Biochem. J. (2012) **448**, 261–271 (Printed in Great Britain) doi:10.1042/BJ20120654

Identification and characterization of a kunzeaol synthase from *Thapsia garganica*: implications for the biosynthesis of the pharmaceutical thapsigargin

Benjamin PICKEL*, Damian P. DREW†, Tom MANCZAK†, Corinna WEITZEL†, Henrik T. SIMONSEN†[1] and Dae-Kyun RO*[1]

*Department of Biological Sciences, University of Calgary, Calgary, Canada, T2N 1N4, and †Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, 1871 Frederiksberg, Denmark

Thapsigargin is a major terpenoid constituent of *Thapsia garganica* root. Owing to its potent antagonistic effect on the sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase, thapsigargin has been widely used to study Ca^{2+} signalling and is also a potential drug for prostate cancer. Despite its importance, thapsigargin biosynthesis in *T. garganica* remains unknown. In order to decipher thapsigargin biosynthesis, deep transcript sequencing (454 and Illumina) of the *T. garganica* root was performed, and two terpene synthases (*TgTPS1/2*) were identified. Functional characterization of their encoded enzymes in a metabolically engineered yeast revealed that *TgTPS1* synthesized δ-cadinene, whereas *TgTPS2* produced ten distinct terpenoids. However, cultivation of the *TgTPS2*-expressing yeast in pH-maintained conditions (pH 6–7) yielded one major oxygenated sesquiterpenoid, suggesting that formation of multiple terpenoids was caused by acidity. The major terpene product from *TgTPS2* was identified as 6β-hydroxygermacra-1(10),4-diene (kunzeaol) by mass-fragmentation pattern, retention index, the nature of its acid-induced degradation and NMR. Also, recombinant TgTPS2 efficiently catalysed the synthesis of kunzeaol *in vitro* from farnesyl diphosphate with a K_m of 2.6 μM and a k_{cat} of 0.03 s^{-1}. The present paper is the first report of a kunzeaol synthase, and a mechanism for the transformation of kunzeaol into the thapsigargin backbone is proposed.

Key words: genomics, 6-hydroxygermacra-1(10),4-diene, kunzeaol, sesquiterpene synthase, *Thapsia garganica*, thapsigargin.

INTRODUCTION

Terpenoids are the largest class of natural products with more than 40000 known structures [1]. Some terpenoids are key constituents for cellular functions (e.g. electron transport chains, steroidal membrane components and hormones), but most of them are specialized products that influence the fitness of the synthesizing organisms in competitive ecosystems. Although microbes and fungi can produce terpenoids, the vast majority of specialized terpenoids reported originate from plants. Plant terpenoids are known to serve as insect and herbivore deterrents, allelochemicals, and as attractants for pollinators and beneficial insects in tritrophic interactions [2,3]. Aside from the eco-physiological roles of terpenoids in nature, they also have benefited humankind as flavours, fragrances, pharmaceuticals, nutraceuticals and industrial chemicals [4].

The wide range of physiological activities and industrial applications exhibited by terpenoids can be attributed to their structural diversity. Central to terpenoid diversity is the catalytic plasticity of TPSs (terpene synthases), which can convert simple linear prenyldiphosphates into complex terpenes, often with cyclic groups and a number of chiral centres. By co-ordination with divalent metal ions (e.g. Mg^{2+}), TPSs cleave the diphosphate moiety of a linear prenyldiphosphate (i.e. C-10 geranyl-, C-15 farnesyl- or C-20 geranylgeranyl-diphosphate) to yield highly reactive carbocations, which can undergo a cascade of reactions, such as allylic rearrangement, hydride- and methyl- shifts, and ring contraction [5]. The carbocation reactions are terminated by either deprotonation or the addition of an hydroxy group, resulting in the formation of the final products. The specific sequence and direction of these carbocation reactions are dictated by the contour of the substrate-binding site in TPS enzymes [6], and it has been shown that mutations in critical residues of the substrate-binding pocket can induce TPSs to synthesize different terpene backbones [6,7].

Although many terpenoids have been used as traditional medicines without knowing their mode-of-action, specific cellular targets of several terpenoid ligands have been identified. The sesquiterpene lactone derivative thapsigargin (Figure 1) was first isolated from *Thapsia garganica* (Apiaceae) roots [8], and has been shown to bind to the SERCA (sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase), thereby blocking its activity. SERCA sequesters cytosolic Ca^{2+} into the lumen of the sarcoplasmic/endoplasmic reticulum [9,10]. Thus blocking SERCA by thapsigargin increases the cytosolic Ca^{2+} level, which results in disrupted cell growth [11–13] and apoptosis of various types of mammalian cells [14–16]. Thapsigargin exhibits affinity at nanomolar levels to various SERCA isoforms [9,10,17] that are ubiquitously present in all mammalian cells [18]. Since the initial discovery of its physiological activity, thapsigargin has been widely used to investigate Ca^{2+} signalling in mammalian cells [19]. Intriguingly, the use of thapsigargin to cure prostate cancer has also been actively pursued [20], and a pro-drug utilizing thapsigargin is currently in Phase 1 clinical trials for the treatment of tumours [21].

Structurally, thapsigargin belongs to the sesquiterpenoid subfamily of 6,12-guaianolides (Figure 1), a subgroup of STLs (sesquiterpene lactones) with a guaiene-type (octahydroazulene) backbone. Depending on the lactone-ring position, guaianolides can have either C-6–C-12 or C-8–C-12 conjugation, and are

Abbreviations used: FPP, farnesyl diphosphate; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy; ORF, open reading frame; RI, retention index; SC, synthetic complete; SERCA, sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase; sesqui-TPS, sesquiterpene synthase; STL, sesquiterpene lactone; TPS, terpene synthase.

[1] Correspondence may be addressed to either of these authors (email hts@life.ku.dk or daekyun.ro@ucalgary.ca).

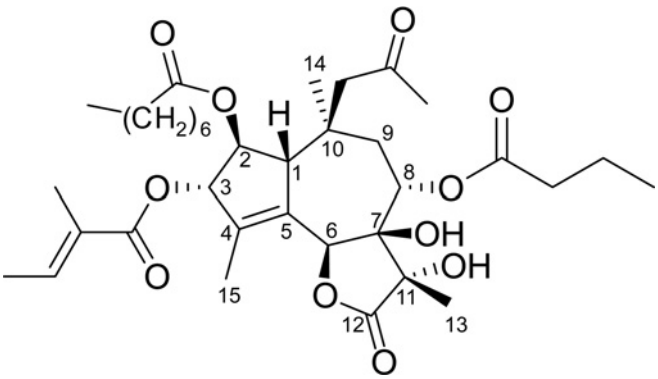

Figure 1 Structure of thapsigargin

generally referred to as 6,12- or 8,12-guaianolides respectively [21]. To date, the mechanism of guaianolide biosynthesis remains incomplete. Biochemical studies of lettuce and the related species chicory have shown that the STL costunolide is first synthesized from FPP (farnesyl diphosphate) by co-ordinated reactions of a sesqui-TPS (sesquiterpene synthase; germacrene A synthase) and two cytochrome P450 monooxygenases (P450s; germacrene A oxidase and costunolide synthase) [22–24]. In chicory, cell-free assays suggested that the transformation of costunolide into a guaianolide backbone is carried out by as yet uncharacterized enzymes [25]. On the other hand, a sesqui-TPS from agarwood can convert FPP directly into a guaiene skeleton. Therefore it appears that biosynthesis of guaianolide backbones can be formed either by transformation of FPP directly into guaiene and subsequent lactone formation, or by skeletal rearrangement of a germacrenolide-type STL, such as costunolide [21]. So far, little is known about the actual biosynthetic mechanism of guaianolides in *Thapsia* and related families.

Although the total chemical synthesis of thapsigargin has been achieved [26], the supply of thapsigargin or its derivatives by synthetic approaches will not be economically viable. Considering the potential use of thapsigargin in medical applications, a cost-effective supply of thapsigargin or its precursors by other means, such as microbial or plant metabolic engineering, will be critical. To accomplish this, the enzymes responsible for the early steps of thapsigargin biosynthesis need to be elucidated.

In the present study we applied a transcriptomics approach to identify two *TPS* genes from *T. garganica*. Biochemical characterization of one of the encoded enzymes (TgTPS2) showed that the terpenoid synthesized from this enzyme is 6β-hydroxygermacra-1(10),4-diene (kunzeaol), and we propose that this unstable compound can be further transformed into the thapsigargin backbone.

EXPERIMENTAL

General procedures

Chemicals were purchased from Sigma–Aldrich and components of yeast media were from Difco.

Plant material

T. garganica roots were collected in July/August 2010, 25 km SSW of Bari, Italy (GPS 40.898625, 16.706139). Living plant material was used for RNA extraction, whereas air-dried tissue was used for chemical extraction. A voucher specimen (HTS 2010-13) has been deposit at the KU-LIFE herbarium (Copenhagen University, Copenhagen, Denmark).

Thapsigargin extraction and HPLC analysis

Extraction and analysis of thapsigargin was performed as described by Jäger et al. [27] with slight modifications. Homogenized root tissue was extracted with 5 ml of ethyl acetate per 100 mg of dry material for 1 h using a sonication bath. Extracts were evaporated to dryness and reconstituted to 1 mg/ml in 83 % (v/v) methanol. HPLC analysis was performed on a Fusion RP column (2.5 μm, 100 mm × 2 mm; Phenomenex) using 70 % (v/v) methanol as the mobile phase with a flow rate of 0.2 ml/min. Eluting compounds were detected at 230 nm. A thapsigargin standard was used as a reference for identification and quantification (donated by Søren Brøgger Christensen, Copenhagen University, Copenhagen, Denmark).

cDNA library construction and sequencing

Total RNA was isolated from *T. garganica* roots using the CTAB method [28]. cDNA was synthesized using SuperScript™ II Reverse Transcriptase (Invitrogen) according to the manufacturer's instruction with an oligo(dT)$_{12}$ primer. A total of 5 μg of cDNA was subjected to 454 sequencing using a Roche GS20 instrument. Total RNA was purified by the same method and was sent to the McGill Innovation Centre (Montréal, Canada) for library preparation, and Illumina sequencing was performed for the paired-end reads of 108 bp using a Genome Analyser (Illumina).

Expression of *TgTPS1* and *TgTPS2* in yeast

The ORFs (open reading frames) of *TgTPS1* and *TgTPS2* were PCR-amplified using high-fidelity polymerases (KAPA Biosystem). Primers used are shown in Supplementary Table S1 (at http://www.BiochemJ.org/bj/448/bj4480261add.htm). The gel-purified PCR amplicons were digested with BamHI and XhoI and ligated to the respective site in pESC-Leu2d. *Saccharomyces cerevisiae* EPY300 strain [34] was transformed with the pESC-Leu2d plasmid expressing either *TgTPS1* or *TgTPS2* using the lithium acetate method, and transformants were selected on His −, Leu −, Met − dropout SC (synthetic complete) medium containing 2 % (w/v) glucose at 30 °C. Individual colonies were cultured overnight in 2 ml of His −, Leu −, Met − dropout SC medium with 2 % (w/v) glucose, and the starter culture was diluted 100-fold into the His −, Leu −, Met − dropout SC medium with 1.8 % (w/v) galactose and 0.2 % glucose. The culture was overlaid with 10 % (v/v) dodecane and 1 mM methionine was added. The yeast cultures were grown at 30 °C and 180 rev./min. For metabolic profiling, aliquots of the dodecane overlays were diluted (1:100) in hexane and 1 μl was analysed by GC-MS as described below.

Heterologous expression and purification of TgTPS2 in *Escherichia coli*

For creation of a N-terminal His$_6$-tagged protein, the ORF of *TgTPS2* was PCR-amplified with the primers *His$_6$-TgTPS2 forward/reverse* (Supplementary Table S1), and the amplified product was cloned into pDONR221 using the BP reaction according to the gateway cloning manual (Invitrogen). Subsequently, *TgTPS2* was cloned into pDEST17 using the LR

gateway cloning reaction, thereby introducing an N-terminal His_6 tag. The resulting expression plasmid was transformed into *E. coli* BL21-AI cells (Invitrogen). For heterologous expression, recombinant bacteria were grown in LB (Luria–Bertani) medium at 37 °C to a D_{600nm} of 0.7 and transferred to 15 °C. Expression was induced by the addition of 0.1 % arabinose. Cells were harvested after 20 h and resuspended in 50 mM Tris/HCl (pH 7.5), 0.2 M NaCl, 5 mM imidazole and 1 mM PMSF. After lysis by sonication, the cleared supernatant was passed through a Bio-Scale Mini Profinity IMAC column (1 ml) on an FPLC system (both Bio-Rad Laboratories). Bound proteins were eluted by gradually increasing the imidazole concentration in lysis buffer from 5 to 250 mM over 10 column volumes. Collected fractions were screened for TgTPS2 by SDS/PAGE. The TgTPS2-containing fractions were pooled and desalted with 50 mM Tris/HCl (pH 7.5) and 10 % (v/v) glycerol using Amicon® Ultra concentrators [Millipore, MWCO (molecular mass cut-off) 10 kDa]. The protein concentration was determined using the Bradford assay.

Biochemical characterization of TgTPS2

For product analysis by GC-MS, enzyme assays were performed in 50 mM Tris/HCl (pH 7.5), 50 mM $MgCl_2$ containing 100 μM FPP and 30 μg of TgTPS2 in a total volume of 0.5 ml. The reaction mixture was overlaid with 1 ml of hexane and incubated for 1 h at 28 °C. Synthesized terpenoids from the enzyme reaction mixtures were partitioned to hexane by vigorous mixing. The hexane layer was transferred to a new vial and concentrated to approximately 200 μl by gentle nitrogen flow. Aliquots (1 μl) were analysed by GC-MS. Enzyme reactions for steady-state kinetics were performed as described above in a reaction volume of 100 μl. The assays contained 0.5 μg of TgTPS2 and 0.5–50 μM FPP spiked with a fixed ratio of [1-^{3}H] FPP (PerkinElmer) to a final specific activity of 0.042 Ci mmol^{-1}. After 10 min, the reaction was stopped by adding EDTA and sodium hydroxide to final concentrations of 0.5 M and 2 M respectively. Products were partitioned to hexane, and radioactivity was measured with a liquid scintillation counter (LS6500, Beckman Coulter). Triplicate assays were carried out for each substrate concentration. To determine the apparent K_m and k_{cat} values, data points obtained were fitted to the Michaelis–Menten model by non-linear regression using R2.13.0 [29].

GC-MS analysis

All GC-MS analyses were carried out on a 6890N gas chromatograph coupled to a 5975B mass spectrometer (Agilent). A sample volume of 1 μl was injected at an inlet temperature of 250 °C, and compounds were separated on a DB-1 MS column [30 m × 250 μm i.d. (internal diameter) × 0.25 μm film thickness, Agilent] using helium as the carrier gas with a constant flow rate of 1 ml/min. The temperature programme consisted of a 1 min hold at 80 °C and a linear gradient of 10 °C min^{-1} to 250 °C. Fragmentation was achieved by electric ionization at 70 eV, and fragmented ions were scanned from *m/z* 50–400. Recorded mass spectra were analysed using the Wiley7/NIST05 library. RIs (retention indices) were determined with alkane standards (C-8–C-40) and calculated using AMDIS32 [30].

Purification of metabolites from yeast culture and NMR analysis

For isolation of kunzeaol, yeast cultures were cultivated in the presence of 0.1 M Hepes (pH 7.5) for 64 h. Cells were harvested by centrifugation at 4000 ***g*** for 5 min and extracted with 50 % (v/v) acetonitrile. The acetonitrile extract was partitioned three times to an equal volume of pentane. The combined pentane extracts were concentrated by a constant nitrogen stream and separated on silica gel 60 (Merck) using increasing concentrations of ethyl acetate in hexane. Fractions were screened by GC-MS. Kunzeaol-containing fractions were pooled and evaporated to dryness by a nitrogen stream. NMR measurements were conducted on an Ultrashield Plus 600 MHz spectrometer (Bruker) in deuterated chloroform at − 20 °C.

In silico analysis of *T. garganica* root transcripts

For quantitative analysis of *T. garganica* root transcripts, the readings of the Illumina sequencing were trimmed for quality, and all reads <21 bp were removed using CLC Genomics Workbench 5.0.1 (CLC Bio). The protein sequences of HMG-CoA (3-hydroxy-3-methylglutaryl-CoA) synthase (AAD00298), isopentenyl-PP-isomerase (AED92292), mevalonate 5-phosphate decarboxylase (AEC09574), actin 2 (AEE76148), ubiquitin 10 (AAA68878), elongation factor 1α (BAC23049), pyruvate dehydrogenase (2019230A), aldolase (AEE85215) and enolase (BAD94751) were used to identify orthologous full-length ORFs within the Velvet assembly of the *T. garganica* Illumina sequencing data using a BLAST search [31]. For *TgTPS1* and *TgTPS2*, the experimentally determined nucleotide sequences were used. The trimmed Illumina reads were mapped to the ORFs using the RNA-seq analysis tool of the CLC Genomics Workbench applying a minimal similarity fraction of 99 %. Estimation of the expression levels were determined by comparison of the numbers of uniquely mapping reads and total mapping reads after normalization by ORF length, similar to RPKM values [32].

RESULTS

Thapsigargin production in *T. garganica* roots

Thapsigargin in root tissue of *T. garganica* plants collected near Bari (Italy) accounted for 0.44 % of dry weight as determined by HPLC analysis. This result is comparable with the amounts previously found in *T. garganica* roots from Libya, Greece, Spain and Italy [33] and indicates that the population investigated is an appropriate target for the elucidation of thapsigargin biosynthesis. Root material of *T. garganica* from the identical geographic location was used for RNA isolation and deep sequencing.

Identification of sesquiterpene synthases in *T. garganica*

454 sequencing (Roche GS20) of the cDNA synthesized from *T. garganica* root mRNA resulted in 293 434 reads averaging 201 nucleotides. Of these reads, 227 880 were assembled into 39 512 contigs with an average length of 361 nucleotides, and the remaining reads were designated as singletons. A BLAST search of the compiled unigenes using the sequence of *Lycopersicon hirsutum* germacrene B synthase (accession number AAG41891) as a query identified six contigs exhibiting similarity to TPS sequences, but no full-length TPS sequences were identified within the contig set. Subsequent BLAST searches using different sesqui-TPS sequences (e.g. amorpha-4,11-diene synthase, AAF61439; 5-*epi*-aristolochene synthase, AAG17667; and germacrene A synthase; AAM11626) did not uncover further potential sesqui-TPS sequences.

When aligned to the characterized sesqui-TPS sequences, contigs 2580 (785 bp) and 16 086 (256 bp) aligned to the 5′-end of the genes and included nucleotides corresponding to putative start codons. Similarly, contigs 2586 (903 bp) and 313 (753 bp) corresponded to the 3′-end of sesqui-TPS coding sequences and included putative stop codons, whereas contigs 7300 (405 bp) and 25 893 (234 bp) corresponded to internal coding sequences. Primers designed around the putative start codons of contigs 2580 and 16 086 (*c2580 forward* and *c16086 forward*; Supplementary Table S1) differed only by a single nucleotide, whereas unique primers could be designed after the putative stop codons (*c2586 reverse* and *c313 reverse*; Supplementary Table S1) in the 3′-untranslated regions where sequence conservation was lower. Combinatorial uses of these primers (*c2580 forward* with *c313 reverse* and *c16086 forward* with *c2586 reverse*) enabled us to retrieve two full-length *TPS* cDNAs by PCR. These were designated as *TgTPS1* (JQ290344) and *TgTPS2* (JQ290345). Furthermore, the sequences of the short internal contigs 7300 and 25 893 were identical with *TgTPS1* and *TgTPS2* respectively, indicating that all six contigs from our 454 sequencing can be assembled into two sesqui-TPS cDNAs. The ORF of *TgTPS1* and *TgTPS2* encoded for proteins with molecular masses of 64.3 and 64.7 kDa, which shared 89 % sequence identity and 94 % similarity to each other in their amino acid sequences (Supplementary Figure S1 at http://www.BiochemJ.org/bj/448/bj4480261add.htm). The absence of chloroplast-targeting sequences is indicative of typical cytosolic sesqui-TPSs.

In vivo characterization of TgTPS1 and TgTPS2

The *Saccharomyces cerevisiae* strain EPY300, metabolically engineered to overproduce FPP [34,35], was used initially to assess the activities of enzymes encoded by *TgTPS1* and *TgTPS2*. The ORFs of the two cDNAs were cloned in pESC-Leu2d under control of a galactose-inducible promoter and transformed into EPY300. Biologically synthesized lipophilic terpenoids were trapped by a dodecane overlay on the cultures and analysed by GC-MS after a cultivation time of 72 h.

In the dodecane layer of the *TgTPS1*-expressing yeast, two major peaks were detected (Figure 2A). The major compound with a retention time of 11.17 min (peak f) also accumulated in the control yeast transformed with the empty vector (Figure 2C). This was identified as the FPP-derivative *trans*-nerolidol, on the basis of its RI and mass spectra (Table 1 and Supplementary Figure S2 at http://www.BiochemJ.org/bj/448/bj4480261add.htm). The second major compound with a retention time of 10.82 min (peak d) exhibited an RI and fragmentation pattern as described for δ-cadinene. This finding was confirmed by an authentic (+)-δ-cadinene standard, synthesized by recombinant (+)-δ-cadinene synthase from *Gossypium arboreum* in EPY300 [36] (Figure 2B). Therefore TgTPS1 exhibits mono-product δ-cadinene synthase activity.

In contrast with *TgTPS1*, the dodecane overlay from *TgTPS2*-expressing yeast showed the occurrence of ten different detectable compounds (Figure 2D). Aside from *trans*-nerolidol (peak f) and δ-cadinene (peak d), the compounds with retention times of 10.37, 10.56, 11.43, 12.14 and 12.27 min respectively, could be identified as germacrene D (peak a), bicyclogermacrene (peak b), 1,6-germacradien-5-ol (peak g), τ-muurolol (peak h) and α-cadinol (peak i) according to their RI values and mass spectra (Table 1 and Supplementary Figure S2). For the minor compounds with retention times of 10.84 (peak e) and 12.73 min (peak j), no hits were found in the MS database.

Interestingly, the quantitative metabolic composition of *TgTPS2*-expressing yeast changed over time. Although the later-eluting unknown compound (peak j) was most abundant at 40 h and 64 h, the corresponding peak had decreased dramatically at 84 h, accompanied by a marked increase of δ-cadinene (peak d), *trans*-nerolidol (peak f), τ-muurolol (peak h) and α-cadinol (peak i) (Figures 3A and 3B). Only marginal increases of germacrene D (peak a), bicyclogermacrene (peak b), two unknown compounds (peak c and e) and 1,6-germacradien-5-ol (peak g) were detected at 84 h. This disproportional change of terpene products could not be easily explained by a multi-product promiscuous activity of TgTPS2. In the case of a multi-product enzyme, one would expect a steady increase of compounds over time, and a previous experiment showed that terpene production in EPY300 rapidly increased in the first 72 h, but reached a plateau after 72 h cultivation [34]. This pattern of terpenoid accumulation was found for the synthesis of germacrene D (peak a), bicyclogermacrene (peak b), an unknown compound (peak e) and 1,6-germacradien-5-ol (peak g). Owing to low abundance and signal overlap, an accurate quantification of peak c could not be achieved.

Since the pH of yeast cultures becomes acidic over time, a lower pH value might influence the product profile at later stages of cultivation. To test whether the formation of δ-cadinene, τ-muurolol and α-cadinol (peaks d, h and i) was caused by

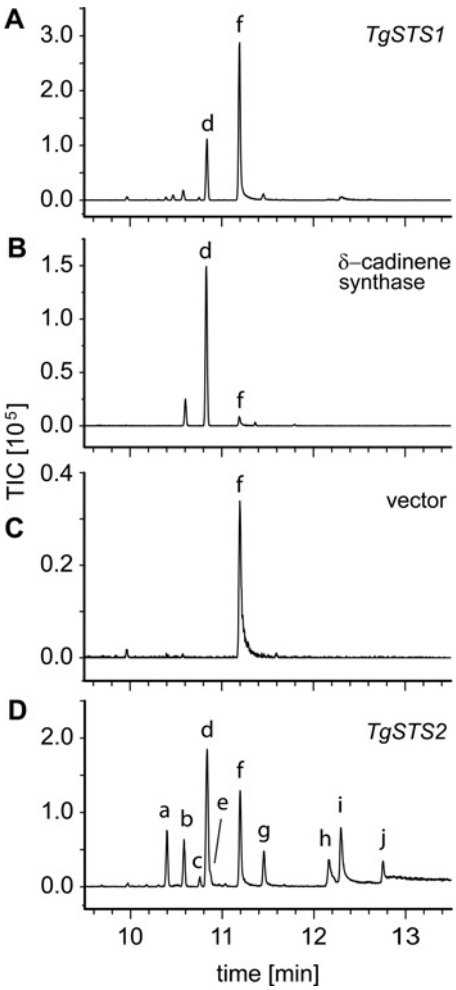

Figure 2 GC-MS analysis of the dodecane overlay of *TgTPS1*- (A), δ-cadinene synthase from *Gossypium arboreteum*- (B) and an empty vector- (C) expressing yeast culture, and a yeast strain expressing *TgTPS2* (D)

For peak annotations, see Table 1.

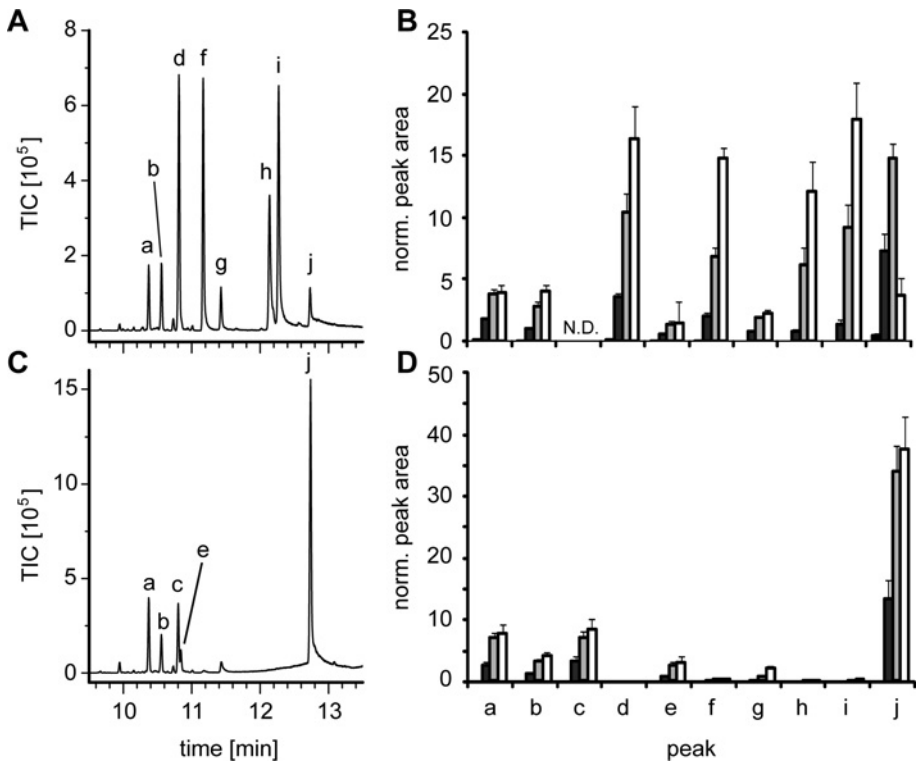

Figure 3 Influence of pH on the metabolic profile of the *TgTPS2*-expressing yeast

The metabolic profile of *TgTPS2*-expressing yeast was compared by GC-MS under normal conditions (**A**) and in the presence of 0.1 M Hepes (pH 7.5) (**C**) at 84 h cultivation. The relative amount of the corresponding compounds is shown after normalization by an internal standard (tetradecane) at different time points (16 h black, 40 h light grey, 64 h dark grey, 84 h white) in the control (**B**) and buffered (**D**) culture. N.D., not determined due to an overlapping peak.

Table 1 Retention times (RT) of compounds detected in the dodecane overlay of the *TgTPS2*-expressing yeast, which were identified by MS database search and comparison of the experimentally determined retention indices (RI_{exp}) with the those from the MassFinder 4 Terpenoid Library (RI_{lib})

N.A., not applicable.

Peak	RT (min)	RI_{exp}	Substance	RI_{lib}
a	10.37	1486	Germacrene D	1479
b	10.56	1501	Bicyclogermacrene	1494
c	10.81	1521	Unknown 1	N.A.
d	10.82	1522	δ-Cadinene	1520
e	10.84	1524	Unknown 2	N.A.
f	11.17	1551	*trans*-Nerolidol	1553
g	11.43	1572	1,6-Germacradien-5-ol	1571
h	12.14	1633	τ-Muurolol	1633
i	12.27	1645	α-Cadinol	1643
j	12.73	1686	6-Hydroxygermacra-1(10),4-diene (or kunzeaol)	1687

low pH, the *TgTPS2*-expressing yeast was cultivated in medium buffered with 0.1 M Hepes (pH 7.5). The supernatant in buffered cultures had a pH of 6 after 84 h. In contrast, non-buffered yeast cultures were greatly acidified and a pH of 3 was observed after the same cultivation time. The pH-adjustment of the medium resulted in a dramatic change of the metabolic profile (Figure 3C). Although the compound with a retention time of 12.73 min (peak j) was the most abundant compound at all time points (Figure 3D), formation of δ-cadinene, *trans*-nerolidol, τ-muurolol and α-cadinol (peaks d, f, h and i respectively) was almost completely suppressed. Germacrene D, bicyclogermacrene, 1,6-germacradien-5-ol (peaks a, b and g) and two unknown compounds (peaks c and e) steadily increased in buffered cultures. The compound with a retention time of 10.81 min (peak c), which could not be clearly distinguished in the unbuffered medium because of its low abundance and a very similar retention behaviour as δ-cadinene, was also detected in the buffered medium.

The differences of the metabolic profile under unbuffered and buffered culture conditions (Figures 3A and 3C) clearly indicate that the unknown compound (peak j) is the major product of TgTPS2. Germacrene D (peak a), bicyclogermacrene (peak b), 1,6-germacradien-5-ol (peak g) and both unknown compounds (peaks c and e) are likely to be secondary products from promiscuous TgTPS2 activity, but may be considered as minor products due to their low abundance. According to their absence in buffered cultures, formation of δ-cadinene (peak d), τ-muurolol (peak h) and α-cadinol (peak i) seems to occur by acid-induced rearrangements of the compound corresponding to peak j. This finding is supported by the accumulation pattern of the compounds in unbuffered medium. The extenuated increase of the unknown compound (peak j) up to 64 h and the following decrease correlate with an increase of the compounds generated by acidic influence (Figure 3B).

Identification of the major TgTPS2 product

The mass spectrum of the major terpene product from TgTPS2 showed a parent ion of 222 typical of oxygenated sesquiterpenes and major fragment ions with *m*/*z* values of 84 (100), 81 (65), 55 (53), 67 (38), 109 (32), 93 (28), 121 (25) and 161 (21) (Figure 4). The dominant fragment ion with an *m*/*z* of 84 is unusual

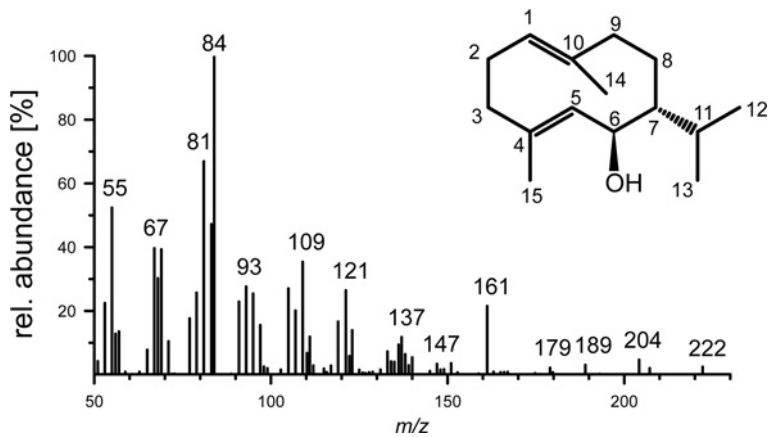

Figure 4 Mass spectra and structure of 6β-hydroxygermacra-1(10),4-diene

in sesquiterpenoids, but has been described as characteristic for kunzeaol [6-hydroxygermacra-1(10),4-diene] [38,39]. Kunzeaol is a major terpenoid constituent from Australian native plants, *Kunzea spp.* [39], but had also been found in leaf trichomes of potato (*Solanum tuberosum*) [40] and *Santalonia rosmarinifolia* [38]. When the mass spectrum of the compound in peak j (Figure 4) was compared with the published spectra of kunzeaol, they showed essentially identical mass fragmenting patterns [38–40]. In addition, comparison of the experimentally determined RI with those from the terpene database MassFinder4 and the literature were in good agreement (Table 1) [39–42]. Intriguingly, kunzeaol is known to be unstable and decomposes to cadinene- and muurolene-type sesquiterpenoids under acidic conditions [39].

Further proof for the identity of 6-hydroxy-germacre-1(10),4-diene was obtained by NMR analysis (see Figure 4 for numbering). Owing to instability of this compound, purification of 6-hydroxygermacra-1(10),4-diene was achieved only to approximately 80 % as estimated by GC-MS. Characteristic downfield signals for the protons H-5 (5.09 and 4.98 p.p.m.), H-1 (4.94 p.p.m.) and H-6 (4.64 p.p.m.) were clearly detectable by ^{1}H-NMR. Also, the protons of the methyl groups on CH_3-12/13 (1.03 and 1.00 p.p.m., doublet), CH_3-14 (1.58 and 1.64 p.p.m., singlet) and CH_3-15 (1.47 p.p.m., singlet) showed clean signals. The remaining signals were broad multiplets. Interpretation of the multiplets was disrupted by doubling and overlapping of several ^{1}H-signals due to the ability of 6-hydroxygermacra-1(10),4-diene to form inter-convertible conformers [38], commonly found in many sesquiterpenoids with germacrene backbones [43]. The different conformers of 6-hydroxygermacra-1(10),4-diene also accounted for the >15 detected ^{13}C-signals, that could be precisely assigned to the signals published previously (Supplementary Table S2 at http://www.BiochemJ.org/bj/448/bj4480261add.htm) [38–40].

To elucidate the configuration at C-6 and C-7, a NOESY (nuclear Overhauser enhancement spectroscopy) spectrum was obtained. The NOESY spectrum, along with specific NOE (nuclear Overhauser effect) difference spectra, showed strong NOE connectivities in H-6–H-12/13, H-6–H-15 and H-7–H-5, along with a weaker signal for H-7–H-1. Calculation of the distances between these protons in minimal energy state revealed that such a NOE pattern would only be possible for a *trans*-configuration and only with a configuration of 6*R* and 7*R*, which is equivalent to a 6β-OH,7α-isopropanyl structure. This pattern also confirms the stereochemical conformation of the 1–10 and 4–5 double bonds as being *trans*. Importantly, such a conformation is also found in the majority of sesquiterpenoids isolated from Apiaceae, including thapsigargins from *Thapsia* species.

In summary, multiple lines of evidence, including mass-fragmenting patterns, RI, acid-induced instability and ^{13}C-spectral match, concluded that the major terpenoid product from TgTPS2 is 6-hydroxygermacra-1(10),4-diene. NOE data indicate that the molecule has the relative configuration of 6β-hydroxygermacra-1(10),4-diene.

In vitro biochemical characterization of TgTPS2

A δ-cadinene synthase has already been identified and characterized [36], whereas studies pertaining to kunzeaol synthase have not been reported to date. In order to further characterize TgTPS2, His-tagged TgTPS2 was expressed in *E. coli* and purified by Ni^{2+}-affinity chromatography. Although the majority of heterologously expressed His_6–TgTPS2 accumulated as insoluble inclusion bodies, 0.3 mg of soluble TgTPS2 could be purified from 1 litre of culture to apparent homogeneity (Supplementary Figure S3 at http://www.BiochemJ.org/bj/448/bj4480261add.htm). The protein size estimated by SDS/PAGE was approximately 65 kDa, which is in agreement with the expected molecular mass of 65.5 kDa. Authenticity of the purified protein was proven by *in vitro* activity in the presence of FPP and Mg^{2+}. GC-MS analysis of the TgTPS2 enzymatic products revealed that the heterologously expressed protein was capable of producing five terpenoids (Figure 5A). Among them, as shown in *in vivo* assays, 6β-hydroxygermacra-1(10),4-diene was the dominant terpene product. Furthermore, germacrene D, bicyclogermacrene and the two unknown compounds (peaks c and e) could also be detected as minor products. Therefore *in vitro* assays of recombinant TgTPS2 resulted in the same product spectrum as obtained from buffered yeast cultures or the early stages of unbuffered yeast cultures. These data substantiated that 6β-hydroxygermacra-1(10),4-diene is the major product of TgTPS2.

Boiled TgTPS2 showed no terpenoid formation *in vitro* (results not shown). Also, the addition of 100 mM EDTA completely abolished TgTPS2 activity (Figure 5B). The presence of divalent cation as an essential cofactor is characteristic for terpene synthases. Using purified enzyme, the kinetic parameters of TgTPS2 were obtained by fitting substrate titration data to the Michaelis–Menten model (Figure 5C). Recombinant TgTPS2 showed an apparent K_m for FPP of 2.6 ± 0.4 μM and a k_{cat} of 0.03 s^{-1}. These values lie in the range of characterized sesqui-TPSs [44].

Table 2 Estimated expression levels of *TgTPS1/2*, and the genes for HMG-CoA synthase (*HCS*), isopentenyl-PP-isomerase (*IPI*), mevalonate 5-phosphate decarboxylase (*MPD*), actin 2 (*Act2*), elongation factor 1α (*Ef1*), ubiquitin 10 (*Ubq10*), pyruvate dehydrogenase (*Pyde*), aldolase (*Aldo*) and enolase (*Enol*) in *T. garganica* roots by comparison of unique and total reads from the Illumina sequencing mapped to the corresponding ORFs and after normalization by cDNA length (RPB, i.e. read number per bp)

Number of unique reads, number of reads uniquely mapped (i.e. single hits) to the targeted gene, and therefore the reads mapped to more than one targets are excluded; Number of total reads, number of reads including both a unique map (i.e. a single hit) and multiple maps (i.e. more than a single hit) (a single read can be counted multiple times in this analysis). The marginal differences in the numbers of unique and total reads in *TgTPS1* and *TgTPS2* are due to the cross-hits of the small portion (< 1 %) of the reads.

Class of gene	Gene	ORF length (bp)	Number of unique reads	RPB_{unique}	Number of total reads	RPB_{total}
Sesquiterpene synthases	*TgTPS1*	1686	4573	2.71	4617	2.74
	TgTPS2	1686	15 796	9.37	15 955	9.46
Mevalonate pathway	*HCS*	1398	484	0.35	484	0.35
	IPI	903	499	0.55	499	0.55
	MPD	1263	1592	1.26	1592	1.26
Housekeeping genes	*Act2*	1134	8714	7.68	8714	7.68
	Ef1	1359	117	0.09	117	0.09
	Ubq10	918	400	0.44	400	0.44
Glycolysis	*Pyde*	1155	1232	1.07	1232	1.07
	Aldo	1074	2294	2.14	2294	2.14
	Enol	771	6082	7.89	6082	7.89

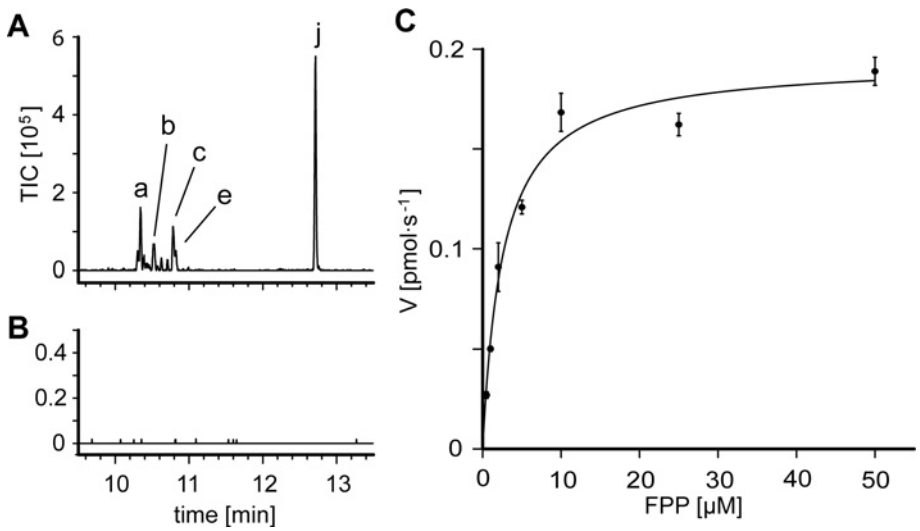

Figure 5 *In vitro* conversion of FPP by TgTPS2

GC-MS profile of hexane-extractable products in the presence of TgTPS2 (**A**) as well as TgTPS2 and EDTA (**B**). Steady-state kinetics of TgTPS2 are shown in (**C**).

Quantitative analysis of *TgTPS2* transcript abundance

At first sight, the monocyclic structure of 6-hydroxygermacra-1(10),4-diene does not appear to reflect the structural core of thapsigargin (i.e. guaiene), and therefore other unique guaiene synthases could be involved in thapsigargin biosynthesis in *T. garganica* roots. However, careful re-analysis of the initial 454 sequencing data did not reveal additional full or partial *TPS* transcripts other than those related to *TgTPS1/2*. Nevertheless, in order to obtain a deeper transcript profile, Illumina paired-end sequencing was used to acquire an additional $\sim 29 \times 10^6$ unique reads with a read-length of 108 bp from the *T. garganica* roots. This represented a 127-fold enrichment of read numbers over the initial 454 sequencing. Owing to the high sequence identity of *TgTPS1* and *TgTPS2* (94 % nucleotide identity) and short read-length of the Illumina sequencing, *de novo* assembly of the Illumina reads by the Velvet algorithm [45] generated a number of chimera contigs constituted with *TgTPS1* and *TgTPS2*. Nonetheless, no other distinct *TPS* transcripts could be found in the Illumina data by BLAST search using different sesqui-TPS sequences as the query.

The high content of thapsigargin implies that its biosynthetic transcripts are also likely to be abundant. Since the Illumina sequencing data can be reliably used to obtain a snapshot of transcript levels [46], quantitative mapping of trimmed Illumina reads was performed with high stringency (99 % identity) on the nucleotide sequences of *TgTPS1* and *TgTPS2*, as well as selected housekeeping genes and representative genes of glycolytic and mevalonate metabolism (Table 2). This *in silico* analysis demonstrated that the relative expression level of *TgTPS2* was equivalent to or significantly higher than the selected reference transcripts. The expression level of *TgTPS1* was estimated to be one-third of *TgTPS2*, but was still relatively high in comparison with other reference genes. Therefore quantitative and qualitative analysis of the Illumina sequencing data suggested that *TgTPS1* and *TgTPS2* are the dominant *TPS* transcripts in *T. garganica* roots. Furthermore, almost all reads mapped to both *TgTPS1* and *TgTPS2* were unique (i.e. single hits, Table 2), indicating

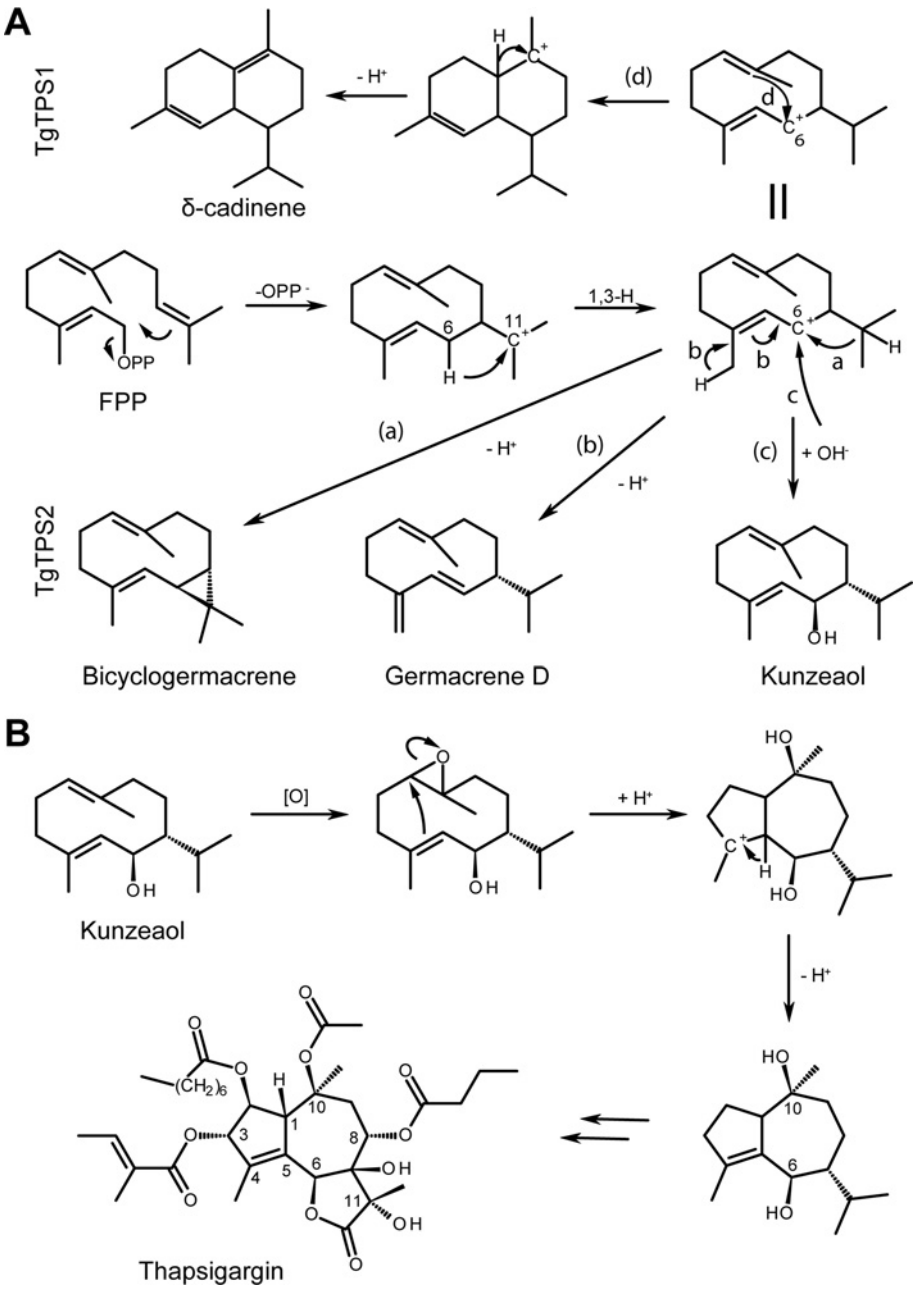

Figure 6 Proposed mechanisms of TgTPS1/2 reactions and the transformation of kunzeaol to thapsigargin precursor

Proposed mechanism for the formation of kunzeaol [6-hydroxygermacra-1(10),4-diene] and δ-cadinene from FPP catalysed by TgTPS1 and TgTPS2 (**A**). A proposed biosynthetic pathway to the guaianolide backbone of thapsigargin in *T. garganica*, that involves kunzeaol as an intermediate (**B**). Formation of the lactone-ring might occur prior to epoxidation.

that the quantitative mapping method was sufficiently stringent to differentiate between the two sequences.

DISCUSSION

In the present study, we have demonstrated that *TgTPS1* encodes for a δ-cadinene synthase, an enzyme also identified in *G. arboreum* [31], and *TgTPS2* for a kunzeaol [or 6β-hydroxygermacra-1(10),4-diene] synthase, a new enzyme that has not been reported previously. A proposed mechanistic transformation of FPP to δ-cadinene and kunzeaol is depicted in Figure 6(A). For kunzeaol synthesis, after initial carbocation formation at C-11 from FPP, re-localization of the positive charge from C-11 to C-6 is likely to occur by an 1,3-hydride shift. The major product, kunzeaol, can be formed by a water quench, whereas deprotonation and/or double-bond rearrangement leads to the formation of germacrene D and bicyclogermacrene, two minor products from TgTPS2. Therefore synthesis of kunzeaol and two known terpenes by TgTPS2 can be explained by simple carbocation reactions.

The identification and characterization of TgTPS1 as a dedicated δ-cadinene synthase is in agreement with the observation that δ-cadinene is the most abundant sesquiterpene in *T. garganica* roots, representing approximately 35 % of the total volatile content [37]. It should be noted that TgTPS1

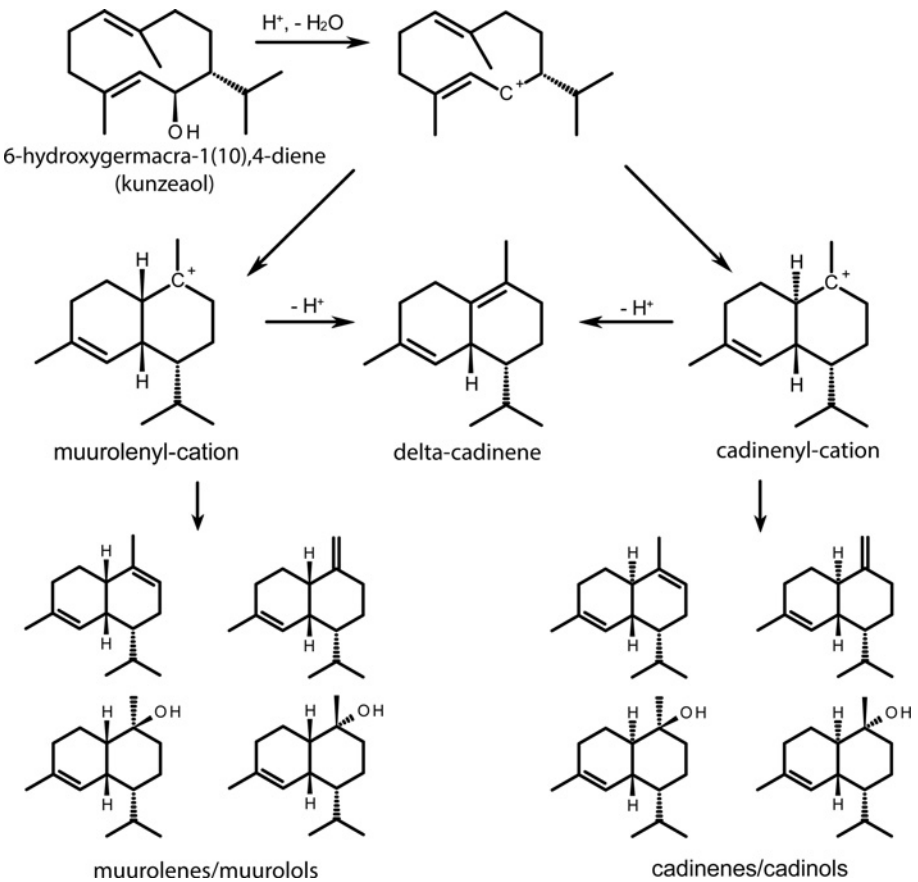

Figure 7 Proposed mechanism of the acid-induced rearrangement of kunzeaol to cadinenes and muurolenes [39]

and TgTPS2 are 89 % identical in their amino acid sequences (Supplementary Figure S1), and the formation of TgTPS1/2 terpene products, kunzeaol and δ-cadinene, shares a common carbocation intermediate (germacrene A bearing a carbocation at C-6). Apparently subtle structural variations by recent diverged evolution of TgTPS1/2 determine the distinct reaction paths shown in Figure 6(A). Therefore TgTPS1/2 can serve as good templates to investigate the diverged evolution and structure–function analysis of the TPS enzyme family.

The present study also showed that multiple terpenoid products can arise from influences other than the actual innate enzyme activity (i.e. low pH). A mechanism for acid-induced rearrangement of kunzeaol was proposed by Cornwell et al. [39] and is shown in Figure 7. After protonation of the hydroxy group, elimination of water leads to the formation of a secondary carbocation, which rearranges to a tertiary cadinenyl cation by electrophilic attack of the double bond. The positive charge is then quenched either by elimination of a proton resulting in the formation of cadinene- and muurolene-type molecules or by electrophilic attack of water leading to cadinols and muurulols. Those rearrangement products from kunzeaol were also identified from *TgTPS2*-expressing yeast cultured in acidic conditions (i.e. δ-cadinene, τ-muurolol and α-cadinol). Although germacrene D could also contribute to the generation of cadinene-type side-products via a similar mechanism [47], it does not seem to be the case under the culture conditions used as judged from its comparable amount in both buffered and unbuffered yeast cultures (Figure 3). The fact that germacrene D is present in *T. garganica* roots, whereas the acidic rearrangement products, τ-muurolol and α-cadinol, have not been detected [37] further supports that the latter two sesquiterpenes are not physiological products of TgTPS2 *in planta*. In a previous study, kunzeaol has not been detected in *T. garganica* [37]. This can be explained by its chemical instability and also because it is a transient intermediate for further modification into more complex oxygenated sesquiterpenes found in *T. garganica* [48].

It has been reported recently [49] that commonly found conifer diterpenes (e.g. levopimaradiene, abietadiene, neoabietadiene and palustradiene) are likely to be non-enzymatically dehydrated products of an unstable diterpene alcohol [i.e. 13-hydroxy-8(14)-abietene], which is directly synthesized by a diterpene synthase. Because of their instability, allylic alcohol terpenoids rearrange easily to structurally related terpenoids without involving enzymatic activities during sample preparation and analytical conditions. The previously reported studies [49] and results of the present study indicate that careful consideration of applied experimental conditions must be undertaken, when dealing with highly reactive terpenoid intermediates.

Although it is not immediately apparent from its chemical structure, kunzeaol serves as an excellent prospective precursor for thapsigargin biosynthesis. The thapsigargin core backbone has a guaiene (octahydroazulene) skeleton with a 4,5-double-bond and a 6,12-lactone ring (Figure 1). In our proposed model (Figure 6B), an epoxidation of kunzeaol occurs at the 1,10 double-bond, which evokes a rearrangement of the 4,5-double-bond, yielding a 5,7-ring guaiene carbocation. Subsequent deprotonation at C-5 yields a 5,7-ring-fused guaiene with hydroxy groups at C-6 and C-10, which is consistent with the thapsigargin structure (Figure 6B). Therefore we propose that the skeletal rearrangement of a germacrenolide, possibly mediated by a cytochrome P450, could

generate the backbone of thapsigargin, similar to the mechanisms previously suggested [21,48]. An analogous mechanism has been proposed for lactucopicrin, a guaianolide STL from lettuce (*Lactuca sativa*) [25]. It is reasonable that the secondary transformation of cyclodecane germacrene-type terpenes into structurally distinct guaienes is the common biochemical mechanisms in Apiaceae and Asteraceae. The present study suggests that STL synthesis via kunzeaol (C-6 hydroxy group) can explain the occurrence of 6,12-guainolides in Apiaceae [48].

Multiple lines of evidence shown in the present study support that *TgTPS2* is involved in thapsigargin biosynthesis. First, the extremely abundant *TgTPS2* transcript in *T. garganica* roots is congruent with the high thapsigargin level in root tissue as experimentally confirmed in the present study as well as previously reported [33]. Secondly, no obvious *TPS* transcript other than *TgTPS1* and *TgTPS2* was identified from the assembled 454 and Illumina sequencing data. Sequencing depth exceeded 29×10^6 reads, and hence we expected that the major TPS involved in the biosynthesis of thapsigargin should be present in these data sets. Although we cannot completely rule out that other *TPS* transcripts may have evaded detection from the current 454 and Illumina sequencings, their extremely low transcript abundance, if present, suggests that other TPSs do not make a significant contribution to the terpenoid profile of *T. garganica* roots. Thirdly, by employing simple epoxidation and subsequent skeletal rearrangement via protonation and deprotonation, which are typical mechanisms in terpenoid chemistry, a guaiene backbone for thapsigargin could be easily generated from kunzeaol. Finally, the relative configuration of C-6 hydroxy and isopropyl groups in kunzeaol matches the one found in the thapsigargin backbone, further implying that kunzeaol is likely to be the precursor of thapsigargin. Although the final conclusion for TgTPS2 being the first enzyme in thapsigargin biosynthesis requires additional experiments (e.g. further enzymatic conversion of kunzeaol into downstream structures resembling the thapsigargin backbone), current biochemical and transcriptome data strongly suggest that the guaiene backbone of thapsigargin is synthesized from kunzeaol. Ultimately, it needs to be examined whether the silencing of the *TgTPS2* transcript by RNA interference results in the decrease of thapsigargin *in planta*, although the lack of transformation techniques for *T. garganica* hinders this conclusive experiment. The elucidation of this step opens the possibility for the future *in planta* or microbial production of thapsigargin, or precursors thereof [50].

Taken together, the results of the present study showcase the successful application of next-generation sequencing in combination with classical biochemistry and metabolic engineering tools for the identification of novel biosynthetic genes in under-studied medicinal plants. The same approach could be used to elucidate complex terpenoid metabolism in various medicinal plants in the future.

AUTHOR CONTRIBUTION

Benjamin Pickel designed and performed the experiments and wrote the paper. Damian Drew generated the 454 sequencing data and cloned *TgTPS1/2* cDNAs, and Corinna Weitzel performed chemical analysis of *T. garganica* root. Tom Manczak characterized TgTPS1/2 in yeast. Dae-Kyun Ro designed the biochemical experiments and wrote the paper, and Henrik Simonsen analysed the NMR data of kunzeaol and wrote the paper.

ACKNOWLEDGEMENTS

We thank Qiao Wu (University of Calgary, Calgary, Canada) for NMR measurement and Soo-Un Kim (Seoul National University, Seoul, South Korea) for valuable discussions.

FUNDING

This work was supported by the Genome Canada [grant number PP780422], Genome Alberta, Government of Alberta [grant number PP789067], Genome Quebec, McGill Innovation Centre, MDEIE (Ministére du Développement économique, Innovation et Exportation), and Canada Research Chair program to D.K.R. and also by the Danish Strategic Research Council Grant 'SPOTLight' to H.T.S.

REFERENCES

1 Buckingham, J. (2011) Dictionary of Natural Products (on-line edition). Chapman and Hall/CRC Press, London
2 Gershenzon, J. and Dudareva, N. (2007) The function of terpene natural products in the natural world. Nat. Chem. Biol. **3**, 408–414
3 Heil, M. (2008) Indirect defence via tritrophic interactions. New Phytol. **178**, 41–61
4 Bohlmann, J. and Keeling, C. I. (2008) Terpenoid biomaterials. Plant J. **54**, 656–669
5 Christianson, D. W. (2006) Structural biology and chemistry of the terpenoid cyclases. Chem. Rev. **106**, 3412–3442
6 Greenhagen, B. T., O'Maille, P. E., Noel, J. P. and Chappell, J. (2006) Identifying and manipulating structural determinates linking catalytic specificities in terpene synthases. Proc. Nat. Acad. Sci. U.S.A. **103**, 9826–9831
7 Yoshikuni, Y., Ferrin, T. E. and Keasling, J. D. (2006) Designed divergent evolution of enzyme function. Nature **440**, 1078–1082
8 Rasmussen, U., Christensen, S. B. and Sandberg, F. (1978) Thapsigargine and thapsigargicine, two new histamine liberators from *Thapsia garganica*. Acta Pharm. Suec. **15**, 133–140
9 Thastrup, O., Cullen, P. J., Drobak, B. K., Hanley, M. R. and Dawson, A. P. (1990) Thapsigargin, a tumor promoter, discharges intracellular Ca^{2+} stores by specific inhibition of the endoplasmic reticulum Ca^{2+}-ATPase. Proc. Nat. Acad. Sci. U.S.A. **87**, 2466–2470
10 Sagara, Y. and Inesi, G. (1991) Inhibition of the sarcoplasmic reticulum Ca^{2+} transport ATPase by thapsigargin at subnanomolar concentrations. J. Biol. Chem. **266**, 13503–13506
11 Waldron, R. T., Short, A. D., Meadows, J. J., Ghosh, T. K. and Gill, D. L. (1994) Endoplasmic reticulum calcium pump expression and control of cell growth. J. Biol. Chem. **269**, 11927–11933
12 Short, A. D., Bian, J., Ghosh, T. K., Waldron, R. T., Rybak, S. L. and Gill, D. L. (1993) Intracellular Ca^{2+} pool content is linked to control of cell growth. Proc. Nat. Acad. Sci. U.S.A. **90**, 4986–4990
13 Ghosh, T. K., Bian, J. H., Short, A. D., Rybak, S. L. and Gill, D. L. (1991) Persistent intracellular calcium pool depletion by thapsigargin and its influence on cell growth. J. Biol. Chem. **266**, 24690–24697
14 Lin, X. S., Denmeade, S. R., Cisek, L. and Isaacs, J. T. (1997) Mechanism and role of growth arrest in programmed (apoptotic) death of prostatic cancer cells induced by thapsigargin. Prostate **33**, 201–207
15 Futami, T., Miyagishi, M. and Taira, K. (2005) Identification of a network involved in thapsigargin-induced apoptosis using a library of small interfering RNA expression vectors. J. Biol. Chem. **280**, 826–831
16 Furuya, Y., Lundmo, P., Short, A. D., Gill, D. L. and Isaacs, J. T. (1994) The role of calcium, pH, and cell proliferation in the programmed (apoptotic) death of androgen-independent prostatic cancer cells induced by thapsigargin. Cancer Res. **54**, 6167–6175
17 Wootton, L. L. and Michelangeli, F. (2006) The effects of the phenylalanine 256 to valine mutation on the sensitivity of sarcoplasmic/endoplasmic reticulum Ca^{2+} ATPase (SERCA) Ca^{2+} pump isoforms 1, 2, and 3 to thapsigargin and other inhibitors. J. Biol. Chem. **281**, 6970–6976
18 Bobe, R., Bredoux, R., Corvazier, E., Martin, V., Gelebart, P., Lacabarztz-Porret, C., Launay, S., Fanchaouy, M., Dally, S., Chaabane, C. et al. (2005) Sarco/endoplasmic reticulum Ca^{2+} ATPase (SERCA) isoforms Curr. Top. Biochem. Res. **7**, 1–16
19 Thastrup, O., Dawson, A. P., Scharff, O., Foder, B., Cullen, P. J., Drobak, B. K., Bjerrum, P. J., Christensen, S. B. and Hanley, M. R. (1989) Thapsigargin, a novel molecular probe for studying intracellular calcium release and storage. Agents Actions **27**, 17–23
20 Christensen, S. B., Skytte, D. M., Denmeade, S. R., Dionne, C., Moller, J. V., Nissen, P. and Isaacs, J. T. (2009) A Trojan horse in drug development: targeting of thapsigargins towards prostate cancer cells. Anti Cancer Agents Med. Chem. **9**, 276–294
21 Simonsen, H. T., Weitzel, C. and Christensen, S. B. (2012) Guaianolide sesquiterpenoids: their pharmacology and biosynthesis. Handbook of Natural Products (Ramawat, K. G. and Merillon, J. M., eds), in press, Springer-Verlag, Berlin
22 Ikezawa, N., Göpfert, J. C., Nguyen, D. T., Kim, S. U., O'Maille, P. E., Spring, O. and Ro, D. K. (2011) Lettuce costunolide synthase (CYP71BL2) and its homolog (CYP71BL1) from sunflower catalyze distinct regio- and stereoselective hydroxylations in sesquiterpene lactone metabolism. J. Biol. Chem. **286**, 21601–21611

23 Liu, Q., Majdi, M., Cankar, K., Goedbloed, M., Charnikhova, T., Verstappen, F. W., de Vos, R. C., Beekwilder, J., van der Krol, S. and Bouwmeester, H. J. (2011) Reconstitution of the costunolide biosynthetic pathway in yeast and *Nicotiana benthamiana*. PLoS ONE **6**, e23255

24 Nguyen, D. T., Gopfert, J. C., Ikezawa, N., MacNevin, G., Kathiresan, M., Conrad, J., Spring, O. and Ro, D. K. (2010) Biochemical conservation and evolution of germacrene A oxidase in asteraceae. J. Biol. Chem. **285**, 16588–16598

25 de Kraker, J. W., Franssen, M. C., Joerink, M., de Groot, A. and Bouwmeester, H. J. (2002) Biosynthesis of costunolide, dihydrocostunolide, and leucodin. Demonstration of cytochrome P450-catalyzed formation of the lactone ring present in sesquiterpene lactones of chicory. Plant Physiol. **129**, 257–268

26 Andrews, S. P., Ball, M., Wierschem, F., Cleator, E., Oliver, S., Hogenauer, K., Simic, O., Antonello, A., Hunger, U., Smith, M. D. and Ley, S. V. (2007) Total synthesis of five thapsigargins: guaianolide natural products exhibiting sub-nanomolar SERCA inhibition. Chemistry **13**, 5688–5712

27 Jäger, A. K., Schottländer, B. and Nyman, U. (1993) Somatic embryogenesis in cell cultures of *Thapsia garganica*: correlation between the state of differentiation and the content of thapsigargins. Plant Cell Rep. **12**, 517–520

28 Meisel, L., Fonseca, B., Gonzalez, S., Baeza-Yates, R., Cambiazo, V., Campos, R., Gonzalez, M., Orellana, A., Retamales, J. and Silva, H. (2005) A rapid and efficient method for purifying high quality total RNA from peaches (*Prunus persica*) for functional genomics analyses. Biol. Res. **38**, 83–88

29 R Development Core Team (2010) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

30 Stein, S. E. (1999) An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data. J. Am. Soc. Mass Spectrom. **10**, 770–781

31 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. J. Mol. Biol. **215**, 403–410

32 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods **5**, 621–628

33 Jäger, A. K., Adsersen, A., Gudiksen, L. and Smitt, U. W. (1992) Comparative studies on *Thapsia garganica* and *Thapsia transtagana*: quantitative investigations of thapsigargins in roots and fruits. Planta Med. **58**, A700–A701

34 Ro, D. K., Ouellet, M., Paradise, E. M., Burd, H., Eng, D., Paddon, C. J., Newman, J. D. and Keasling, J. D. (2008) Induction of multiple pleiotropic drug resistance genes in yeast engineered to produce an increased level of anti-malarial drug precursor, artemisinic acid. BMC Biotechnol. **8**, 83

35 Ro, D. K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J. et al. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature **440**, 940–943

36 Chen, X. Y., Chen, Y., Heinstein, P. and Davisson, V. J. (1995) Cloning, expression, and characterization of (+)-δ-cadinene synthase: a catalyst for cotton phytoalexin biosynthesis. Arch. Biochem. Biophys. **324**, 255–266

37 Drew, D. P., Rasmussen, S. K., Avato, P. and Simonsen, H. T. (2012) A comparison of headspace solid-phase microextraction and classic hydrodistillation for the identification of volatile constituents from *Thapsia* spp. provides insights into guaianolide biosynthesis in Apiaceae. Phytochem. Anal. **23**, 44–51

38 Barrero, A. F., Herrador, M. M., Quilez, J. F., Alvarez-Manzaneda, R., Portal, D., Gavin, J. A., Gravalos, D. G., Simmonds, M. S. and Blaney, W. M. (1999) Bioactive sesquiterpenes from *Santolina rosmarinifolia* subsp. *canescens*. A conformational analysis of the germacrane ring. Phytochemistry **51**, 529–541

39 Cornwell, C. P., Reddy, N., Leach, D. N. and Wylie, S. G. (2001) Germacradienols in the essential oils of the Myrtaceae. Flavour Frag. J. **16**, 263–273

40 Szafranek, B., Chrapkowska, K., Waligora, D., Palavinskas, R., Banach, A. and Szafranek, J. (2006) Leaf surface sesquiterpene alcohols of the potato (*Solanum tuberosum*) and their influence on colorado beetle (*Leptinotarsa decemlineata* Say) feeding. J. Agric. Food Chem. **54**, 7729–7734

41 Karlsson, M. F., Birgersson, G., Cotes Prado, A. M., Bosa, F., Bengtsson, M. and Witzgall, P. (2009) Plant odor analysis of potato: response of guatemalan moth to above- and belowground potato volatiles. J. Agric. Food Chem. **57**, 5903–5909

42 Takeoka, G. R., Hobbs, C. and Park, B. S. (2010) Volatile constituents of the aerial parts of *Salvia apiana* Jepson. J. Essent. Oil Res. **22**, 241–244

43 Faraldos, J. A., Wu, S., Chappell, J. and Coates, R. M. (2007) Conformational analysis of (+)-germacrene A by variable temperature NMR and NOE spectroscopy. Tetrahedron **63**, 7733–7742

44 Cane, D. E. (1990) Enzymatic formation of sesquiterpenes. Chem. Rev. **90**, 1089–1103

45 Zerbino, D. R. and Birney, E. (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**, 821–829

46 Hamilton, J. P. and Robin Buell, C. (2012) Advances in plant genome sequencing. Plant J. **70**, 177–190

47 Bulow, N. and Konig, W. A. (2000) The role of germacrene D as a precursor in sesquiterpene biosynthesis: investigations of acid catalyzed, photochemically and thermally induced rearrangements. Phytochemistry **55**, 141–168

48 Drew, D. P., Krichau, N., Reichwald, K. and Simonsen, H. T. (2009) Guaianolides in apiaceae: perspectives on pharmacology and biosynthesis. Phytochem. Rev. **8**, 581–599

49 Keeling, C. I., Madilao, L. L., Zerbe, P., Dullat, H. K. and Bohlmann, J. (2011) The primary diterpene synthase products of *Picea abies* levopimaradiene/abietadiene synthase (PaLAS) are epimers of a thermally unstable diterpenol. J. Biol. Chem. **286**, 21145–21153

50 Simonsen, H. T., Drew, D. P. and Lunde, C. (2009) Perspectives on using *Physcomitrella patens* as an alternative production platform for thapsigargin and other terpenoid drug candidates. Perspec. Med. Chem. **3**, 1–6

Received 30 April 2012/30 August 2012; accepted 31 August 2012
Published as BJ Immediate Publication 31 August 2012, doi:10.1042/BJ20120654

Biochem. J. (2012) **448**, 261–271 (Printed in Great Britain) doi:10.1042/BJ20120654

SUPPLEMENTARY ONLINE DATA

Identification and characterization of a kunzeaol synthase from *Thapsia garganica*: implications for the biosynthesis of the pharmaceutical thapsigargin

Benjamin PICKEL*, Damian P. DREW†, Tom MANCZAK†, Corinna WEITZEL†, Henrik T. SIMONSEN†[1] and Dae-Kyun RO*[1]

*Department of Biological Sciences, University of Calgary, Calgary, Canada, T2N 1N4, and †Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, 1871 Frederiksberg, Denmark

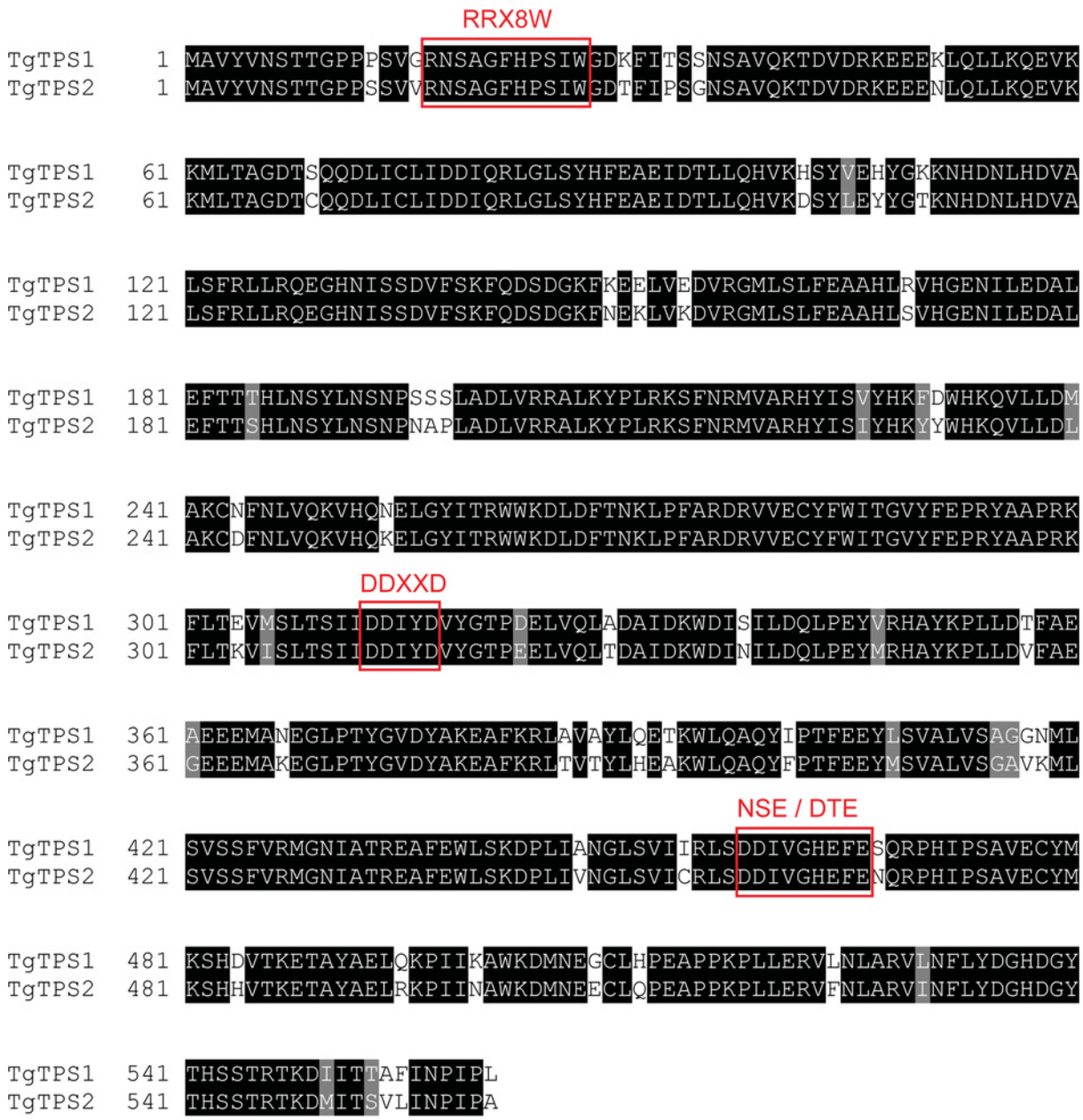

Figure S1 An alignment of amino acid sequences of TgTPS1 and TgTPS2

The conserved motifs in sesqui-TPSs are indicated by the red boxes. The DDXXD motif in TgTPS1/2 is perfectly conserved, but slight deviations of the RRX8W and NSE/DTE motifs [i.e. (D/N)DXX(T/S)XXX(E/D)] are observed in TgTPS1/2.

[1] Correspondence may be addressed to either of these authors (email hts@life.ku.dk or daekyun.ro@ucalgary.ca).

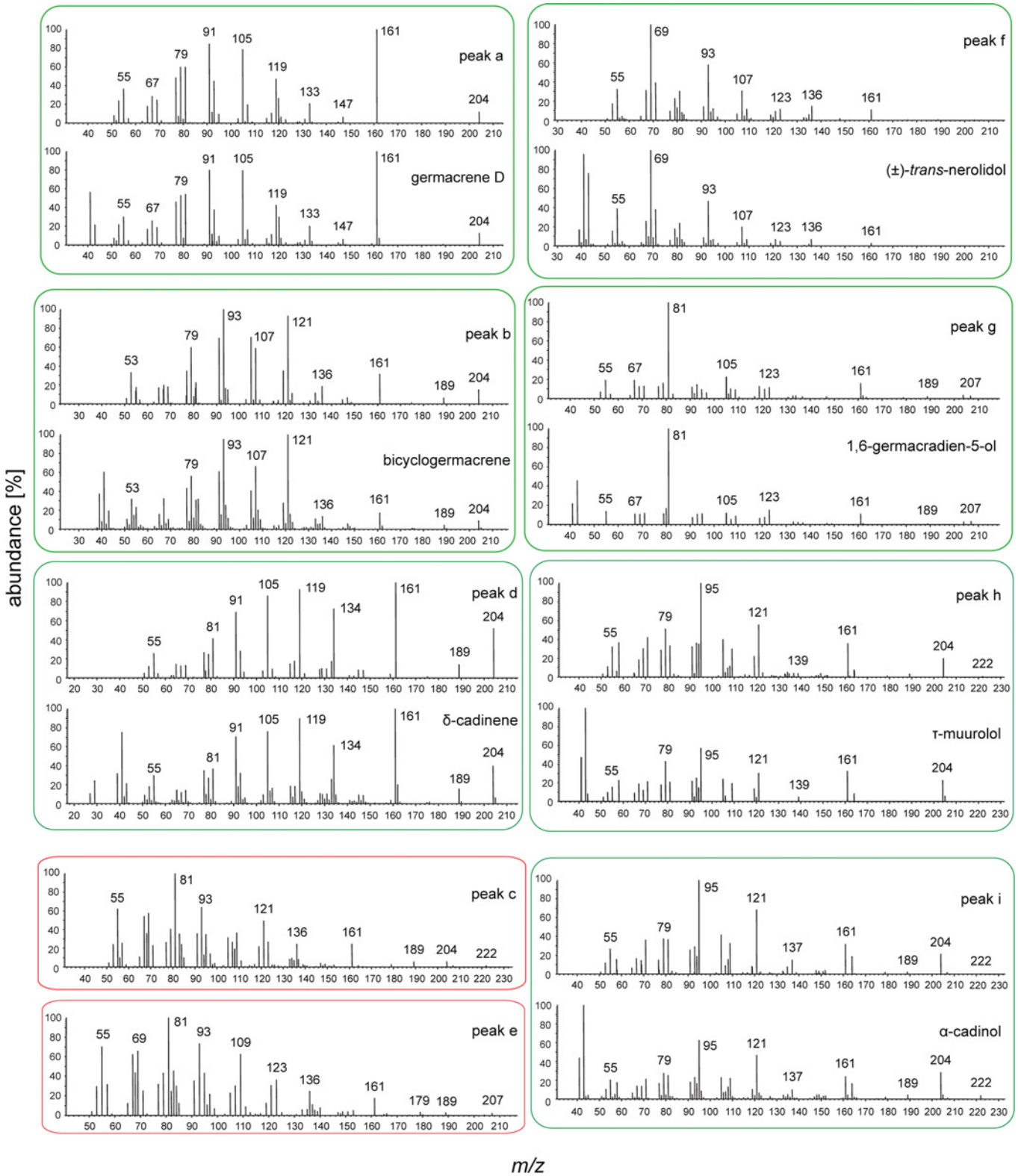

Figure S2 Mass spectra of peaks a–i from *TgTPS2*-expressing yeast with available corresponding reference spectra

Peaks c and e boxed by red lines do not have matching reference spectra.

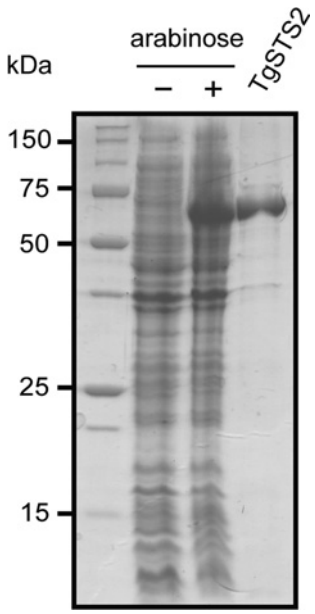

Figure S3 SDS/PAGE of total protein extract of *TgTPS2*-expressing *E. coli* (total protein corresponding to 0.2 ml of a culture with $D_{600} = 1$) and purified TgTPS2 (2.5 μg)

Table S1 Primer sequences

Primer	Sequence (5′→3′)
c2580 forward	ATGGCTGTGTATGTTAACTCTACAAC
c16806 forward	ATGGCTGTGTATGTTAACTCTACAAC
c2586 reverse	AGCTGGGCTCCTGCCTGTC
c313 reverse	ACGGGCTCCTGCCTCAAAG
TgTPS1 forward	GAGAGGATCCAGCATGGCTGTGTATGTTAATTCTACAAC
TgTPS2 reverse	CACACCTCGAGCTATAGTGGAATGGGATTTATGAAAG
TgTPS2 forward	GAGAGGATCCAGCATGGCTGTGTATGTTAACTCTAC
TgTPS2 reverse	CACACCTCGAGCTACGCTGGAATGGGATTTATG
His_6-TgTPS2 forward	AAAAAGCAGGCTTCGCTGTGTATGTTAACTCTACAAC
His_6-TgTPS2 reverse	AGAAAGCTGGGTCCTACGCTGGAATGGGATTTATG

Table S2 Comparison of ^{13}C-signals (δ_C in p.p.m.) from purified 6β-hydroxygermacra-1(10),4-diene with the ones from kunzeaol described in literature [1,2]

The annotation is according to Cornwell et al. [1], and signals in parentheses belong to the second conformer of 6β-hydroxygermacra-1(10),4-diene.

Carbon number	The present study	Cornwell et al. [1]	Szafranek et al. [2]
C-1	121.3 (128.8)	121.1 (128.5)	121.4 (128.8)
C-2	41.4 (35.8)	41.2 (35.6)	41.4 (35.8)
C-3	37.1 (39.1)	36.9 (38.8)	37.1 (39.1)
C-4	133.2 (132.9)	132.3 (132.2)	132.4
C-5	131.5 (133.6)	131.7 (133.8)	131.5 (133.6)
C-6	68.8	68.5	68.8
C-7	52.2 (49.4)	51.9 (49.2)	52.2 (49.4)
C-8	25.0 (30.1)	24.9 (29.9)	25.1 (30.1)
C-9	24.3	24.3	24.3
C-10	138.7 (135.7)	138.4 (135.3)	138.9
C-11	32.2 (31.7)	31.9 (31.5)	32.2 (31.7)
C-12	21.3 (21.0)	20.0	21.1
C-13	21.2 (21.1)	20.0	21.1
C-14	16.9 (21.9)	16.7 (21.7)	16.9 (22.0)
C-15	16.3	16.2	16.4

REFERENCES

1 Cornwell, C. P., Reddy, N., Leach, D. N. and Wylie, S. G. (2001) Germacradienols in the essential oils of the Myrtaceae. Flavour Frag. J. **16**, 263–273

2 Szafranek, B., Chrapkowska, K., Waligora, D., Palavinskas, R., Banach, A. and Szafranek, J. (2006) Leaf surface sesquiterpene alcohols of the potato (*Solanum tuberosum*) and their influence on colorado beetle (*Leptinotarsa decemlineata* Say) feeding. J. Agric. Food Chem. **54**, 7729–7734

Received 30 April 2012/30 August 2012; accepted 31 August 2012
Published as BJ Immediate Publication 31 August 2012, doi:10.1042/BJ20120654