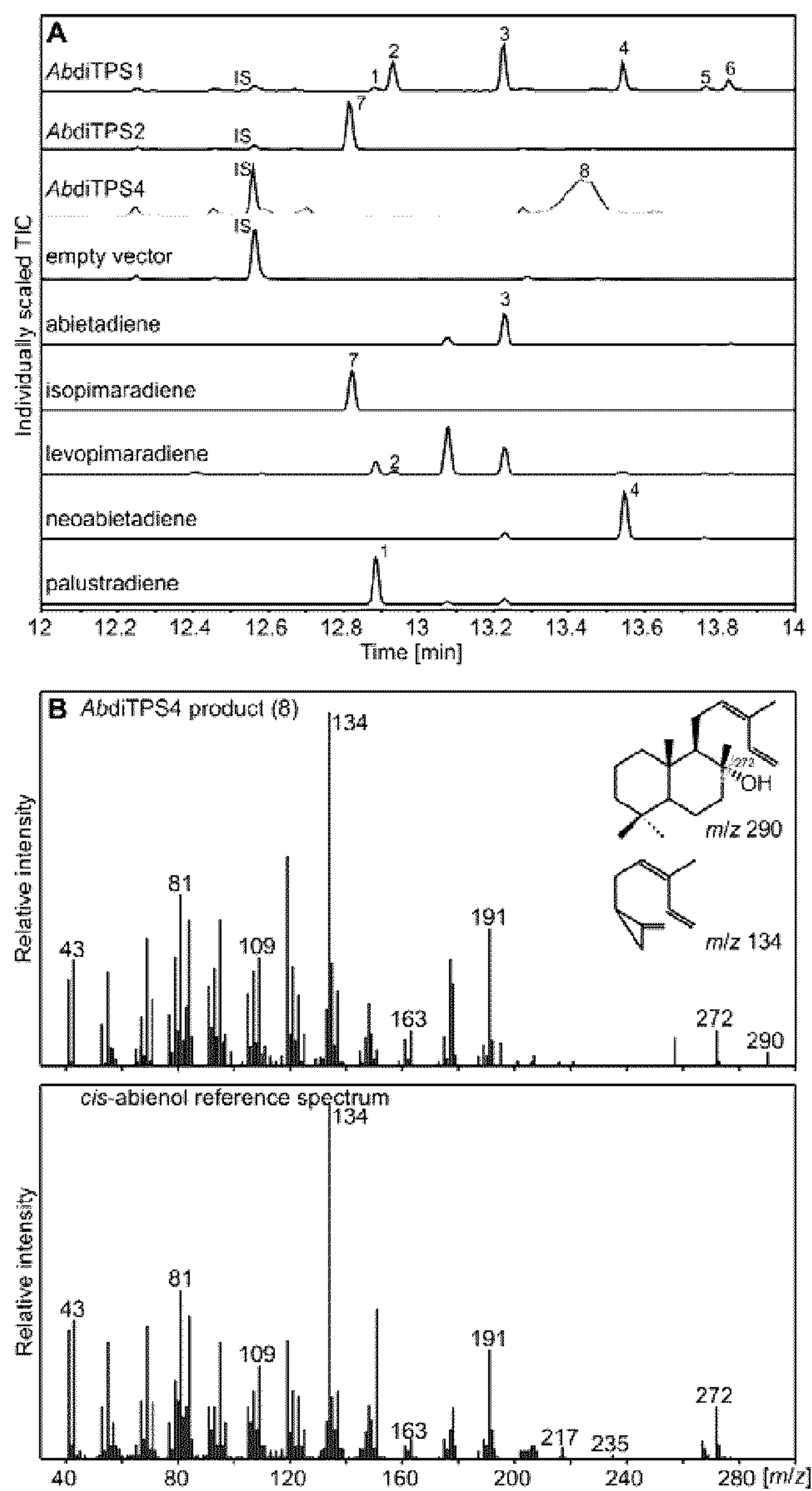


Figure 3

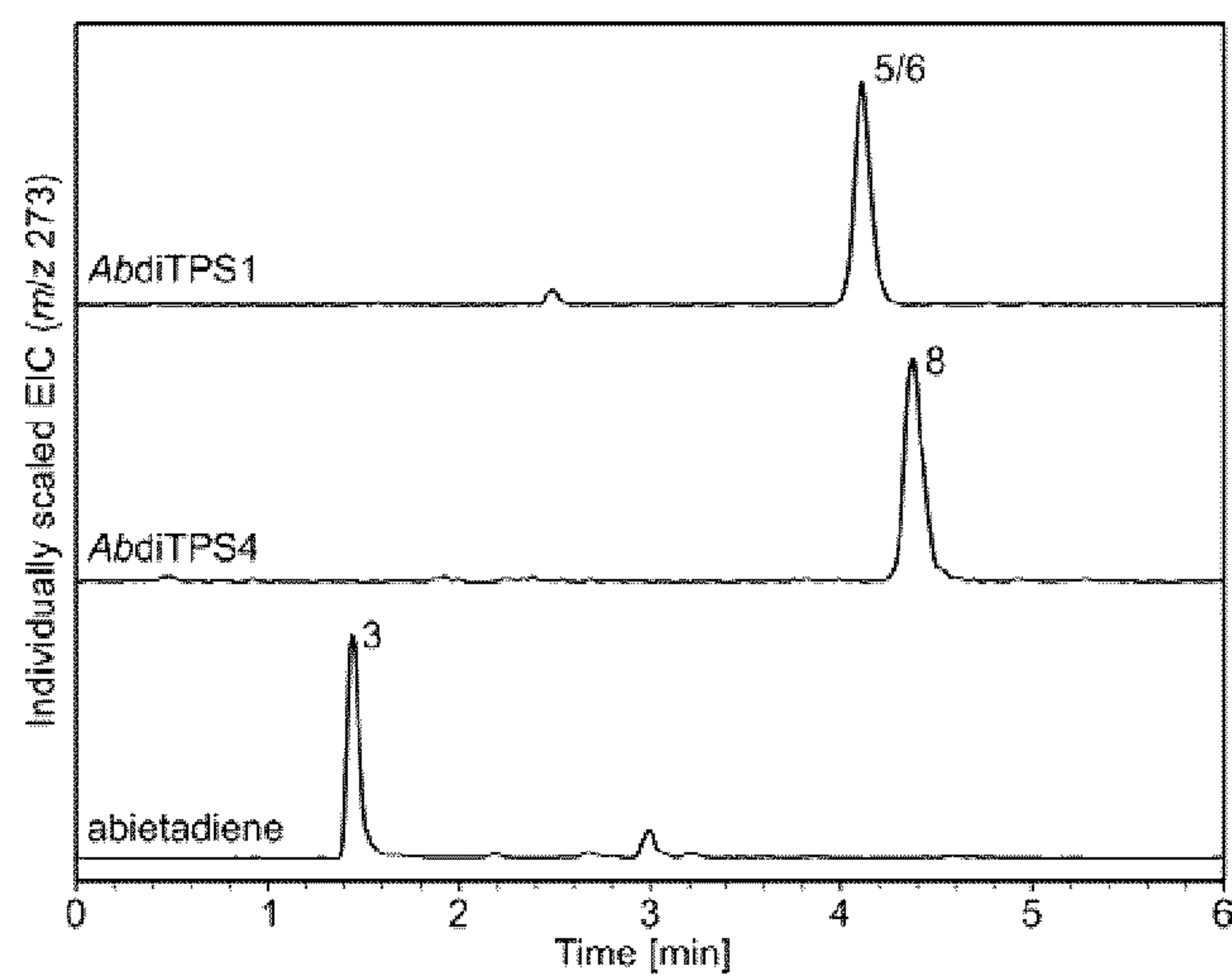
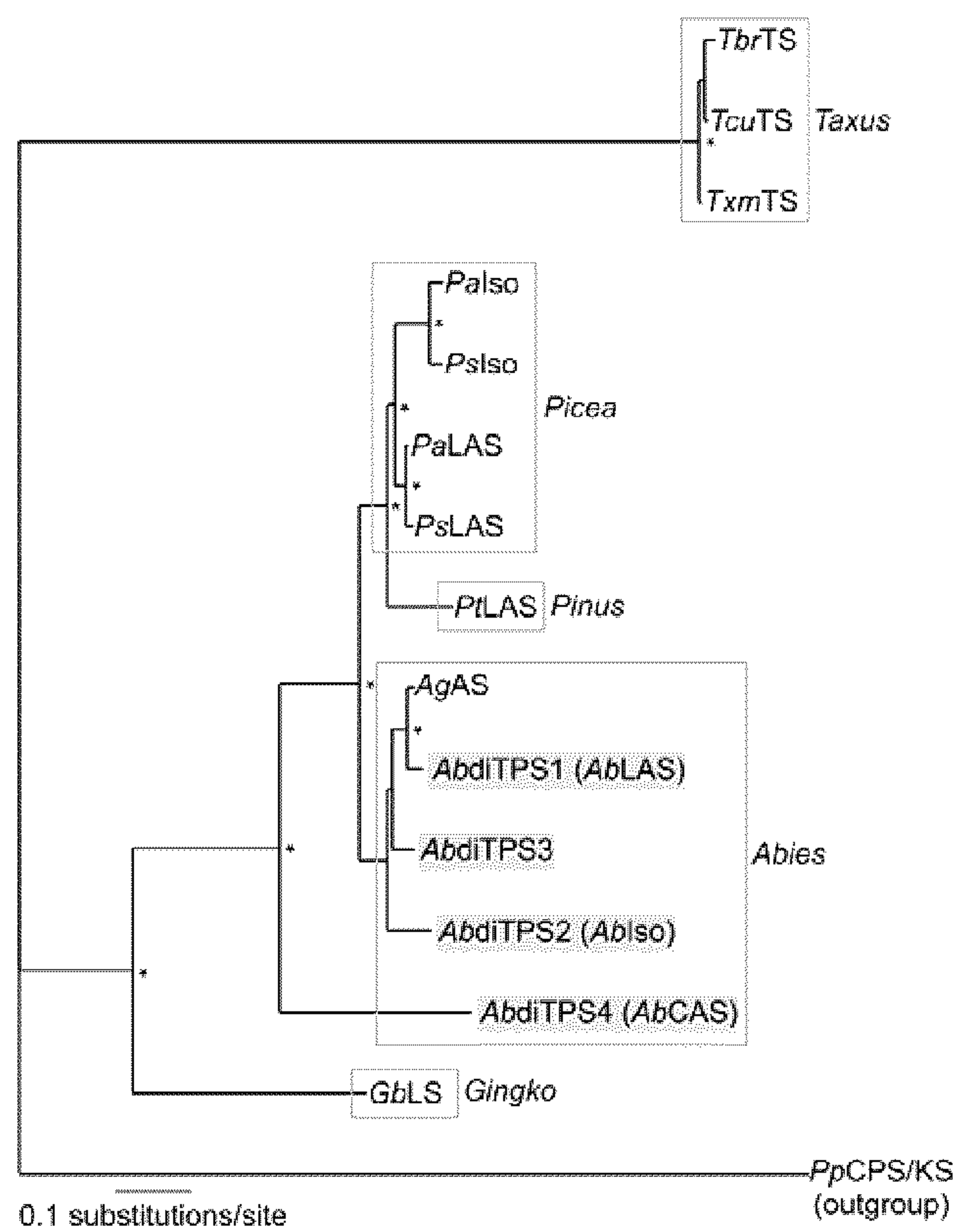


Figure 4



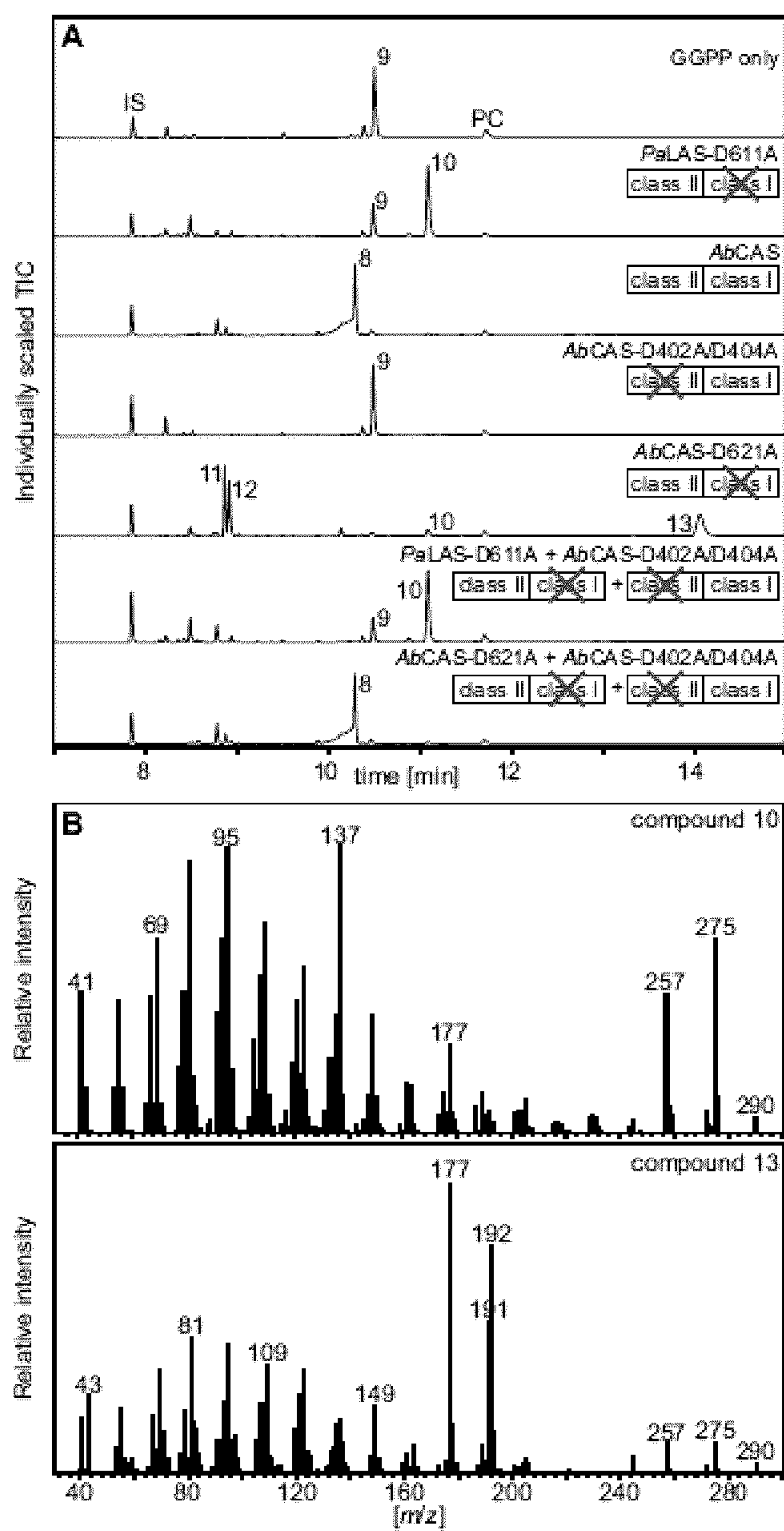
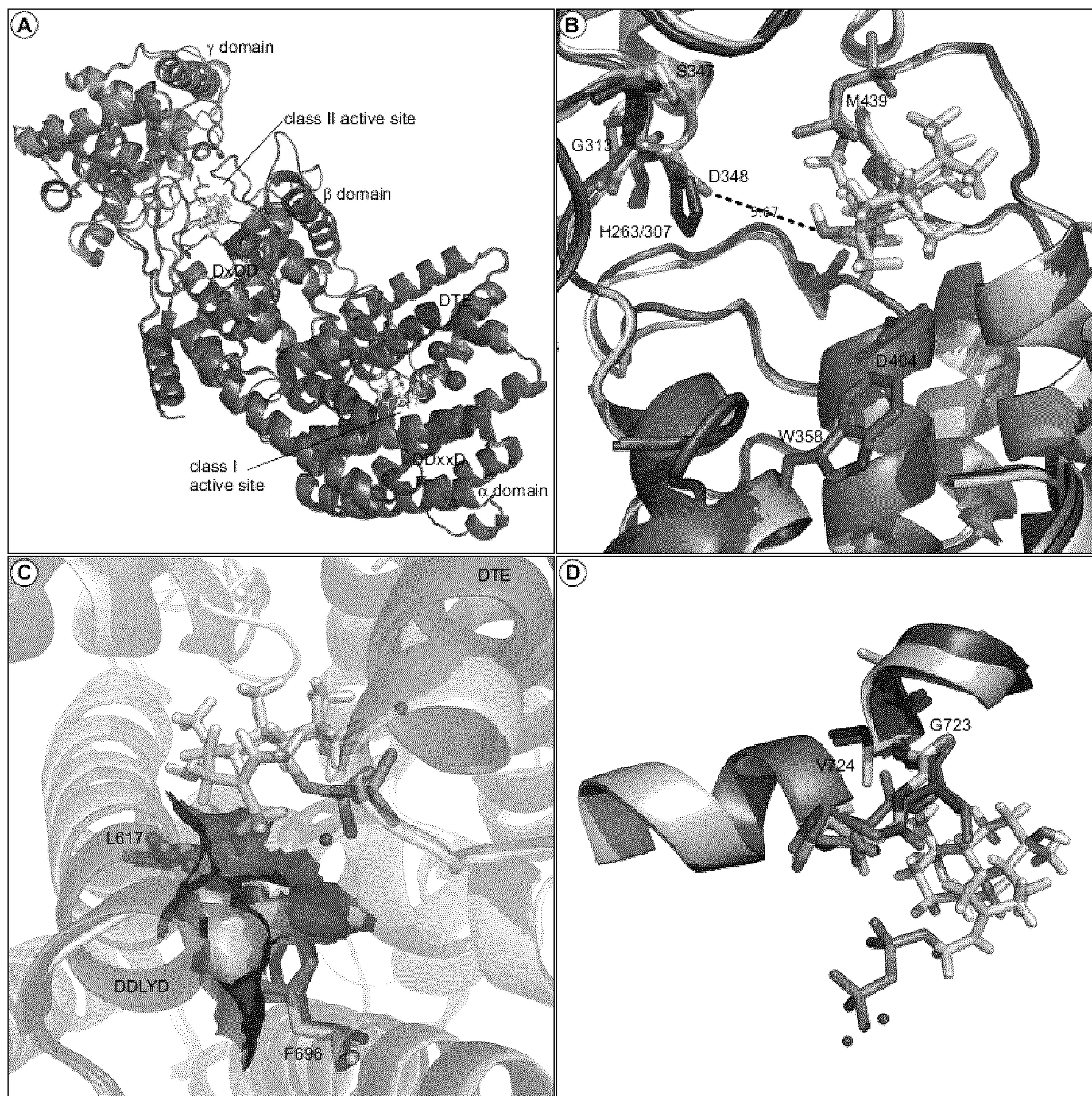


Figure 6



PstISO MALL-----SSSLSSQIPTGSHPLT-----HTQCIPHFSTTINAGISAQK--PRSFYLRWGK-----GSMKIIACVGE--TTS-----LPY 68  
 PstISO MALL-----SSSLSSRIPTGPHPLT-----MTQCIPHFSTTINAGISAQK--PRSFYLRWGKDSOPKNLGSNKIIACVGE--TTS-----LPY 75  
 GdLS MAGVLFANLPGSLQLSPKVP-----FRQSTNIIIPFHKRSSFQFNAQHCVRSHLRWN-----CVGIIHASAAETR-----PD 68  
 PstAS MALL-----SSSLSSQIHTGA-----ITQCIPHFHGSINAGTSAGK--RRSIIYLRWGK-----GPSKIYACAGQ-----57  
 PstLAS MALL-----SSSLSSQIPTGAHMLTLNAYANTQCIPHFSTTINAGTSAGK--RRSIIYLRWGK-----GSMKIIACVGE-----66  
 PstLAS MALL-----SSSLSSSHIPTGAHMLTLNAYANTQCIPHFSTTINAGTSAGK--RRSIIYLRWGK-----GSMKIIACVGE-----66  
 AgAS MAMP-----SSSLSQIPTAAHMLT-----ANAQSIPIHFSTTINAGSSASK--RRSIIYLRWGK-----GSMKIIACVGE--GATS-----VPY 70  
 AbdiTPS1 MAS-----MT-----GGQOM-----GR-----DPN-----15  
 AbdiTPS2 HMLT-----ANTQSIPIHFSTTINAGSSARK--RRSIIYLRWGK-----GSMKIIACVGE--ATS-----VPY 83  
 AbdiTPS3 NTQSIPIHFSTTINAGSSARK--RRSIIYLRWGK-----GSMKIIACVGE--ATS-----VPY 48  
 AbdiTPS4 MALL-----VYSIKSHIP-----ITIASAKMHYISNKGITAK--GRSRGIR-----LSPNKIYACAGGADATFPSQSLKLTALFPD 70

PstISO QSAEKTDSLSAPTILKREFPAGFKDHYIDSLTS--SHKVSAA--EE--KMETIISEKNIRSMGYCEINPSAYDTAWVARIPAVDGSSEHREFEEL 161  
 PstISO QSAEKTDSLSAPTILKREFPAGFKDHYIDSLTS--SHKVSAA--EE--KRIETIISOKNIRSMGYCEINPSAYDTAWVARIPAVDGSSEHREFEEL 168  
 GdLS QLPQEEERFYSR--LNADYHRAVKKDDFIDSLTSPNHSATSKSSVDETINX--IOTVKEOCMOSMGDCEINPSAYDTAWVARIPSDGSGAPQEQEEL 165  
 PstAS MALL-----DPSVPTILKREFPAGFKDHYIDSLTSPNHSATSKSSVDETINX--IOTVKEOCMOSMGDCEINPSAYDTAWVARIPAVDGSSEHREFEEL 144  
 PstLAS MALL-----DLSAPTILKREFPAGFKDHYIDSLTS--SHKVAAS--DE--KRIETIISEKNMRSMGYCEINPSAYDTAWVARIPAVDGSSEHREFEEL 153  
 PstLAS MALL-----DSVSAPTILKREFPAGFKDHYIDSLTS--SHKVAAS--DE--KRIETIISEKNMRSMGYCEINPSAYDTAWVARIPAVDGSSEHREFEEL 153  
 AgAS QSAEKNDSLSSTILKREFPAGFKDDIDSLTS--SHKVAAS--DE--KRIETIISEKNMRSMGYCEINPSAYDTAWVARIPAVDGSSEHREFEEL 163  
 AbdiTPS1 MALL-----SSVCKREFPAGFKDDIDSLTS--SHKVAAS--DE--KRIETIISEKNMRSMGYCEINPSAYDTAWVARIPAVDGSSEHREFEEL 99  
 AbdiTPS2 QSAEKNDSLSSTILKREFPAGFKDDIDSLTS--SHKVAAS--DE--KRIETIISEKNMRSMGYCEINPSAYDTAWVARIPAVDGSSEHREFEEL 146  
 AbdiTPS3 QSAEKNDSLSSTILKREFPAGFKDDIDSLTS--SHKVAAS--DE--KRIETIISEKNMRSMGYCEINPSAYDTAWVARIPAVDGSSEHREFEEL 141  
 AbdiTPS4 PFSKNGTFSNFTPNREFPAGFKDDIDSLTS--SHKVAAS--DE--KRIETIISEKNMRSMGYCEINPSAYDTAWVARIPAVDGSSEHREFEEL 163

PstISO ENILQNGQKDGSGWGEFYFLAYDRILATLACIITLTLRTSETIRKGIIEFFKTQAGKIEDEADSHRPSGFEIVFPAMLKSAKVLGLDLPYELPIIKQII 261  
 PstISO ENILQNGQKDGSGWGEFYFLAYDRILATLACIITLTLRTSETIRKGIIEFFKTQAGKIEDEADSHRPSGFEIVFPAMLKSAKVLGLDLPYELPIIKQII 268  
 GdLS QWLENNQDPGSGWGEFYFLAYDRILATLACIITLTLRTSETIRKGIIEFFKTQAGKIEDEADSHRPSGFEIVFPAMLKSAKVLGLDLPYELPIIKQII 265  
 PstAS ENILQNGQKDGSGWGEFYFLAYDRILATLACIITLTLRTSETIRKGIIEFFKTQAGKIEDEADSHRPSGFEIVFPAMLKSAKVLGLDLPYELPIIKQII 244  
 PstLAS ENILQNGQKDGSGWGEFYFLAYDRILATLACIITLTLRTSETIRKGIIEFFKTQAGKIEDEADSHRPSGFEIVFPAMLKSAKVLGLDLPYELPIIKQII 253  
 PstLAS ENILQNGQKDGSGWGEFYFLAYDRILATLACIITLTLRTSETIRKGIIEFFKTQAGKIEDEADSHRPSGFEIVFPAMLKSAKVLGLDLPYELPIIKQII 253  
 AgAS ENILQNGQKDGSGWGEFYFLAYDRILATLACIITLTLRTSETIRKGIIEFFKTQAGKIEDEADSHRPSGFEIVFPAMLKSAKVLGLDLPYELPIIKQII 263  
 AbdiTPS1 ENILQNGQKDGSGWGEFYFLAYDRILATLACIITLTLRTSETIRKGIIEFFKTQAGKIEDEADSHRPSGFEIVFPAMLKSAKVLGLDLPYELPIIKQII 196  
 AbdiTPS2 ENILQNGQKDGSGWGEFYFLAYDRILATLACIITLTLRTSETIRKGIIEFFKTQAGKIEDEADSHRPSGFEIVFPAMLKSAKVLGLDLPYELPIIKQII 246  
 AbdiTPS3 ENILQNGQKDGSGWGEFYFLAYDRILATLACIITLTLRTSETIRKGIIEFFKTQAGKIEDEADSHRPSGFEIVFPAMLKSAKVLGLDLPYELPIIKQII 241  
 AbdiTPS4 ENILQNGQKDGSGWGEFYFLAYDRILATLACIITLTLRTSETIRKGIIEFFKTQAGKIEDEADSHRPSGFEIVFPAMLKSAKVLGLDLPYELPIIKQII 263

PstISO EKREAKLERLPTNIXYALPTITLLYSLEGQEIYDQWKIKLQSKDGSFLTSASTAAVFRITGKKKEFLNFKKFGNHHPCHPDLPFERLWADTV 361  
 PstISO EKREAKLERLPTNIXYALPTITLLYSLEGQEIYDQWKIKLQSKDGSFLTSASTAAVFRITGKKKEFLNFKKFGNHHPCHPDLPFERLWADTV 368  
 GdLS QKRGKXKOKIPLNVHNNQALLYSLEGQEIYDQWKIKLQSKDGSFLTSASTAAVFRITGKKKEFLNFKKFGNHHPCHPDLPFERLWADTV 365  
 PstAS EKREAKLERLPTNIXYALPTITLLYSLEGQEIYDQWKIKLQSKDGSFLTSASTAAVFRITGKKKEFLNFKKFGNHHPCHPDLPFERLWADTV 344  
 PstLAS EKREAKLERLPTNIXYALPTITLLYSLEGQEIYDQWKIKLQSKDGSFLTSASTAAVFRITGKKKEFLNFKKFGNHHPCHPDLPFERLWADTV 353  
 PstLAS EKREAKLERLPTNIXYALPTITLLYSLEGQEIYDQWKIKLQSKDGSFLTSASTAAVFRITGKKKEFLNFKKFGNHHPCHPDLPFERLWADTV 353  
 AgAS EKREAKLERLPTNIXYALPTITLLYSLEGQEIYDQWKIKLQSKDGSFLTSASTAAVFRITGKKKEFLNFKKFGNHHPCHPDLPFERLWADTV 363  
 AbdiTPS1 EKREAKLERLPTNIXYALPTITLLYSLEGQEIYDQWKIKLQSKDGSFLTSASTAAVFRITGKKKEFLNFKKFGNHHPCHPDLPFERLWADTV 298  
 AbdiTPS2 EKREAKLERLPTNIXYALPTITLLYSLEGQEIYDQWKIKLQSKDGSFLTSASTAAVFRITGKKKEFLNFKKFGNHHPCHPDLPFERLWADTV 346  
 AbdiTPS3 EKREAKLERLPTNIXYALPTITLLYSLEGQEIYDQWKIKLQSKDGSFLTSASTAAVFRITGKKKEFLNFKKFGNHHPCHPDLPFERLWADTV 341  
 AbdiTPS4 EKREAKLERLPTNIXYALPTITLLYSLEGQEIYDQWKIKLQSKDGSFLTSASTAAVFRITGKKKEFLNFKKFGNHHPCHPDLPFERLWADTV 363

PstISO ERLGIDRHKEEKIDALDYVYSHND--ERGIGWARENPVPIIDGTAMGLRILRLHGYNYSDDVKTFRDENGEEFCGLGQTORGVDTMLNVNRCSSHVA 460  
 PstISO ERLGIDRHKEEKIDALDYVYSHND--ERGIGWARENPVPIIDGTAMGLRILRLHGYNYSDDVKTFRDENGEEFCGLGQTORGVDTMLNVNRCSSHVA 467  
 GdLS ERLGIDRYKKKEIKESLDYVYRYND--ERGIGWARENPVPIIDGTAMGLRILRLHGYNYSDDVKTFRDENGEEFCGLGQTORGVDTMLNVNRCSSHVA 465  
 PstAS ERLGIDRHKEEKIDALDYVYSHND--ERGIGWARENPVPIIDGTAMGLRILRLHGYNYSDDVKTFRDENGEEFCGLGQTORGVDTMLNVNRCSSHVA 443  
 PstLAS ERLGIDRHKEEKIDALDYVYSHND--ERGIGWARENPVPIIDGTAMGLRILRLHGYNYSDDVKTFRDENGEEFCGLGQTORGVDTMLNVNRCSSHVA 452  
 PstLAS ERLGIDRHKEEKIDALDYVYSHND--ERGIGWARENPVPIIDGTAMGLRILRLHGYNYSDDVKTFRDENGEEFCGLGQTORGVDTMLNVNRCSSHVA 452  
 AgAS ERLGIDRHKEEKIDALDYVYSHND--ERGIGWARENPVPIIDGTAMGLRILRLHGYNYSDDVKTFRDENGEEFCGLGQTORGVDTMLNVNRCSSHVA 462  
 AbdiTPS1 ERLGIDRHKEEKIDALDYVYSHND--ERGIGWARENPVPIIDGTAMGLRILRLHGYNYSDDVKTFRDENGEEFCGLGQTORGVDTMLNVNRCSSHVA 397  
 AbdiTPS2 ERLGIDRHKEEKIDALDYVYSHND--ERGIGWARENPVPIIDGTAMGLRILRLHGYNYSDDVKTFRDENGEEFCGLGQTORGVDTMLNVNRCSSHVA 445  
 AbdiTPS3 ERLGIDRHKEEKIDALDYVYSHND--ERGIGWARENPVPIIDGTAMGLRILRLHGYNYSDDVKTFRDENGEEFCGLGQTORGVDTMLNVNRCSSHVA 440  
 AbdiTPS4 ERLGIDRHKEEKIDALDYVYSHND--ERGIGWARENPVPIIDGTAMGLRILRLHGYNYSDDVKTFRDENGEEFCGLGQTORGVDTMLNVNRCSSHVA 462

PstISO ETINEEAKLCERYRNALEDVGFQKWAALKNIRGEVEYALKPNHRSMPLEARSYEHYGPNDVWLGKTMVMMPYISNEKYLEAKLDFNHVSLHQ 560  
 PstISO ETINEEAKLCERYRNALEDVGFQKWAALKNIRGEVEYALKPNHRSMPLEARSYEHYGPNDVWLGKTMVMMPYISNEKYLEAKLDFNHVSLHQ 567  
 GdLS ETINEEAKLCERYRNALEDVGFQKWAALKNIRGEVEYALKPNHRSMPLEARSYEHYGPNDVWLGKTMVMMPYISNEKYLEAKLDFNHVSLHQ 565  
 PstAS ETINEEAKLCERYRNALEDVGFQKWAALKNIRGEVEYALKPNHRSMPLEARSYEHYGPNDVWLGKTMVMMPYISNEKYLEAKLDFNHVSLHQ 543  
 PstLAS ETINEEAKLCERYRNALEDVGFQKWAALKNIRGEVEYALKPNHRSMPLEARSYEHYGPNDVWLGKTMVMMPYISNEKYLEAKLDFNHVSLHQ 552  
 PstLAS ETINEEAKLCERYRNALEDVGFQKWAALKNIRGEVEYALKPNHRSMPLEARSYEHYGPNDVWLGKTMVMMPYISNEKYLEAKLDFNHVSLHQ 552  
 AgAS ETINEEAKLCERYRNALEDVGFQKWAALKNIRGEVEYALKPNHRSMPLEARSYEHYGPNDVWLGKTMVMMPYISNEKYLEAKLDFNHVSLHQ 562  
 AbdiTPS1 ETINEEAKLCERYRNALEDVGFQKWAALKNIRGEVEYALKPNHRSMPLEARSYEHYGPNDVWLGKTMVMMPYISNEKYLEAKLDFNHVSLHQ 497  
 AbdiTPS2 ETINEEAKLCERYRNALEDVGFQKWAALKNIRGEVEYALKPNHRSMPLEARSYEHYGPNDVWLGKTMVMMPYISNEKYLEAKLDFNHVSLHQ 545  
 AbdiTPS3 ETINEEAKLCERYRNALEDVGFQKWAALKNIRGEVEYALKPNHRSMPLEARSYEHYGPNDVWLGKTMVMMPYISNEKYLEAKLDFNHVSLHQ 540  
 AbdiTPS4 ETINEEAKLCERYRNALEDVGFQKWAALKNIRGEVEYALKPNHRSMPLEARSYEHYGPNDVWLGKTMVMMPYISNEKYLEAKLDFNHVSLHQ 562

PstISO KELRDRLRRWKS3SGFTELKTRERVTETSPASFIIEPERATORAVTNTNFTYVLDLYCAHSTLDLKLPSDSYKKWDLVYDRM-PQDMKICFVG 659  
 PstISO KELRDRLRRWKS3SGFTELKTRERVTETSPASFIIEPERATORAVTNTNFTYVLDLYCAHSTLDLKLPSDSYKKWDLVYDRM-PQDMKICFVG 666  
 GdLS KETQIVSWRESQFNDLTTRQPVEMTSVAVSMIEPERAACHIAKTCGLAYLDLYCAHSTLDLKLPSDSYKKWDLVYDRM-PQDMKICFVG 665  
 PstAS QELQDIRWNS3SGFTELKTRERVTETSPASFIIEPERATORAVTNTNFTYVLDLYCAHSTLDLKLPSDSYKKWDLVYDRM-PQDMKICFVG 642  
 PstLAS KELRDRLRRWKS3SGFTELKTRERVTETSPASFIIEPERATORAVTNTNFTYVLDLYCAHSTLDLKLPSDSYKKWDLVYDRM-PQDMKICFVG 651  
 PstLAS KELRDRLRRWKS3SGFTELKTRERVTETSPASFIIEPERATORAVTNTNFTYVLDLYCAHSTLDLKLPSDSYKKWDLVYDRM-PQDMKICFVG 651  
 AgAS TELQDLRRWKS3SGFTELKTRERVTETSPASFIIEPERATORAVTNTNFTYVLDLYCAHSTLDLKLPSDSYKKWDLVYDRM-PQDMKICFVG 661  
 AbdiTPS1 TELQDLRRWKS3SGFTELKTRERVTETSPASFIIEPERATORAVTNTNFTYVLDLYCAHSTLDLKLPSDSYKKWDLVYDRM-PQDMKICFVG 596  
 AbdiTPS2 TELQDLRRWKS3SGFTELKTRERVTETSPASFIIEPERATORAVTNTNFTYVLDLYCAHSTLDLKLPSDSYKKWDLVYDRM-PQDMKICFVG 644  
 AbdiTPS3 TELQDLRRWKS3SGFTELKTRERVTETSPASFIIEPERATORAVTNTNFTYVLDLYCAHSTLDLKLPSDSYKKWDLVYDRM-PQDMKICFVG 639  
 AbdiTPS4 EIRRELVRWKS3SGFTELKTRERVTETSPASFIIEPERATORAVTNTNFTYVLDLYCAHSTLDLKLPSDSYKKWDLVYDRM-PQDMKICFVG 661

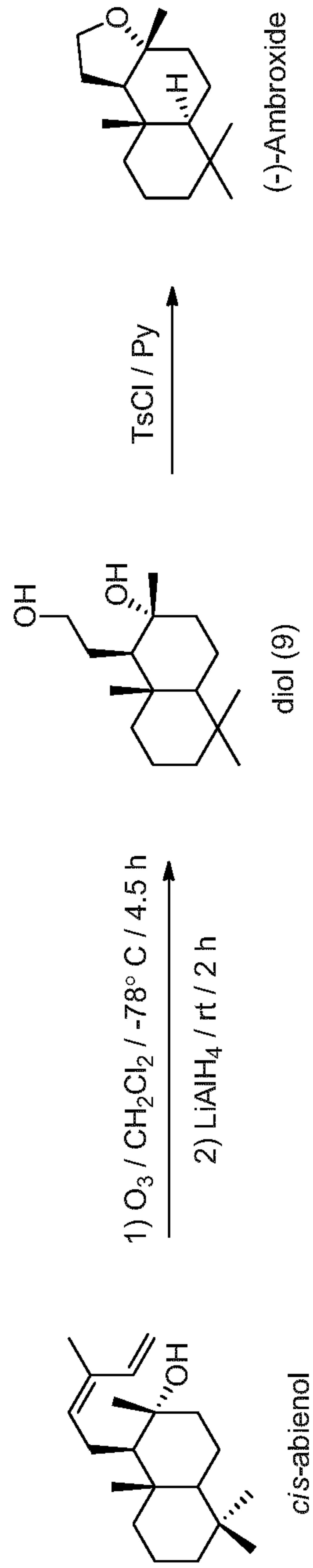
PstISO LYNVNEIAVEGRKRQGRDVLGYIRNVLEILAAHKEAEWSAARYVSFDEYIENASVSIALGTVVLSVLTGEILTDDVLSKIGRCSRLOLMGTG 759  
 PstISO LYNVNEIAVEGRKRQGRDVLGYIRNVLEILAAHKEAEWSAARYVSFDEYIENASVSIALGTVVLSVLTGEILTDDVLSKIGRCSRLOLMGTG 766  
 GdLS LYNVNGFGKDLKEDGRDVLGYIRNVLEILAAHKEAEWSAARYVSFDEYIENASVSIALGTVVLSVLTGEILTDDVLSKIGRCSRLOLMGTG 765  
 PstAS FYNVNEIAVEGRKRQGRDVLGYIRNVLEILAAHKEAEWSAARYVSFDEYIENASVSIALGTVVLSVLTGEILTDDVLSKIGRCSRLOLMGTG 742  
 PstLAS FYNVNEIAVEGRKRQGRDVLGYIRNVLEILAAHKEAEWSAARYVSFDEYIENASVSIALGTVVLSVLTGEILTDDVLSKIGRCSRLOLMGTG 751  
 PstLAS FYNVNEIAVEGRKRQGRDVLGYIRNVLEILAAHKEAEWSAARYVSFDEYIENASVSIALGTVVLSVLTGEILTDDVLSKIGRCSRLOLMGTG 751  
 AgAS FYNVNEIAVEGRKRQGRDVLGYIRNVLEILAAHKEAEWSAARYVSFDEYIENASVSIALGTVVLSVLTGEILTDDVLSKIGRCSRLOLMGTG 761  
 AbdiTPS1 FYNVNEIAVEGRKRQGRDVLGYIRNVLEILAAHKEAEWSAARYVSFDEYIENASVSIALGTVVLSVLTGEILTDDVLSKIGRCSRLOLMGTG 696  
 AbdiTPS2 FYNVNEIAVEGRKRQGRDVLGYIRNVLEILAAHKEAEWSAARYVSFDEYIENASVSIALGTVVLSVLTGEILTDDVLSKIGRCSRLOLMGTG 744  
 AbdiTPS3 LYNVNEIAVEGRKRQGRDVLGYIRNVLEILAAHKEAEWSAARYVSFDEYIENASVSIALGTVVLSVLTGEILTDDVLSKIGRCSRLOLMGTG 724  
 AbdiTPS4 LYNVNEIAVEGRKRQGRDVLGYIRNVLEILAAHKEAEWSAARYVSFDEYIENASVSIALGTVVLSVLTGEILTDDVLSKIGRCSRLOLMGTG 758

PstISO RLINDTKTYEAGGQGEVAAYOCYMKDHEISEEALKHNYTYMENALDELNREFVN-NRDVDSCRRLVETARIMOLYMEGGDLTSLHETEKHEV 858  
 PstISO RLINDTKTYEAGGQGEVAAYOCYMKDHEISEEALKHNYTYMENALDELNREFVN-NRDVDSCRRLVETARIMOLYMEGGDLTSLHETEKHEV 865  
 GdLS RLINDTKTYEAGGQGEVAAYOCYMKDHEISEEALKHNYTYMENALDELNREFVN-NRDVDSCRRLVETARIMOLYMEGGDLTSLHETEKHEV 864  
 PstAS RLINDTKTYEAGGQGEVAAYOCYMKDHEISEEALKHNYTYMENALDELNREFVN-NRDVDSCRRLVETARIMOLYMEGGDLTSLHETEKHEV 841  
 PstLAS RLINDTKTYEAGGQGEVAAYOCYMKDHEISEEALKHNYTYMENALDELNREFVN-NRDVDSCRRLVETARIMOLYMEGGDLTSLHETEKHEV 850  
 PstLAS RLINDTKTYEAGGQGEVAAYOCYMKDHEISEEALKHNYTYMENALDELNREFVN-NRDVDSCRRLVETARIMOLYMEGGDLTSLHETEKHEV 850  
 AgAS RLINDTKTYEAGGQGEVAAYOCYMKDHEISEEALKHNYTYMENALDELNREFVN-NRDVDSCRRLVETARIMOLYMEGGDLTSLHETEKHEV 859  
 AbdiTPS1 RLINDTKTYEAGGQGEVAAYOCYMKDHEISEEALKHNYTYMENALDELNREFVN-NRDVDSCRRLVETARIMOLYMEGGDLTSLHETEKHEV 794  
 AbdiTPS2 RLINDTKTYEAGGQGEVAAYOCYMKDHEISEEALKHNYTYMENALDELNREFVN-NRDVDSCRRLVETARIMOLYMEGGDLTSLHETEKHEV 843  
 AbdiTPS3 RLINDTKTYEAGGQGEVAAYOCYMKDHEISEEALKHNYTYMENALDELNREFVN-NRDVDSCRRLVETARIMOLYMEGGDLTSLHETEKHEV 843  
 AbdiTPS4 RLINDTKTYEAGGQGEVAAYOCYMKDHEISEEALKHNYTYMENALDELNREFVN-NRDVDSCRRLVETARIMOLYMEGGDLTSLHETEKHEV 857

PstISO KNCLEQOVA 867  
 PstISO KNCLEQOVA 874  
 GdLS SRTLEQOVA 873  
 PstAS KNCLEQOVA 850  
 PstLAS KNCLEQOVA 859  
 PstLAS KNCLEQOVA 859  
 AgAS KNCLEQOVA 868  
 AbdiTPS1 KNCLEQOVA 803  
 AbdiTPS2 KNCLEQOVA 852  
 AbdiTPS3 KNCLEQOVA 866  
 AbdiTPS4 KNCLEQOVA 866

Figure 7

## A. Scheme I



## B. Scheme II

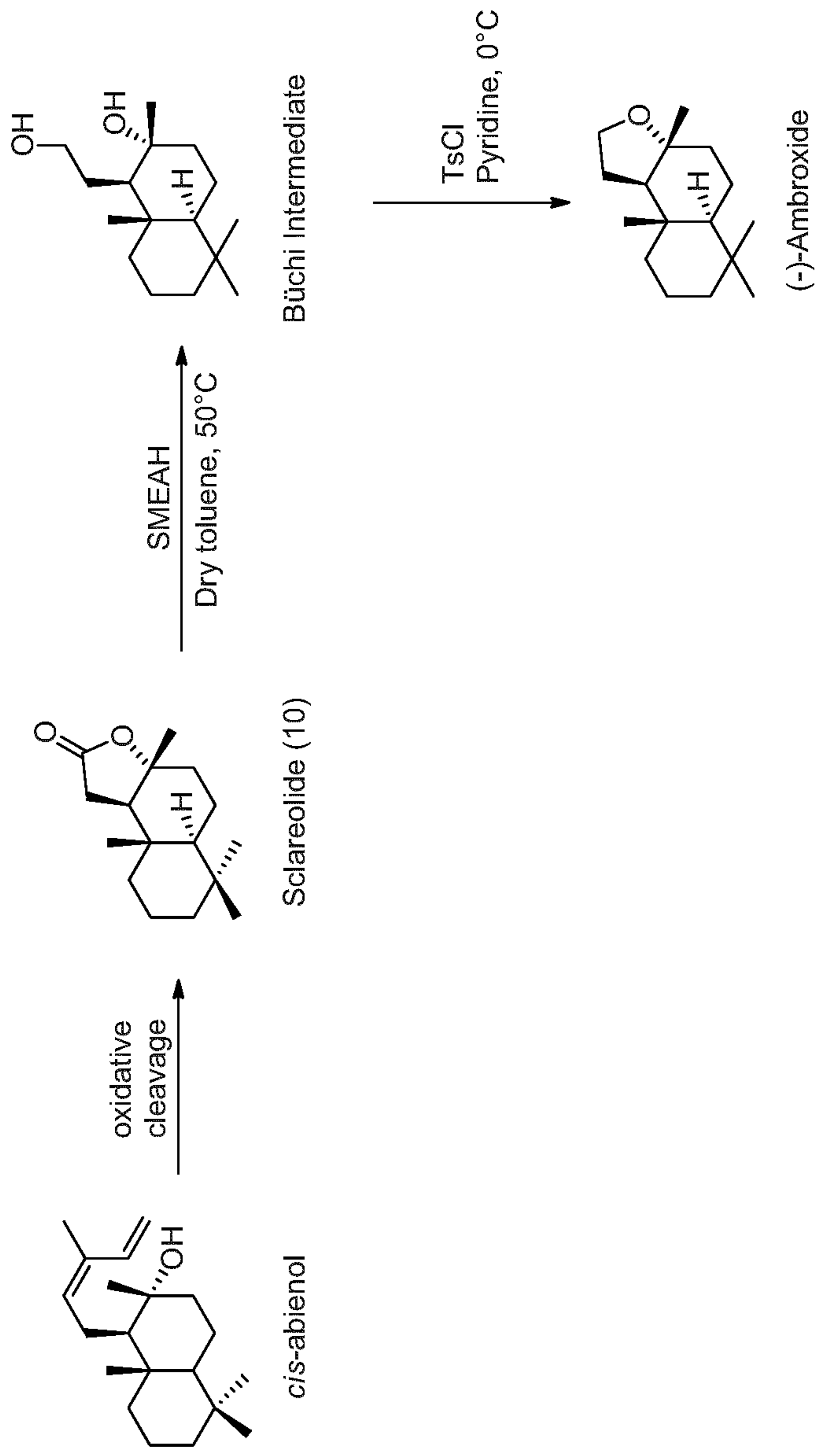


FIGURE 8

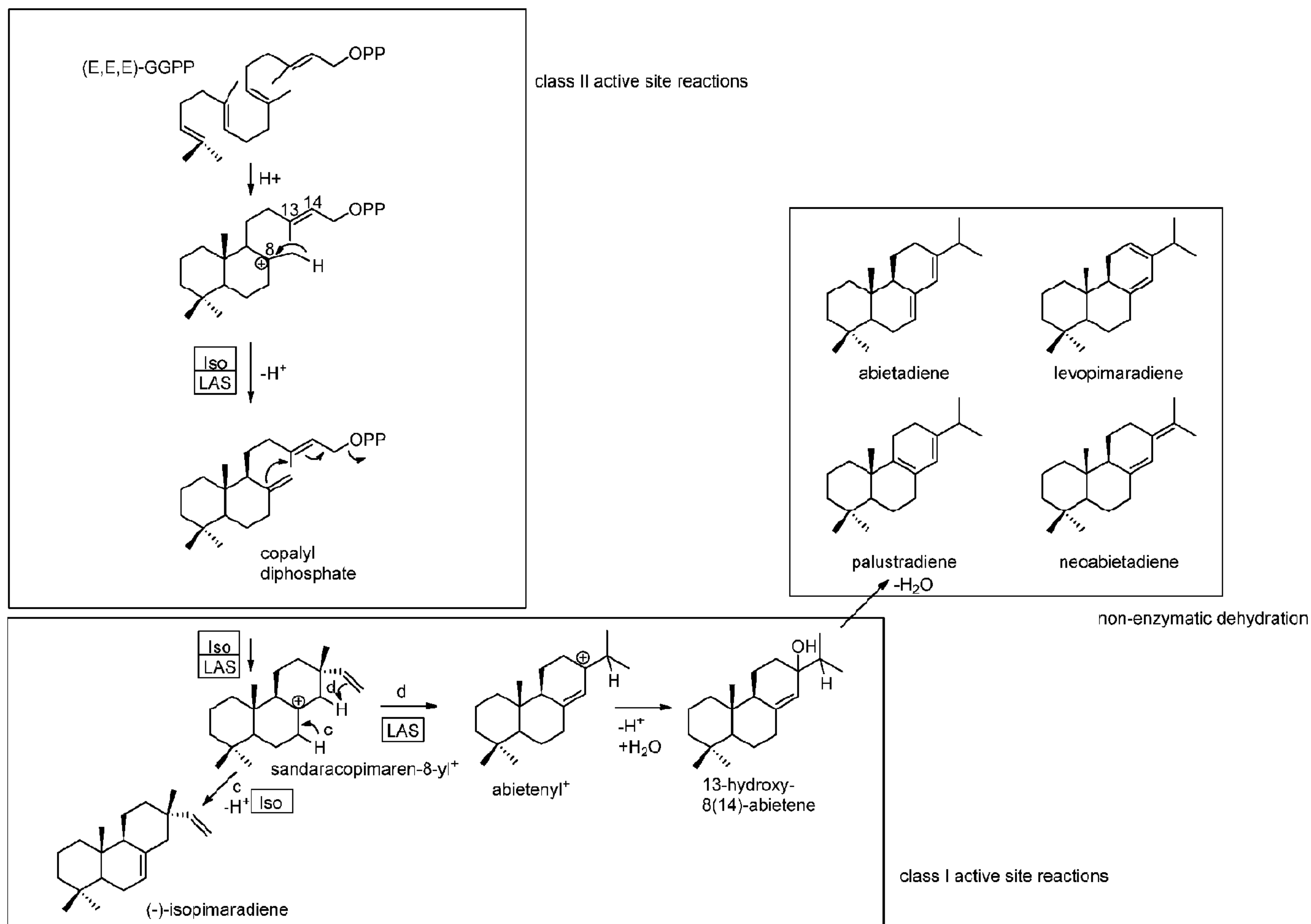


FIGURE 1A

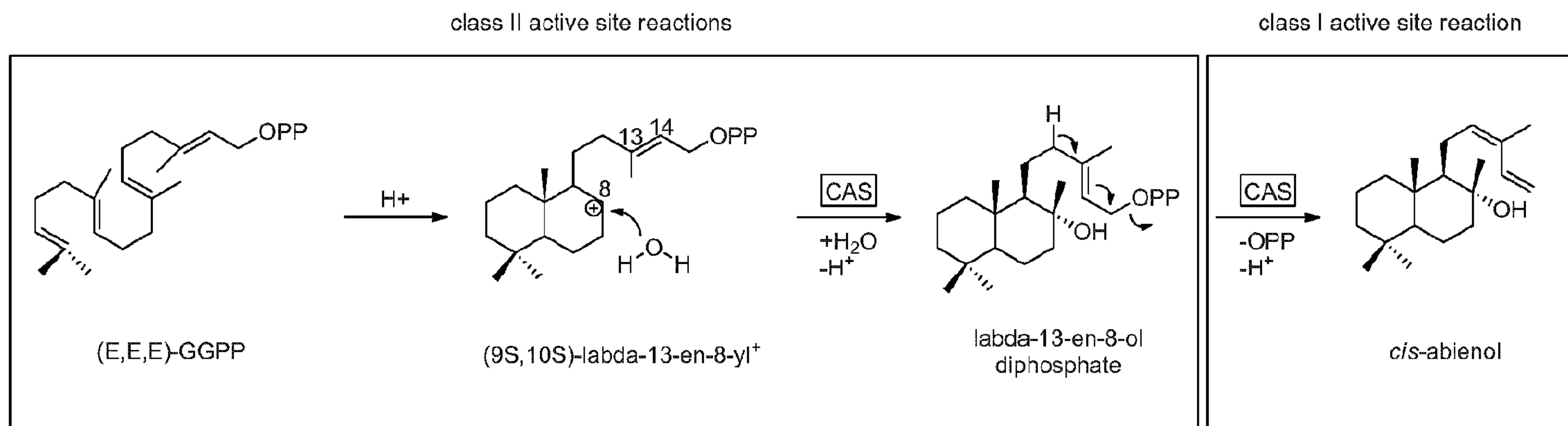


FIGURE 1B