Molecular cloning and functional characterization of a monoterpene synthase isolated from the aromatic wild shrub *Thymus albicans*

Alexandra Filipe[a], João CR Cardoso[a], Graça Miguel[b], Liliana Anjos[a], Helena Trindade[c], Ana Cristina Figueiredo[c], José Barroso[c], Deborah M. Power[a], Natália T. Marques[d*]

[a] Centro de Ciências do Mar, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal.

[b] Centre for Mediterranean Bioresources and Food (MeditBio), Faculdade de Ciências e Tecnologia, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal.

[c] Centro de Estudos do Ambiente e do Mar Lisboa, Faculdade de Ciências, Universidade de Lisboa, CBV, DBV, 1749-016 Lisboa, Portugal.

[d] Center of Electronics, Optoelectronics and Telecommunications, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal.

* Corresponding author: Natália Tomás Marques.

Address: Faculdade de Ciências e Tecnologia, Edif. 8, Universidade do Algarve, Campus de Gambelas, 8005-139 Faro, Portugal

e-mail: nmarques@ualg.pt

Phone: +351 289800900, ext. 7164; Fax: +351 289818419

Alexandra Filipe (xanafilippe@gmail.com), João C. R. Cardoso (jccardo@ualg.pt), Graça Miguel (mgmiguel@ualg.pt), Liliana Anjos (lanjos@ualg.pt), Helena Trindade (htrindade@fc.ul.pt), A. Cristina Figueiredo (acsf@fc.ul.pt), José G. Barroso (jmbarroso@fc.ul.pt), D. M. Power (dpower@ualg.pt), Natália Marques (nmarques@ualg.pt)

Abstract

The essential oil of *Thymus albicans* Hoffmanns. & Link, a native shrub from the Iberian Peninsula, is mainly composed of monoterpenes. In this study, a 1,8-cineole synthase was isolated from the 1,8-cineole chemotype. A partial sequence that lacked the complete plastid transit peptide but contained an extended C-terminal when compared to other related terpene synthases was generated by PCR and Rapid Amplification of cDNA Ends (RACE). The predicted mature polypeptide was 593 amino acids in length and shared 78% and 77% sequence similarity with the homologue 1,8-cineole synthase from *Rosmarinus officinalis* and *Salvia officinalis*, respectively. The putative protein possessed the characteristic conserved motifs of plant monoterpene synthases including the RRx_8W and DDxxD motifs and phylogenetic analysis indicated that the amplified 1,8-cineole synthase bears greater sequence similarity with other 1,8-cineole synthases from Lamiaceae family relative to the terpene synthases from the genus *Thymus*. Functional expression of the recombinant protein in *Escherichia coli* revealed that in the presence of geranyl diphosphate (GPP) 1,8-cineole was the major product but that its production was too low for robust quantification. Other minor conversion products included α-pinene, β-pinene, sabinene and β-myrcene suggesting the isolated 1,8-cineole synthase may be a multi-product enzyme. To our knowledge, this is the first report of a functionally characterized monoterpene synthase from *Thymus albicans*.

Keywords: *Thymus albicans*, heterologous expression, 1,8-cineole, terpene synthase, geranyl diphosphate.

Abbreviations

DAB	3,3´-Diaminobenzidine
DTT	Dithiothreitol
cDNA	Complementary DNA
EO	Essential oil
FPP	Farnesyl diphosphate
GC	Gas Chromatography

GC-MS	Gas Chromatography coupled with Mass Spectrometry
GGPP	Geranyl geranyldiphosphate
GPP	Geranyl diphosphate
IPTG	Isopropyl- β-D-1-thiogalactopyranoside
LB	Luria-Bertani
mTPS	Monoterpene synthase
ORF	Open reading frame
PCR	Polymerase chain reaction
RACE	Rapid amplification of cDNA ends
ROS	Reactive oxygen species
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPME	Solid phase micro extraction
TBS	Tris-buffered saline
TPS	Terpene synthase

Introduction

Essential oils (EO) are complex mixtures of volatile compounds, mostly mono- and sesquiterpenes, either hydrocarbons or oxygen-containing (Miguel, 2010). EOs are of high economic value being used as additives in the food, cosmetics and pharmaceutical industries. In the plants of the Lamiaceae family, the compounds isolated in essential oils are produced by specialized secretory cells, the glandular trichomes occurring in the leaves, stems and calyx (Gershenzon et al., 1992). The production of essential oils is mediated by a group of specialized enzymes of the terpene synthase (TPS) family that are responsible for catalyzing the synthesis of monoterpenes (C_{10}), sesquiterpenes (C_{15}) or diterpenes (C_{20}) from the substrates geranyl diphosphate (GPP), farnesyl diphosphate (FPP) and geranyl geranyldiphosphate (GGPP), respectively (Degenhardt et al., 2009; Trapp and Croteau, 2001).

The oxygen-containing monoterpene 1,8-cineole (also known as eucalyptol) is one of the components of essential oils widely used in the pharmaceutical industry and in medicine as an antiseptic and analgesic, with immunostimulant activity, antitussive bronchodilator effects and myorelaxant properties (Bastos et al., 2010; Caldas et al., 2015). 1,8-Cineole was also reported to have protective effects against genotoxicity of ROS-inducing agents (Nikolić et al., 2011). This monoterpene also has important ecological effects and it acts as

both a deterrent of some phytophagous insects or as a fragrant attractant to pollinators (Al-Alawi, 2014). Due to the variety of applications of 1,8-cineole, there is a great interest in characterizing the mechanism of its biosynthesis in respect to the genes that encode the proteins responsible for its production.

1,8-Cineole-rich essential oils have been isolated from some *Thymus* species (Lamiaceae) grown in Portugal, namely *T. albicans*, *T. camphoratus*, *T. capitellatus*, *T. lotocephalus*, *T.* x *mourae*, *T. mastichina* subsp. donyanae, *T. mastichina* subsp. *mastichina*, *T. villosus* subsp. *villosus* and *T. zygis* subsp. *sylvestris* (Figueiredo et al, 2008; Miguel et al., 2004a; Salgueiro et al., 1997). *T. albicans* Hoffmanns. & Link is an aromatic wild shrub that grows in the South of the Iberian Peninsula in dry and sandy soils consolidated near the coast or in clearings of pine forests. *T. albicans* EOs have previously been analyzed and three main chemotypes, 1,8-cineole, linalool and linalool/1,8-cineole identified (Miguel et al., 2004b; Salgueiro et al., 1997). The 1,8-cineole content in the chemotype generally ranged from 29% to 43% (Salgueiro et al., 1997) reaching maximal levels of up to 68% (Miguel et al., 2003).

The present study reports the isolation of a cDNA encoding a putative 1,8-cineole synthase from *T. albicans* leaves of a 1,8-cineole chemotype, using the RACE technique. A recombinant 1,8-cineole synthase was generated and functionally characterized in three distinct strains of *Escherichia coli*. To gain insight into the evolution of the monoterpene synthase family in the Lamiaceae, phylogenetic analysis of sequences from 34 species of the family was performed.

Materials and Methods

Plant material and reagents

Thymus albicans Hoffmanns. & Link branches, with essential oil rich in 1,8-cineole, were collected from a plant from a wild population at the Algarve University (Faro, Portugal) campus. Branches were collected during the flowering season (from May to July) to ensure correct specimen identification. Leaves were detached and used for total RNA extraction.

RNA extraction and cDNA synthesis

Aproximately 100 mg of leaves were used for total RNA (tRNA) extraction using Tri Reagent™ (Sigma, Spain) and following the manufacturer´s instructions. Contaminating genomic DNA was removed from 10 µg of tRNA using a TURBO DNA-free™ Kit (Ambion, Germany). The quality and concentration of tRNA was assessed using a Nanodrop 1000 (Thermo Scientific, Waltham, MA, USA) at 260 nm and by visualization after agarose gel electrophoresis. Reverse transcription reactions were performed for a final reaction volume of 40 µL using 5 µg of tRNA with 1 µL of random hexamers (200 ng/µl, Jena Bioscience, Germany), 1 µL of 10 mM dNTP mix, 4 U of the RNasin® RNase Inhibitor (Promega, USA) and 40 U of M-MLV Reverse Transcriptase (Promega, USA). RNA was denatured at 65 ºC for 10 min and ice cooled for 5 min, then the reaction mix was added and samples were incubated for 2 h at 37 ºC.

PCR

T. albicans 1,8-cineole synthase was isolated from cDNA by PCR using the primers Mono 1 Fw and Mono D1 Rv (Table 1) that were designed for conserved nucleotide regions identified using a multiple sequence alignment [using clustal X program available from the Geneious v6.1.5 software (http://www.geneious.com, Kearse et al., 2012).

of the homologue transcripts from related plant species, *Rosmarinus officinalis* (DQ839411), *Salvia officinalis* (AF051899) and *Salvia fruticosa* (DQ785793) (Table 1). Degenerate primers were designed using the OLIGO Primer Analysis software v.6 (Rychlik, 2007).

The reaction was carried out in a final volume of 25 µL containing 2 µL of cDNA (0.5-1 µg), 200 nmol of each primer, 1× Advantage HD buffer (Clontech, USA), 100 µM of each dNTP (Invitrogen) and 0.25 U of Advantage® HD Polymerase (Clontech, USA) and amplification reactions performed in a thermal cycler MyCycler™ (BioRad, USA) using the following cycles: 95 ºC for 2 min, followed by 35 cycles at 95 ºC for 30 s, 48 ºC for

1 min and 72 ºC for 90 s, with an extension time of 7 min at 72 ºC. A negative control reaction without cDNA was also included.

PCR products with the predicted amplicon size were gel purified using the Illustra GFX PCR DNA and a Gel Band Purification Kit (GE Healthcare, UK), cloned into the pGEM T-Easy vector (Promega, USA), sequenced and their identity confirmed by searching against the NCBI database (http://blast.ncbi.nlm.nih.gov/Blast.cgi) using the blastx algorithm.

RACE-PCR

RACE-PCR was used to amplify the missing 5' and 3' ends of the *T. albicans* putative TPS using a 5′/3′RACE kit 2nd Generation (Roche, Germany) and primers designed using the sequence of the amplified PCR fragment. cDNA was synthesized from 5 µg of tRNA using the gene specific antisense primer Cin10 Rv (Table 1), 10 µM dNTPs mix (Invitrogen), 5 µM of DTT (Invitrogen), 40 U of RNasin® RNase Inhibitor (Promega) and 200 U of SuperScript™ III Reverse Transcriptase (Invitrogen, USA). RNA was eliminated using RNase A (Panreac Applichem, Germany). An homopolymeric A-tail was added to the 3' end of the cDNA (Table 1) using Recombinant Terminal Deoxynucleotidyl Transferase (Promega, USA) and 4 mM dATP (Invitrogen). PCR was then performed using the dA-tailed cDNA, the Oligo d(T) Anchor Primer (Table 1) and the antisense specific primer Cin10 Rv. Subsequently nested PCR reactions were performed using the Anchor BP and the Anchor BP Nested as forward primers and the gene specific reverse primers Cin8 Rv and Cin6 Rv for the second and third amplifications, respectively (Table 1). The PCR reactions were performed for a final volume of 25 µL as follows: 1 µL of PCR product, 1× DyNazyme™ Ext buffer provided with 1.5 mM $MgCl_2$ (Thermo Scientific, USA), 1.25 U of DyNazyme™ Ext DNA Polymerase (Thermo Scientific, USA), 200 µM dNTPs (Invitrogen) and 200 nM of each primer. The PCR cycling conditions were 94 ºC for 2 min, 94 ºC for 30 s, annealing at 45 ºC for 45 s and elongation at 72 ºC for 90 s, with a final extension at 72 ºC for 7 min. The amplified products were gel

purified and cloned into the pGEM T-Easy vector (Promega) and sequenced to confirm identity.

For 3´-end amplification, cDNA was synthesized using the Oligo d(T) Anchor Primer with conditions similar to those used for the 5´-end amplification. PCR reactions were performed with the Anchor BP and the Anchor BP Nested primers in combination with the gene specific forward primers Cin13 Fw and Cin3 Fw, respectively (Table 1). The amplified products of the expected length were cloned into the pGEM T-Easy vector and sequenced. The information obtained by 3´- and 5´- RACE-PCR was used to design primer pairs that amplify the complete sequence of the putative TPS (Table 1). TargetP 1.1 Server (http://www.cbs.dk/services/TargetP/) (Emanuelsson et al., 2000), PWOLF PSORT (https://wolfpsort.hgc.jp/) (Horton et al., 2007) and PREDOTAR (http://urgi.versailles.inra.fr/predotar) (Small et al., 2004) were used to predict the putative plastidial transit peptide and to analyse the C-terminal region in the deduced protein of the cloned *TalbCin* cDNA. The isolated *TalbCin* was deposited in GenBank with the accession number KX452344.

Construction of the recombinant expression vector

The ORF of *TalbCin* terpene synthase excluding the sequence encoding the plastidial transit peptide was amplified by PCR using DyNAzyme™ Ext DNA Polymerase (Thermo Scientific, USA) with the primers FullCin Fw and FullCin Rv containing restriction digestion sites (Table 1) to allow directional cloning in the expression vector. The PCR reaction was performed as described above for 5´-end amplification except that the annealing temperature used was 58 ºC for 45 s. The PCR product was exposed to restriction enzymes, gel purified and ligated into the pGEM T-Easy vector and sequenced to confirm its identity. To obtain a recombinant 1,8-cineole fusion protein for functional characterization, the *TalbCin* transcript sequence was subcloned into the expression vector pCR®T7/NT-TOPO® (Invitrogen) in frame with the T7 promoter and the N-terminal 6xHis-tag sequence. The recombinant expression vector was inserted into One Shot DH5α

competent *Escherichia coli* cells and plasmid DNA was purified and the authenticity of the vector insert and the correct reading frame confirmed by sequencing.

Recombinant protein expression

The pCR®T7/NT-TOPO®/*TalbCin* expression construct (10 ng of each) was used to transform three different *Escherichia coli* strains, BL21(DE3)Star, BL21-CodonPlus(DE3) (both from Invitrogen) and Rosetta(DE3)pLysS (Novagen, Germany) to identify the best heterologous system for recombinant protein expression. Transformed *E. coli* were grown in LB medium supplemented with ampicillin (100 µg/mL) in the absence or presence of chloramfenicol (30 µg/mL, Rosetta cells). Then 1 mL of the starter culture was used to inoculate 100 mL of LB medium containing the appropriated antibiotics and the culture was incubated at 37 ºC and 200 rpm.

For protein expression several temperatures (16 ºC, 20 ºC, 25 ºC and 37 ºC) and induction times (10 h, 16 h, 24 h and 48 h) were tested. In face of the results obtained, all induction assays were subsequently performed overnight for 14-16 h at 18 ºC. IPTG (Sigma-Aldrich, Spain) was added at a final concentration of 1 mM to transformed bacteria ($OD_{600\ nm}$ = 0.5-0.6). Recombinant protein production was analysed in soluble and insoluble fractions after removal of cell debris by centrifugation at 13,000 g for 15 min at 4 ºC, according to Lima et al. (2013).

SDS-polyacrylamide gel electrophoresis and Western Blot

Insoluble and soluble protein extracts of TalbCin protein was analyzed by 12% SDS-polyacrylamide gel electrophoresis (PAGE) and Western blot analysis. IPTG induced *E. coli* cells transformed with an empty pCR®T7/NT-TOPO® vector were used as a negative control (C^-). Non-induced IPTG bacterial cultures were also used for comparison. About 20 µg (per lane) of total protein (soluble or insoluble protein extracts), estimated using a Nanodrop 1000 (Thermo Scientific, Waltham, MA, USA), was separated by SDS-PAGE at a constant power (100 V). Two gels were run and one was stained with Coomassie

Brilliant Blue [0.1% Coomassie Blue R, methanol 50% (v/v), glacial acetic acid 10% (v/v)] for visual inspection and the other was used for Western blot analysis.

The Western blot methodology was as described in Anjos et al. (2013) with some modifications. In brief, the blocking step was performed overnight at 4 ºC in 10% (w/v) milk, 0.1% (v/v) Tween 20 in 1x Tris-buffered saline [TBS, 100 mM Tris-HCl pH 7.5, 0.9% (w/v) NaCl]. The immunoblotting step was carried out using as a primary antibody mouse anti-His-tag (1/3,000, Amersham GE Healthcare, UK), and the secondary antibody was biotinylated anti-mouse IgG (1/400, Amersham GE Healthcare, UK) followed by streptavidin-horseradish peroxidase conjugate (1/1,000, Amersham GE Healthcare, UK). All the antibody incubation steps were carried out for 1 h and the washing steps between the primary and secondary antibody incubations were 3 x 10 min in 1x TBS-T [TBS buffer with 0.1% (v/v) Tween 20]. Membranes were developed with 3,3´-diaminobenzidine (DAB, 0.2 mg/ml in 100 mM Tris-HCl pH7.5, 0.003% H_2O_2) with nickel chloride intensification (0.40 mg/ml $NiCl_2$). The prestained SDS-PAGE standard (BioRad, USA, Cat. No. 161-0318) and the PageRuler™ Prestained Protein Ladder (Thermo Scientific, USA, Cat. No. 26616) were used as molecular weight markers.

The soluble fraction of the recombinant 1,8-cineole synthase was purified using HisTrap™ FF columns (Amersham GE Healthcare, UK) following the manual instructions. The column was equilibrated with binding buffer (20 mM sodium phosphate buffer containing 500 mM NaCl and 5 mM imidazole at pH 7.5). The recombinant TalbCin was eluted with elution buffer (20 mM sodium phosphate buffer with 500 mM NaCl, pH 7.5) containing increasing concentrations of imidazole in a step-wise fashion from 25 mM to 400 mM. Eluted fractions were analysed by SDS-PAGE and Western blot.

Enzymatic assays and headspace solid phase micro extraction (SPME)

Enzyme activity of recombinant 1,8-cineole synthase expressed and assayed at least twice in several different *E. coli* strains were assayed following the approach of Mendes et al. (2014), and using 50 µM GPP or FPP (Echelon, USA) as the substrate.

Protein extracts from IPTG induced *E. coli* cells transformed with the empty expression vector were used as a negative control. Terpene products were collected by solid phase micro extraction (SPME) after incubation for 45 min, as described in Mendes et al. (2014). The assays were performed at 32 ºC and the products analysed by GC and GC-MS. Each enzymatic assay was performed twice.

Gas Chromatography (GC) analysis

Immediately after sampling, an SPME needle was introduced into the injector of a PerkinElmer Clarus 400 gas chromatograph equipped with two flame ionization detectors (FIDs), a data-handling system and a vaporizing injector port into which two columns of different polarities were installed: a DB-1 fused-silica column (polydimethylsiloxane, 30 m x 0.25 mm i.d., film thickness 0.25 μm; J & W Scientific Inc., Rancho Cordova, CA, USA) and a DB-17HT fused-silica column [(50% phenyl)-methylpolysiloxane, 30 m x 0.25 mm i.d., film thickness 0.15 μm; J & W Scientific Inc.]. Analytes desorption was achieved in the splitless mode at 250 ºC for 1 min. The oven temperature was programmed for 45-175 ºC, at 3 ºC/min, subsequently at 15 ºC/min up to 300 ºC, and then held isothermal for 10 min; injector and detector temperatures were 250 ºC and 300 ºC respectively; and the carrier gas (hydrogen) was adjusted to a linear velocity of 30 cm/s. Analytes desorption was confirmed by subjecting the SPME fibre to a second run, which always showed no carry-over peaks. The percentage of the detected compounds was determined by integration of the peak areas without the use of a correction factor.

Gas Chromatography-Mass Spectrometry (GC-MS) analysis

Directly after sampling, the SPME needle was introduced into the split/splitless injector of the GC-MS unit that consisted of a PerkinElmer Clarus 600 gas chromatograph, equipped with DB-1 fused-silica column (30 m x 0.25 mm i.d., film thickness 0.25 μm; J & W Scientific, Inc.), and interfaced with a Clarus 600T Mass Spectrometer (software version 5.4.2.1617, Perkin Elmer, Shelton, CT, USA). Injector and oven temperatures were as

above; transfer line temperature, 280 ºC; ion source temperature, 220 ºC; carrier gas, helium, adjusted to a linear velocity of 30 cm/s; splitless mode; ionization energy, 70 eV; scan range, 40-300 u; scan time, 1 s. The identities of the detected moleculaes were assigned by comparing their retention indices, relative to *n*-alkane indices and GC-MS spectra from a lab-made library, which was constructed based on reference essential oils, and commercially available standards.

Multiple sequence comparisons and phylogeny

The predicted amino acid sequence of the isolated TalbCin was compared with that of other mTPS from other Lamiaceae, retrieved from GenBank. Sequences were aligned using ClustalW software (http://www.genome.jp/tools/clustalw/). The percent of sequence similarity was determined using GeneDoc software (version 2.7.0) (Nicholas et al., 1997) on sequences starting at the RRx8W motif onwards to the stop codon. The aligned sequences were submitted to ProTest 2.4 Server (Abascal et al., 2005) to select the best model to study protein family evolution. Evolutionary analysis was performed using the maximum likelihood method using the ATGC Montpellier Bioinformatics platform (http://www.atgc-montpellier.fr/) (Guindon et al., 2010) using a JTT matrix with a proportion of invariable sites (0.072) and a gamma shape of 2.042 with 100 bootstrap replicates. The output file was opened in FigTree v.1.4.2 and the tree was edited using Inkscape 0.91 program. The deduced amino acid sequences of 1,8-cineole synthase from *T. albicans* and the 1,8-cineole synthase homologues retrieved from GenBank, were aligned and compared as described above, using a LG matrix, a proportion of invariable sites (0.044) and a gamma shape of 1.667 with 100 bootstrap replicates. Both trees were rooted with the amino acid sequence of a 9-*epi*-caryophyllene sesquiterpene synthase from *Lavandula* x *intermedia* (GenBank accession no. AGU13712).

Results and discussion

Thymus albicans 1,8-cineole synthase

The cDNA from *T. albicans* 1,8-cineole chemotype was amplified and one transcript with a partial sequence of 540 bp in length was obtained. *TalbCin* full-length sequence containing a partial 5´-end shared 78% amino acid sequence similarity with the 1,8-cineole synthase from *Rosmarinus officinalis* (AFZ41794) and 77% with *S. officinalis* (AAC26016) and *Rosmarinus officinalis* (ABI20515). The highest sequence similarity between TalbCin and mTPS from other *Thymus* species was 63% for *T. caespititius* (AID51194), *T. vulgaris* (ALB78115) and *T. serpyllum* (AGT29345) and 62% for *T. citriodorus* (AGT29350).

The N-terminus of mTPS sequences contains a 50-70 bp long transit peptide, that directs the pre-protein to the plastids where proteolytic processing occurs to release the mature protein. There are indications that the cleavage site *in vivo* is immediately upstream of the N-terminal RRx_8W motif (Bohlmann et al., 1998; Williams et al., 1998). Several studies have reported that, when expressed in *E. coli,* mTPSs are active both as full-length proteins and after deletion of the predicted plastidial targeting sequence [Crocoll et al., 2010; Mendes et al., 2014 (Tctps4)]. Other studies refer to a high activity of the mTPSs only in the absence of the transit peptide [Crowell et al., 2002; Lima et al., 2013; Mendes et al., 2014 (Tctps2)]. Thus the transit peptide is usually removed to generate a functional protein following heterologous expression, and to prevent inclusion body formation, a frequent problem when mTPSs are expressed in bacterial system (Bohlmann et al. 1998; Landmann et al., 2007; Williams et al., 1998). The expressed TalbCin putative protein started with the two amino acids (Ser^6Thr^7) that precede the double-arginine (RRw8W) motif (Fig. 1). The signal peptide of sequences coding for TPSs characteristically contain low amounts of acidic amino acids and a high serine and threonine residue content (Degenhardt et al., 2009; von Heijne et al., 1989; Wise et al., 1998). In the predicted TalbCin protein, a small sequence of seven aminoacids $T^1DEASST^7$ was found before the RRx_8R motif (Fig. 1) and was rich in serine and threonine residues. However, sequence analysis using TargetP 1.1,

PWOLF PSORT and PREDOTAR failed to identify the $T^{1}DEASST^{7}$ sequence as being a chloroplast transit peptide, possibly due to its short length. The initial methionine residue of the transit peptide was not found using RACE-PCR so it was concluded that TalbCin isolated sequence is partial and lacks the complete plastid transit peptide characteristic of mTPSs. Short transit peptides are not common in mTPSs although the one from γ-terpinene synthase from *T. caespititius* (tctps4) was predicted to have 3-4 amino acids (Mendes et al., 2014). In general, mTPS with the N-terminal transit peptide are 600-650 amino acids long (Bohlmann et al., 1998; Degenhardt et al., 2009). The cloned cDNA sequence of the TPS TalbCin was 1803 bp in length which corresponded to a predicted 593 amino acid protein from the RR motif to the Stop codon. The predicted molecular mass was 69.18 kDa and the pI 4.76 and multiple sequence alignment revealed that the TalbCin was 51 amino acids longer than the other identified plant 1,8-cineole synthases (Fig. 1) due to the presence of an extended C-terminus. Sequence analysis of TalbCin using TargetP 1.1, PWOLF PSORT and PREDOTAR failed to detect any active site motifs in the C-terminal extension. Multiple sequence alignments of mTPS sequences from plants of the Lamiaceae family (see Fig. 2) and of 1,8-cineole synthases from several plant families, retrieved from GenBank (see Fig. 3), revealed that an extended C-terminal sequence is not a common feature of the cineole synthases, with *Solanum lycopersicum* (AEM05857) being the only other species with an extended sequence of 22 amino acids.

TPS are classified into six subfamilies (Bohlmann et al., 1998), from *Tps-a* to *Tps-g* and those that are associated with secondary metabolism are members of the *Tps-a*, *Tps-b* and *Tps-d* subfamilies. The mTPS from the Lamiaceae family cluster with the *Tps-b* subfamily and despite their sequence diversity, they share several conserved amino acid residues in both the N- and C-terminal domains. According to Bohlmann et al. (1998), *Tps-b* (and also *Tps-a* and *Tps-d*) subfamilies along with the conserved motifs described below, also possess conserved positions within the mature protein that corresponded to TalbCin $Trp^{18, 249, 322, 381}$, $Asp^{445, 527}$, $Arg^{110, 263, 441}$, $Glu^{242, 268, 378, 392}$, $Pro^{331, 388}$, His^{75} and Cys^{464}. Conserved aromatic amino acids were also found in TalbCin sequences, namely Phe^{261} and $Tyr^{270, 338, 522}$ (Fig. 1). Other highly conserved regions of mTPS are the RWW, YMQ and

CYMxE motifs that were also present in TalbCin (Fig. 1) (Wise et al.,1998). Characteristic motifs in the N-terminal domain of monoterpene and some sesquiterpene synthases are the RR(x)8W motif that is involved in the first isomerization step to convert the substrate geranyl diphosphate to linalyl diphosphate prior to cyclization (Dudareva et al., 2003; Williams et al., 1998) and the (M/L)L(S/Q/N)L(F/Y)EAS motif, thought to be part of the active site (Li and Fan, 2011; Wise et al., 1998) (Fig. 1). Conserved motifs present in the C-terminal domain of enzymes using prenyl diphosphate substrates such as the aspartate-rich DDxxD motif and the NSE/DTE [(L/V)(V/L/A)(N/D)Dx2(S/T)x3E] motif known to be involved in the binding of a divalent metal ion cofactor during the enzymatic action (Bohlmann et al., 1998; Christianson, 2006; Degenhardt et al., 2009; Williams et al. 1998), were completely conserved in the TalbCin amino acid sequence.

The deduced amino acid sequence of 1,8-cineole synthase and biochemically characterized mTPS from 34 species of the Lamiaceae family were aligned and a phylogenetic tree was constructed. In general, mTPS tended to cluster according to the genus and not according to function (Bohlmann et al., 1998; Degenhardt et al., 2009; Demissie et al., 2012; Landmann et al., 2007; Lima et al., 2013), as occurrs in *Lavandula* and *Ocimum* groups. Phylogenetic analysis confirmed the identity of the isolated TalbCin, as it was clustered with 1,8-cineole synthase from *Salvia, Rosmarinus* and *Lavandula* spp. (Fig. 2). The exception was *S. fruticosa* (ACM89961) 1,8-cineole synthase that grouped with the limonene group suggesting that conservation of function does not necessarily equal sequence conservation and the amino acid sequence of enzymes that have retained similar functions may have evolved differently.

The 1,8-cineole synthases have been isolated and functionally characterized from a wide range of plant species. To gain insight into relatedness of the TalbCin amino acid sequence to other cineole synthases, a phylogenetic tree was constructed with TalbCin and the amino acid sequence of 1,8-cineole homologues retrieved from GenBank (Fig. 3). The cineole synthases were clustered according to the plant family they represent, namely the Lamiaceae, the Brassicaceae, the Solanaceae, the Asteraceae and the Pinaceae, thus revealing that amino acid sequence identities or similarities are higher within each plant

family. The Pinaceae 1,8-cineole synthases form a separate branch from the angiosperms and the pattern of bifurcation from a common ancestor implies independent functional specialization after separation of the angiosperm and gymnosperm lineages, which is in accordance with the TPS classification reported by Bohlmann et al. (1988).

Functional characterization of 1,8-cineole synthase

Heterologous expression of the recombinant fusion TalbCin protein was assayed in *E. coli* BL21(DE3)Star, BL21-CodonPlus(DE3) and Rosetta(DE3)pLysS and the soluble fraction of bacterial extracts were used for enzymatic assays. Both insoluble and soluble bacterial extracts were assayed by Western blot using an anti-His-tag monoclonal antibody to compare levels of expression of the recombinant fusion protein from 10 h to 48 h at 20 ºC, after induction with IPTG. Western blot analysis showed that the recombinant fusion protein was successfully expressed and its molecular weight (MW, 72.8 kDa) was similar to that predicted and it was present in both the soluble and insoluble fraction [see Fig. 4A and 4B for BL21-CodonPlus(DE3) strain]. Insoluble fractions obtained from non-induced bacteria BL21-CodonPlus(DE3) also contained immunoreactive protein and suggested there was leaky expression of the protein (Rosano and Ceccarelli, 2014). No protein expression was detected in the soluble bacterial protein extracts transformed with the empty vector (C^-, Fig. 4C). Furthermore, since Western blots didn't show major differences on protein production after 10 h or 48 h induction (Fig. 4B), all induction assays were subsequently performed overnight for a total of 14-16 h. From the set of temperatures tested (16 ºC, 20 ºC, 25 ºC and 37 ºC), a diminished protein expression, detected by Western blot analysis, was obtained only at 37 ºC (data not shown), and since there were no distinction between the other temperatures tested, all the assays were subsequently performed at 18 ºC, a low temperature choosed to promote a high solubility of recombinant proteins (Rosano and Ceccarelli, 2014).

Heterologous expression of the TalbCin recombinant fusion protein performed in three *E. coli* strains [BL21(DE3)Star, BL21-CodonPlus(DE3) and Rosetta(DE3)pLysS] was not

significantly different. The TalbCin protein displayed a low activity in the presence of the substrate GPP, producing 1,8-cineole as the main product (Fig. 5A). The low enzymatic activity detected may be related to the low protein yield in the soluble fraction, possibly caused by protein misfolding or codon usage bias, despite using the Rosetta(DE3)pLysS host strain that is suited for target genes containing rare *E. coli* codons (Chen and Texada, 2006). The results may indicate that for proper functioning, mTPS proteins may require processing or secondary modifications that are not performed by bacteria (Jia et al., 1999). The His-tag (6xHis-tag) is one of the simplest and most widely used purification tag for its small size and charge with minimal or no impact on folding and biological functions of the target protein, although sometimes can interfere in these processes (Yadav et al, 2016). Results from the *in vitro* tests contrast with the high enzymatic activity of the enzyme *in vivo* with up to 68% of 1,8-cineole content in the essential oil of *T. albicans* being of the 1,8-cineole chemotype (Miguel et al., 2003). The soluble protein fractions extracted from the *E. coli* strains transformed with the empty expression vector (negative control) had no enzymatic activity (data not shown), which is coherent with the failure to detect the protein by Western blot analysis (Fig. 4C). Further, the attempts to purify 1,8-cineole synthase with HisTrap™ FF columns failed owing to the low yield of protein, or to weak binding of the tagged protein to the resin, a problem sometimes associated with His-tagged proteins (Yadav et al, 2016).

The formation of multiple products by specific mTPS has been associated with the succession of unstable carbon cation intermediates in electrophilic cyclization of GPP (Fig. 6) (Davis and Croteau, 2000). The 1,8-cineole synthases are typical multi-product enzymes. The 1,8-cineole synthases isolated thus far from plants of the Solanaceae family such as *Nicotiana noctiflora* and *Nicotiana* species from section *Alatae* synthesized the so called “cineole cassette” that comprised seven monoterpenes: 1,8-cineole, limonene, myrcene, β-pinene, α-pinene, sabinene and α-terpineol (Fähnrich et al., 2012; Fähnrich et al., 2014). The cineole synthases of the Pinaceae family (Fig. 3) do not produce limonene or sabinene and instead γ-terpinene is synthesized by *Picea glauca* × *P. engelmannii* (Keeling et al., 2011). *Salvia officinalis* from the Lamiaceae family synthesizes all

monoterpenes of the "cineole cassette" (Wise et al., 1998), while the 1,8-cineole synthase from *Lavandula x intermedia* produces only three of the compounds, namely sabinene, α-terpineol and limonene, along with β-phellandrene as minor product (Demissie et al., 2012). Monoterpene production has been directly correlated with the transcriptional activity of the mTPSs (Demissie et al., 2012) and the absence of some products of the "cineole cassette" in some plant species and in *Citrus unshiu* Marc that produces mainly cineole (Shimada et al., 2005) may be related to the use of *in vitro* synthesis systems and the way the enzyme is proteolytically processed in the *E. coli* host that may compromise substrate and intermediate binding conformations (Crocoll et al., 2010). Whether TalbCin is a multi-product enzyme remains to be clarified, since minor products of α-pinene, β-pinene, sabinene, β-myrcene and 1,8-cineole were also detected during GC-MS analysis (Fig. 5A), indicating the involvement of 1,8-cineole synthase in multiple reaction mechanisms (Fig. 6). However, owing to the poor activity of the enzyme it is likely that the recombinant enzyme produced may have been insufficient to enable the detection of other products. No product was formed upon exposure to FPP (Fig. 5B), thus confirming that the putative TalbCin has no sesquiterpene synthase activity.

The results of *in vitro* enzyme activity, combined with the conserved amino acids and protein motifs in the deduced protein that are common to other TPS and specific to mTPS, confirmed that TalbCin is a real mTPS.

Conclusion

In the present work, a 1,8-cineole synthase involved in essential oil production in *T. albicans*, chemotype 1,8-cineole, was isolated and functionally characterized constituting the first report for this species. The putative amino acid sequence of 1,8-cineole synthase shared all characteristic conserved motifs with the homologues from other plants TPS. Phylogenetic analysis indicates that 1,8-cineole synthase from *T. albicans* is closer to the other 1,8-cineole synthases from the Lamiaceae family than to the TPS from the genus *Thymus*. Recombinant 1,8-cineole synthase expression in bacteria resulted in low

enzymatic activity with production of 1,8-cineole as the main product and also other minor products, suggesting that 1,8-cineole synthase may be a multi-product enzyme.

Acknowledgements

Alexandra Filipe was funded by Ceratonia Project MONOTHYMUS, Universidade do Algarve, Portugal. João CR Cardoso was funded by an auxiliary research contract (UID/Multi/04326/2013) and Liliana Anjos by Post-doctoral fellowship [SFRH/BPD/79105/2011] from FCT, the Ministry of Science, Portugal. This study was partially funded by FCT, under UID/AMB/50017/2013, FEDER PT2020-Compete 2020 and by UID/Multi/00631/2013/CEOT.

References

Abascal, F., Zardoya R., Posada D., 2005. ProtTest: Selection of best-fit models of protein evolution. Bioinformatics. 21(9), 2104-2105.

Al-Alawi, M.S., 2014. Efficacy of essential oils from medicinal plants in control of the hairy rose beetle, tropinota squalid (scopoli) and their comparative toxicity to the honey bee, *Apis mellifera* L. AJABS. 9(3), 284-288. DOI:10.3844/ajabssp.2014.284.288.

Anjos L., Gomes A.S., Melo E.P., Canario A.V., Power D.M., 2013. Cartilage acidic protein 2 a hyperthermostable, high affinity calcium-binding protein. Biochim. Biophys. Acta. 1834, 642-650. http://dx.doi.org/10.1016/j.bbapap.2012.12.012.

Bastos, V.P.D., Gomes, A.S., Lima, F.J.B., Brito, T.S., Soares, P.M.G., Pinho, J.P.M., Silva, C.S., Santos, A.A., Souza, M.H.L.P., Magalhães, P.J.C., 2010. Inhaled 1,8-cineole reduces inflammatory parameters in airways of ovalbumin-challenged guinea pigs. Basic Clin. Pharmacol. Toxicol. 108, 34-39. DOI:10.1111/j.1742-7843.2010.00622.x

Bohlmann, J., Meyer-Gauen, G., Croteau, R., 1998. Plant terpenoid synthases: molecular biology and phylogenetic analysis. PNAS. 95, 4126-4133.

Caldas, G.F.R., da Silva Oliveira, A.R., Araújo, A.V., Lafayette, S.S.L., Albuquerque, G.S., Silva-Neto, J.C., Costa-Silva, J.H., Ferreira, F., da Costa, J.G.M., Wanderley,

A.G., 2015. Gastroprotective mechanisms of the monoterpene 1, 8-cineole (eucalyptol), PloS One 10(8). DOI:10.1371/journal.pone.0134558.

Chen, D., Texada, D.E., 2006. Low-usage codons and rare codons of *Escherichia coli*, mini review. Gene Ther. Mol. Biol. 10, 1-12.

Christianson, D.W., 2006. Structural biology and chemistry of the terpenoid cyclases. Chem. Rev. 106, 3412-3442.

Crocoll, C., Asbach, J., Novak, J., Gershenzon, J., Degenhardt, J., 2010. Terpene synthases of oregano (*Origanum vulgare* L.) and their roles in the pathway and regulation of terpene biosynthesis. Plant Mol. Biol. 73, 587-603. DOI:10.1007/s11103-010-9636-1.

Crowell, A.L., Williams, D.C., Davis, E.M., Wildung, M.R., Croteau, R., 2002. Molecular cloning and characterization of a new linalool synthase. Arch. Biochem. Biophys. 405, 112-121.

Davis, E.M., Croteau, R., 2000. Cyclization enzymes in the biosynthesis of monoterpenes, sesquiterpenes, and diterpenes, in: Leeper, F.J., Vederas, J.C. (Eds.), Biosynthesis, series Top. Curr. Chem. Springer-Verlag, Berlin, Heidelberg, 209, pp. 53-95.

Degenhardt, J., Köllner, T.G., Gershenzon, J., 2009. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants. Phytochemistry. 70, 1621-1637. DOI:10.1016/j.phytochem.2009.07.030.

Demissie, Z.A., Cella, M.A., Sarker, L.S., Thompson, T.J., Rheault, M.R., Mahmoud,S.S., 2012. Cloning, functional characterization and genomic organization of 1,8-cineole synthases from *Lavandula*. Plant Mol. Biol. 79, 393-411. DOI:10.1007/s11103-012-9920-3.

Dudareva, N., Martin, D., Kish, C.M., Kolosova, N., Gorenstein, N., Fäldt, J., Miller, B., Bohlmann, J., 2003. (E)-β-Ocimene and myrcene synthase genes of floral scent biosynthesis in snapdragon: function and expression of three terpene synthase genes of a new terpene synthase subfamily. Plant Cell. 15, 1227-1241. DOI/10.1105/tpc.011015.

Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300, 1005-16.

Fähnrich, A., Brosemann, A., Teske, L., Neumann, M., Piechulla, B., 2012. Synthesis of 'cineole cassette' monoterpenes in *Nicotiana* section *Alatae*: gene isolation, expression, functional characterization and phylogenetic analysis. Plant Mol. Biol. 79, 537-553. DOI:10.1007/s11103-012-9933-y.

Fähnrich, A., Neumann, M., Piechulla, B., 2014. Characteristic alatoid 'cineole cassette' monoterpene synthase present in *Nicotiana noctiflora*. Plant Mol. Biol. 85, 135-145. DOI: 10.1007/s11103-014-0176-y.

Figueiredo, A.C., Barroso, J.G., Pedro, L.G., Salgueiro, L., Miguel, M.G., Faleiro, M.L., 2008. Portuguese *Thymbra* and *Thymus* species volatiles: chemical composition and biological activities. Curr. Pharm. Des. 14, 3120-3140.

Gershenzon, J., McCaskill, D., Rajaonarivony, J.I.M., Mihaliak, C., Karp, F., Croteau, R., 1992. Isolation of secretory cells from plant glandular trichomes and their use in biosynthetic studies of monoterpenes and other gland products. Anal. Biochem. 200, 130-138.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59(3), 307-21.

Horton, P., Park, K.-J, Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., Nakai, K., 2007. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 35, W585–W587. DOI:10.1093/nar/gkm259.

Jia, J.-W, Crock, J., Lu, S., Croteau, R., Chen, X.-Y., 1999. (3R)-Linalool synthase from *Artemisia annua* L.: cDNA isolation, characterization, and wound induction. Arch. Biochem. Biophys. 372(1), 143-149.

Landmann, C., Fink, B., Festner, M., Dregus, M., Engel, K.H., Schwab, W., 2007. Cloning and functional characterization of three terpene synthases from lavender (*Lavandula angustifolia*). Arch. Biochem. Biophys. 465(2), 417-429.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software

platform for the organization and analysis of sequence data. Bioinformatics. 28(12), 1647-1649.

Keeling, C.I., Weisshaar, S., Ralph, S.G., Jancsik, S., Hamberger, B., Dullat, H.K., Bohlmann, J., 2011. Transcriptome mining, functional characterization, and phylogeny of a large terpene synthase gene family in spruce (*Picea* spp.). BMC Plant Biol. 11:43. http://www.biomedcentral.com/1471-2229/11/43.

Li, R., Fan, Y., 2011. Molecular cloning and expression analysis of a terpene synthase gene, HcTPS2, in *Hedychium coronarium*. Plant Mol. Biol. Rep. 29, 35-42. DOI:10.1007/s11105-010-0205-1.

Lima, A.S., Schimmel, J., Lukas, B., Novak, J., Barroso, J.G., Figueiredo, A.C., Pedro, L.G., Degenhardt, J., Trindade, H., 2013. Genomic characterization, molecular cloning and expression analysis of two terpene synthases from *Thymus caespititius* (Lamiaceae). Planta. 238, 191-204. DOI:10.1007/s00425-013-1884-2.

Mendes, M.D., Barroso, J.G., Oliveira, M.M., Trindade, H., 2014. Identification and characterization of a second isogene encoding γ-terpinene synthase in *Thymus caespititius*. J. Plant Physiol. 171, 1017–1027. http://dx.doi.org/10.1016/j.jplph.2014.04.001.

Miguel, G., Simões, M., Figueiredo, A.C., Barroso, J.G., Pedro, L.G., Carvalho, L., 2004a. Composition and antioxidant activities of the essential oils of *Thymus caespititius*, *Thymus camphoratus* and *Thymus mastichina*. Food Chem. 86, 183-188. DOI:10.1016/j.foodchem.2003.08.031.

Miguel, M.G., 2010. Antioxidant and anti-inflammatory activities of essential oils: a short review. Molecules. 15, 9252-9287. DOI:10.3390/molecules15129252.

Miguel, M.G., Duarte, F., Venâncio, F., Tavares, R., 2004b. Composition of the essential oils from portuguese *Thymus albicans* collected at different regions of Ria Formosa (Algarve). J. Essent. Oil Res. 16, 308-311. http://dx.doi.org/10.1080/10412905.2004.9698729.

Miguel, M.G., Figueiredo, A.C., Costa, M.M., Martins, D., Duarte, J., Barroso, J.G., Pedro, L.G., 2003. Effect of the volatile constituents isolated from *Thymus albicans*, *T.*

mastichina, *T. carnosus* and *Thymbra capitata* in sunflower oil. Nahrung/Food. 47(6), 397-402.

Nicholas, K.B., Nicholas Jr., H.B. Deerfield II, D.W., 1997. GeneDoc: Analysis and visualization of genetic variation. Embnet News. 4, 1-4.

Nikolić, B., Mitić-Ćulafić, D., Vuković-Gačić, B., Knežević-Vukčevic, J., 2011. Modulation of genotoxicity and DNA repair by plant monoterpenes camphor, eucalyptol and thujone in *Escherichia coli* and mammalian cells. Food Chem. Toxicol. 49, 2035–2045. DOI:10.1016/j.fct.2011.05.015.

Rosano, G.L., Ceccarelli, E.A., 2014. Recombinant protein expression in *Escherichia coli*: advances and challenges. Front. Microbiol. 5(172), 1-17. DOI:10.3389/fmicb.2014.00172.

Rychlik, W., 2007. OLIGO 7 Primer Analysis Software, in: Methods in Molecular Biology Vol. 402, PCR Primer Design, Yuryev, A. (Edt), Humana Press Inc., Totowa, NJ. pp. 35-59.

Salgueiro, L.R., Vila, R., Tomàs, X., Cañigueral, S., Proença da Cunha, A., Adzet, T., 1997. Composition and variability of the essential oils of *Thymus* species from Section *Mastichina* from Portugal. Biochem. Syst. Ecol. 25, 659-672. DOI:10.1016/S0305-1978(97)00046-X.

Shimada, T., Endo, T., Fujii, H., Hara, M., Omura, M., 2005. Isolation and characterization of (E)-beta-ocimene and 1,8 cineole synthases in *Citrus unshiu* Marc. Plant Sci. 168, 987-995. DOI:10.1016/j.plantsci.2004.11.012.

Small, I., Peeters, N., Legeai, F., Lurin, C., 2004. Predotar: a tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics. 4, 1581-1590. DOI:10.1002/pmic.200300776.

Trapp, S.C., Croteau, R.B., 2001. Genomic organization of plant terpene synthases and molecular evolutionary implications. Genetics. 158, 811-832.

von Heijne, G., Steppuhn, J., Herrmann, R.G., 1989. Domain structure of mitochondrial and chloroplast targeting peptides. Eur. J. Biochem. 180, 535-545.

Williams, D.C., McGarvey, D.J., Katahira, E.J., Croteau, R., 1998. Truncation of limonene

synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair. Biochemistry. 37, 12213-12220.

Wise, M.L., Savage, T.J., Katahira, E., Croteau, R., 1998. Monoterpene synthases from common sage (*Salvia officinalis*). cDNA isolation, characterization, and functional expression of (+)-sabinene synthase, 1,8-cineole synthase, and (+)-bornyl diphosphate synthase. J. Biol. Chem. 273, 24(12), 14891-14899.

Yadav, D.K., Yadav, N., Yadav, S., Haque, S., Tuteja, N., 2016. An insight into fusion technology aiding efficient recombinant protein production for functional proteomics. Arch. Biochem. Biophys. 612, 57-77. http://dx.doi.org/10.1016/j.abb.2016.10.012.

Table 1. Primer sequences used for *TalbCin* amplification. Enzyme restriction sites are underlined.

Name	Primer	Sequences (5´- 3´)
cDNA amplification	Mono1 Fw	GATTGTTACAATTGTATGAAGC
	Mono D1 Rv	AX(A/G)(C/T)ATC(A/G)(A/T)AXA(C/T)A(C/T)ATC(A/G)TC
5' RACE-PCR reverse primers	Cin6	GA(A/G)GCTTC(A/G)TACAATTGTAACA
	Cin8	CATCTCGAATCGATGACAAGATATATTG
	Cin10	AGTTTGGCTTCTAGTACAGTTGG
3' RACE-PCR forward primers	Cin3	GA(C/T)AGGATTGTGGAATGCTAC
	Cin13	TTTCGTGAGGGACAGGATT
	Cin15	ATGAATCCAACTGTACTAGAGC
RACE-PCR standard primers	Oligo d(T) Anchor*	GCTGTCAACGATACGCTACGTAACGGCATGACAGTGTTTTTTT TTTTTTTTTTT
	Anchor BP*	GCTGTCAACGATACGCTACGTAACG
	Anchor BP Nested*	CGCTACGTAACGGCATGACAGTG
Full length *TalbCin* amplification	FullCin Fw	GCGGATCCCACGCGACGATCGGGAAATT
	FullCin Rv	GCGAATTCTACGGCGCTTAATTAGGACG

* primer sequences according to Landmann *et al*. (2007). Primer FullCyn Fw with a restriction site for *Bam*HI underlined. Primer FullCyn Rv with a restriction site for *Eco*RI underlined. X = Inosin.

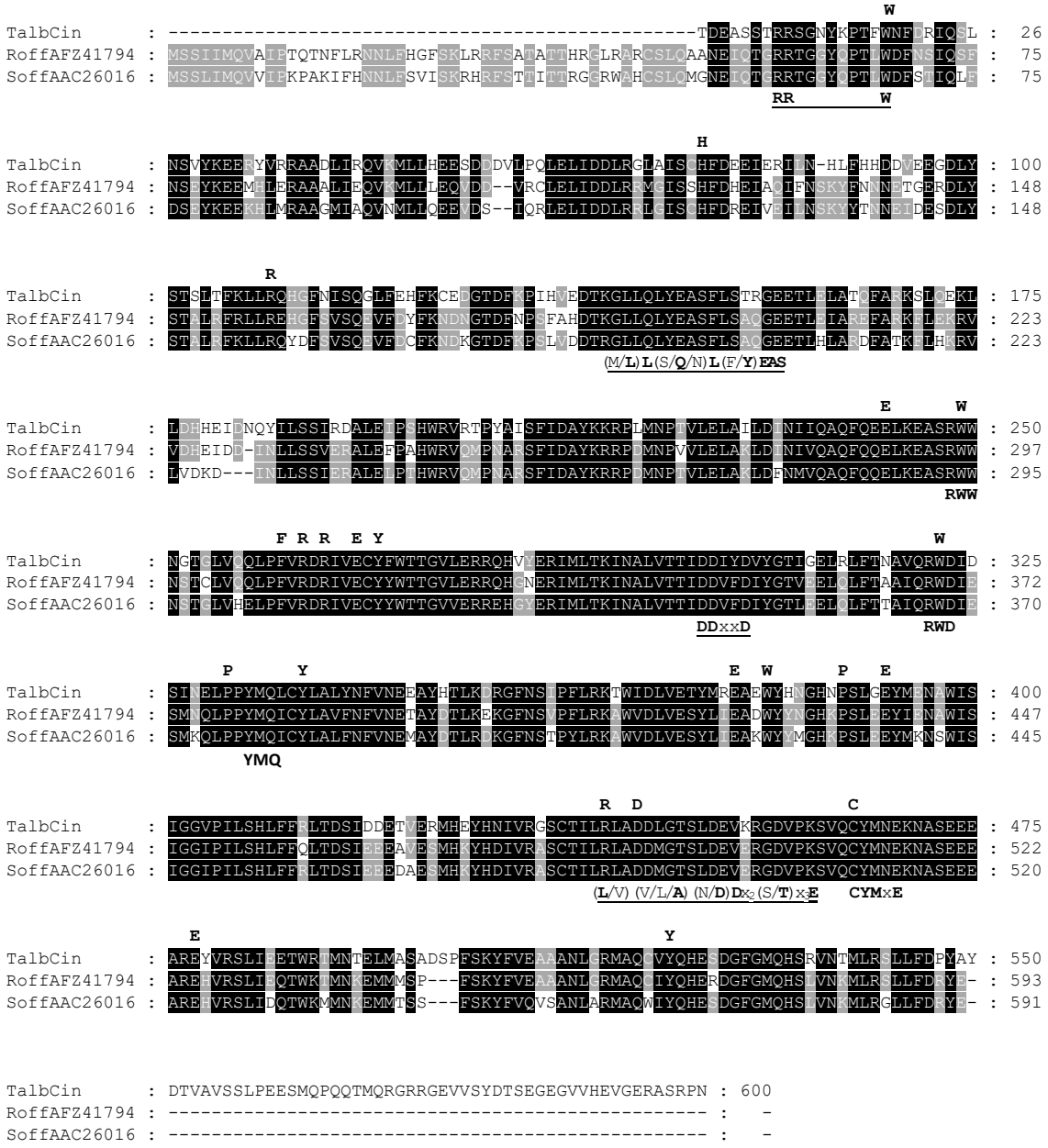

Fig. 1. Multiple sequence alignment (MSA) of the predicted amino acid sequence of *Thymus albicans* terpene synthase TalbCin with the putative 1,8-cineole synthase from *Rosmarinus officinalis* (AFZ41794) and *Salvia officinalis* (AAC26016). Note it was not possible to amplify the 5′-end of *TalbCin* and so this region is missing from the alignment. The MSA was constructed using ClustalW and was edited using GeneDoc software. Conserved residues are indicated with a black background. The four conserved motifs RRx$_8$W, (M/L)L(S/Q/N)L(F/Y)EAS, DDxxD and NSE/DTE characteristic of terpene synthases are underlined, with the amino acids present in TalbCin shown in bold. Other

conserved motifs RWW, YMQ and CYMxE and specific absolutely conserved amino acids in terpene synthases of *Tps-b* subfamily are also highlighted below and above the MSA, respectively.

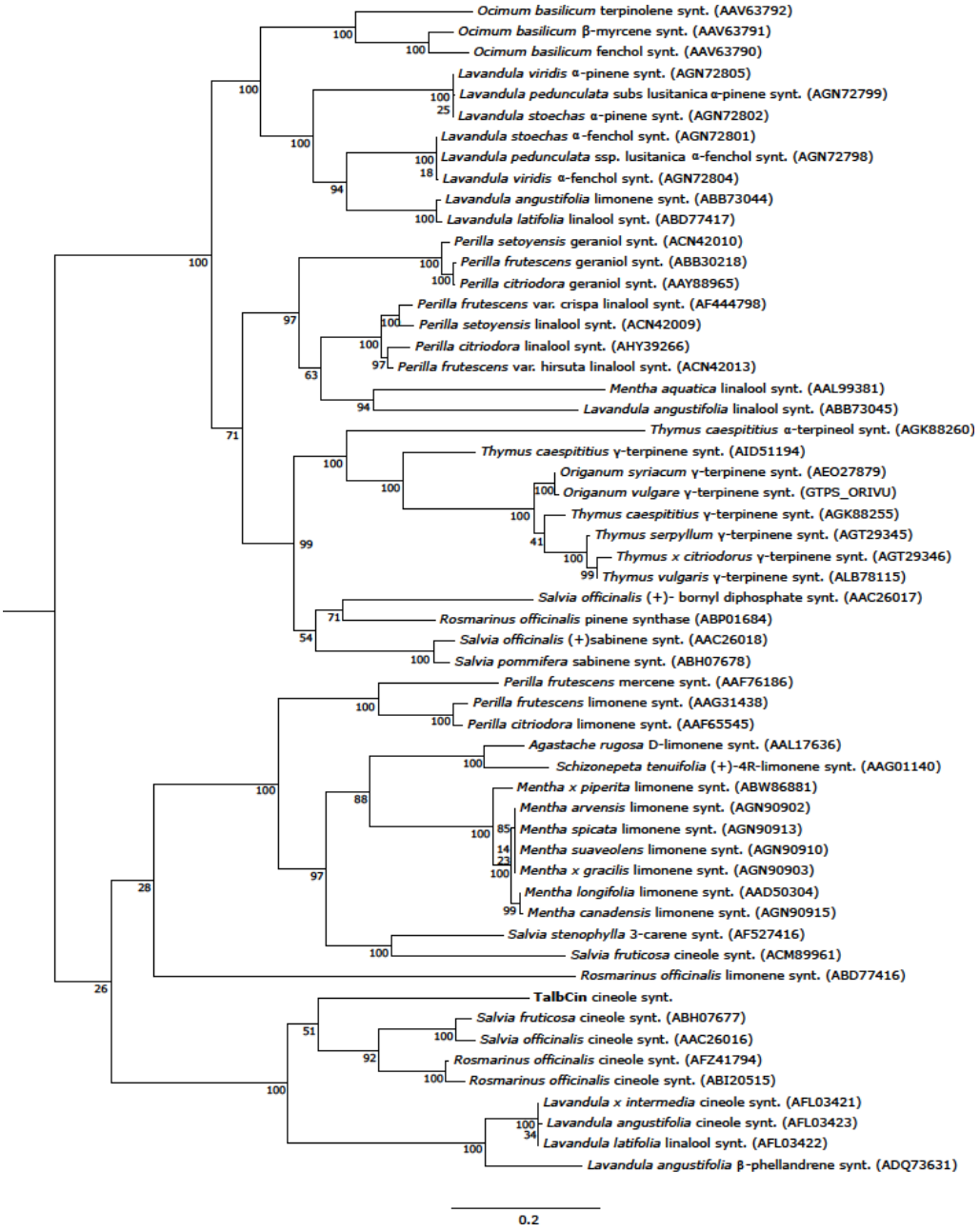

Fig. 2. Maximum likelihood evolutionary phylogenetic tree of the TalbCin with monoterpene synthases of other plants from Lamiaceae family. TalbCin is highlighted in bold and the plant species names are followed by the characterized terpene synthase name. The GenBank protein accession number of each gene is given in brackets. The tree was rooted with the 9-*epi*-caryophyllene sesquiterpene synthase from *Lavandula* x *intermedia* (AGU13712) (not displayed on the tree). The tree was constructed using the deduced protein sequence from the RRx_8W motif to the stop codon. The amino acid sequences were

aligned using ClustalW software. A JTT substitution model was selected assuming an estimated proportion of invariant sites (0.072) and gamma shape of 2.042. The reliability of internal branches was assessed using the bootstrapping method (100 bootstrap replicates). The tree was constructed using the ATGC Montpellier Bioinformatics platform. Graphical representation and editing of the phylogenetic tree was performed with FigTree v.1.4.2 and the Inkscape 0.91 program.

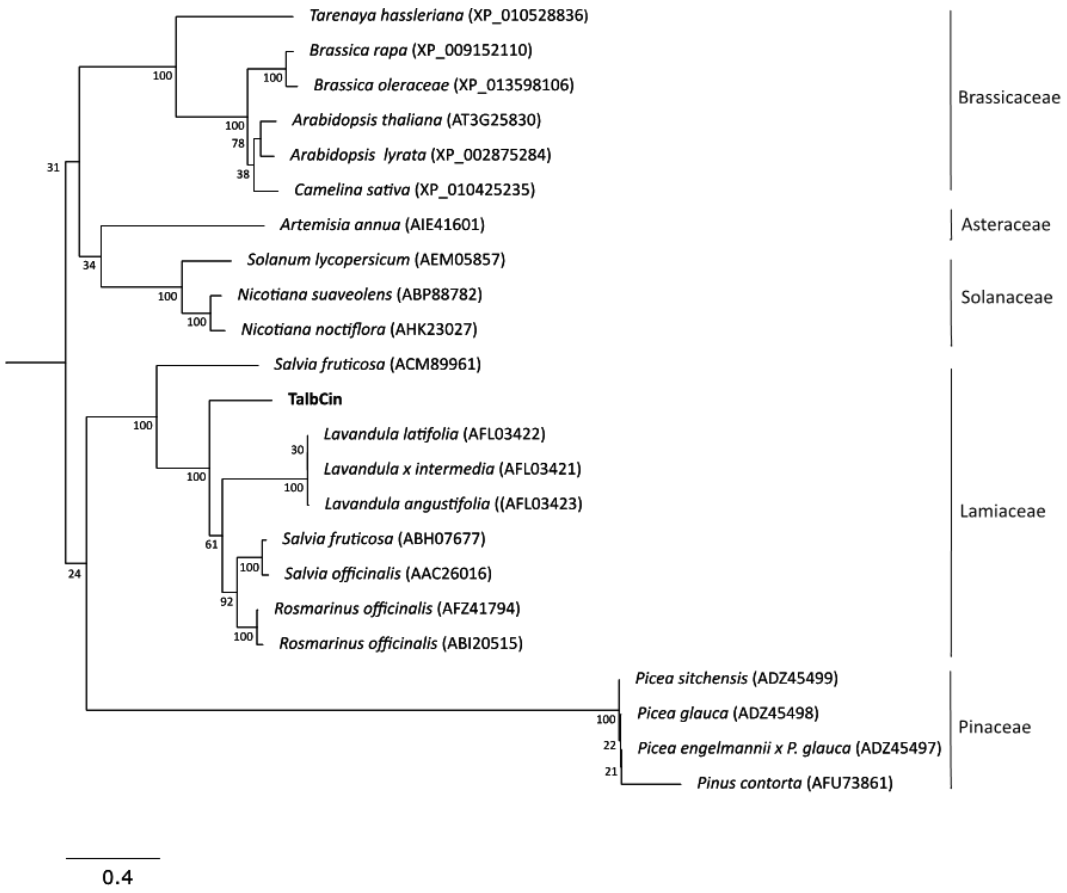

Fig. 3. Phylogenetic analysis of the TalbCin with 1,8-cineole synthases from several plant families, retrieved from GenBank. Not all of the 1,8-cineole amino acid sequences in GenBank have been functionally characterized. The phylogenetic tree was constructed using the Maximum likelihood estimation. TalbCin is highlighted in bold and the plant species names are followed by the GenBank protein accession number given in brackets. The tree was rooted with the 9-*epi*-caryophyllene sesquiterpene synthase from *Lavandula* x *intermedia* (AGU13712) (not displayed on the tree). The tree was constructed using the deduced protein sequence from the RRx_8W motif to the stop codon. The amino acid sequences were aligned using the ClustalW software. An LG substitution model was selected assuming an estimated proportion of invariant sites (0.044) and gamma shape of 1.667. The reliability of the internal branches was assessed using the bootstrapping method (100 bootstrap replicates). The tree was constructed using the ATGC Montpellier

Bioinformatics platform. Graphical representation and editing of the phylogenetic tree was performed with FigTree v.1.4.2 and the Inkscape 0.91 program.

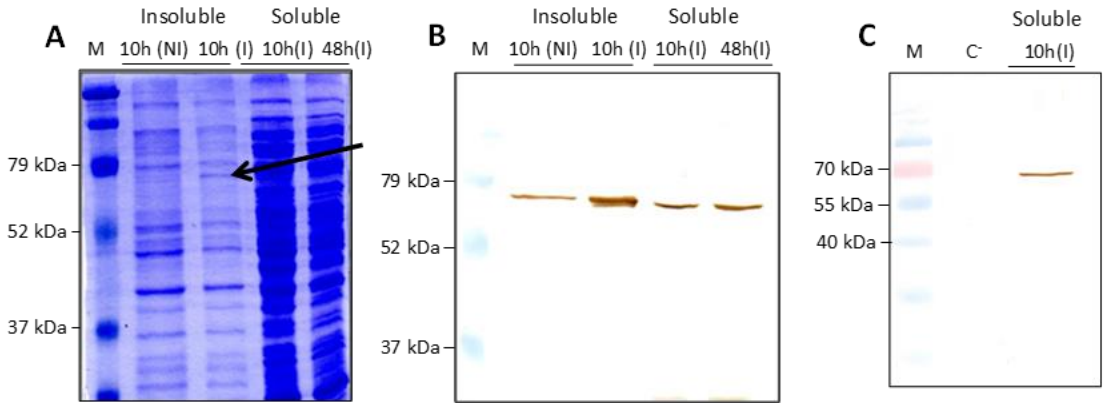

Fig. 4. SDS-PAGE and Western blot analysis of bacterial lysates of *Escherichia coli* BL21-CodonPlus(DE3) following expression of the recombinant fusion 6xHis-tag TalbCin protein. **A**. Coomassie blue-stained SDS-PAGE (12%) gel showing the protein profile of the insoluble and soluble fractions of bacterial extracts after expression of the recombinant TalbCin without addition of IPTG (non-induced, NI) for 10h of expression and with IPTG induction (I), during 10 h or 48 h, at 20 ºC; The arrow indicates the recombinant TalbCin with a Mw of ≅ 72.8 kDa. **B**. Western blot of the gel on A using antibody (1/3000) raised against 6xHis tag, revealing a immunoreactive band of the expected size (~72.8 kDa). **C**. Western blot analysis of the soluble fraction from bacterial extracts transformed with the control vector pCR®T7/NT-TOPO® (C^-) or the recombinant pCR®T7/NT-TOPO®/*TalbCin* construct, IPTG induced for 10 h. Note the absence of an immunoreactive band in the bacteria transformed with the control vector without insert. M: Molecular weight marker (kDa) as indicated. M: Molecular weight marker (kDa) as indicated.

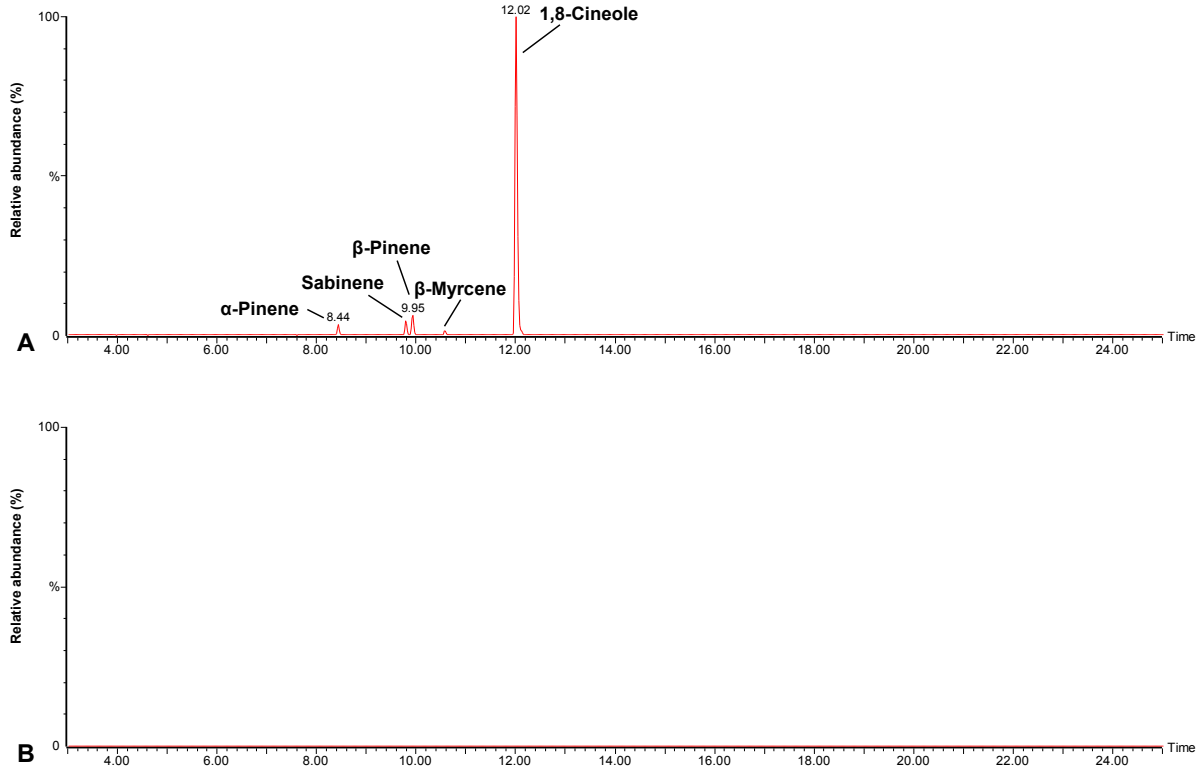

Fig. 5. SPME-GC-MS chromatograms of products obtained after the assays with TalbCin in the presence of the substrate GPP (A) and FPP (B).

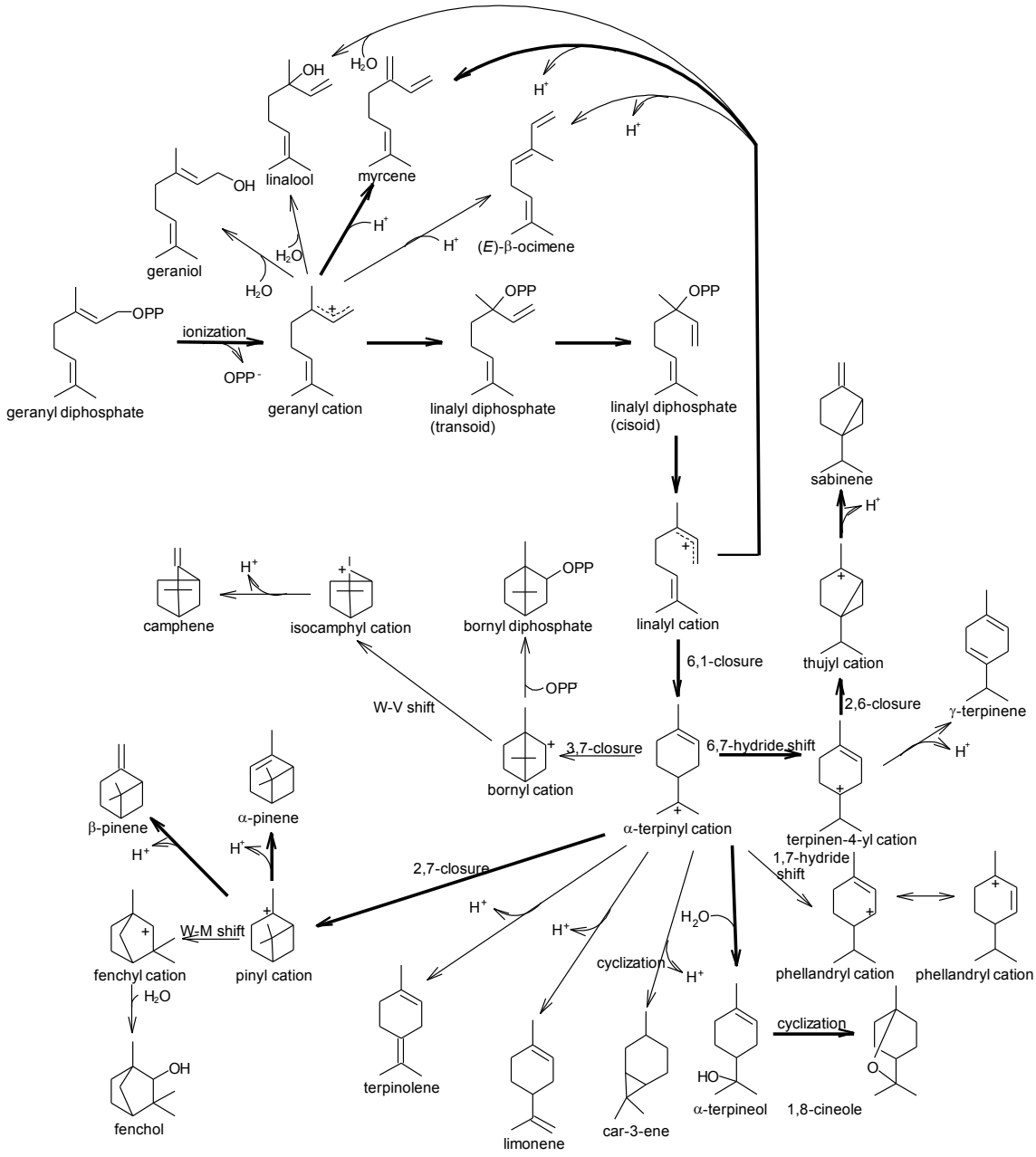

Fig. 6. Reaction mechanism of mTPS from ionization of GPP as substrate. The multi-product formation of the acyclic monoterpenes may proceed via either the geranyl cation or via the linalyl cation pathways. Possible reaction mechanisms of 1,8-cineole synthase of *Thymus albicans* are highlighted with bold arrows. Adapted from Bohlmann et al. (1998) and Degenhardt et al. (2009).