

Rational engineering of plasticity residues of sesquiterpene synthases from *Artemisia annua*: product specificity and catalytic efficiency

Jian-Xu LI*^{†1}, Xin FANG*¹, Qin ZHAO*[†], Ju-Xin RUAN*[†], Chang-Qing YANG*, Ling-Jian WANG*, David J. MILLER[‡], Juan A. FARALDOS[‡], Rudolf K. ALLEMANN[‡], Xiao-Ya CHEN*^{§2} and Peng ZHANG*^{||2}

*National Key Laboratory of Plant Molecular Genetics and National Plant Gene Research Center, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China, [†]Graduate School of Chinese Academy of Sciences, Beijing 100049, China, [‡]School of Chemistry and Cardiff Catalysis Institute, Cardiff University, Main Building, Park Place, Cardiff CF10 3AT, U.K., [§]Plant Science Research Center, Shanghai Chenshan Botanical Garden, Shanghai 201602, China, and ^{||}Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China

Most TPSs (terpene synthases) contain plasticity residues that are responsible for diversified terpene products and functional evolution, which provide a potential for improving catalytic efficiency. Artemisinin, a sesquiterpene lactone from *Artemisia annua* L., is widely used for malaria treatment and progress has been made in engineering the production of artemisinin or its precursors. In the present paper, we report a new sesquiterpene synthase from *A. annua*, AaBOS (*A. annua* α -bisabolol synthase), which has high sequence identity with AaADS (*A. annua* amorpho-4,11-diene synthase), a key enzyme in artemisinin biosynthesis. Comparative analysis of the two enzymes by domain-swapping and structure-based mutagenesis led to the identification of several plasticity residues, whose alteration changed the product profile of AaBOS to include γ -humulene as the major product. To elucidate

the underlying mechanisms, we solved the crystal structures of AaBOS and a γ -humulene-producing AaBOS mutant (termed AaBOS-M2). Among the plasticity residues, position 399, located in the substrate-binding pocket, is crucial for both enzymes. In AaBOS, substitution of threonine for leucine (AaBOS_{L339T}) is required for γ -humulene production; whereas in AaADS, replacing the threonine residue with serine (AaADS_{T399S}) resulted in a substantial increase in the activity of amorpho-4,11-diene production, probably as a result of accelerated product release. The present study demonstrates that substitution of plasticity residues has potential for improving catalytic efficiency of the enzyme.

Key words: *Artemisia annua*, catalytic efficiency, domain swapping, plasticity residue, product profile, terpene synthase.

INTRODUCTION

Terpenoids constitute a large and structurally diverse class of natural products. The plant terpenoids not only contribute to plant adaptation to the environment [1], but also provide a valuable source of lead compounds for medicine [2,3]. For instance, artemisinin, an endoperoxide sesquiterpene lactone isolated from the traditional herb *Artemisia annua* L., is an effective component against malaria [4,5]. Each individual terpene compound is often present in small quantities in the native organism, and chemical syntheses of these compounds are usually difficult owing to their complex structure, hindering their usage by humans. These problems prompted investigation of alternative methodologies to obtain valuable terpenoids, among which synthetic biology is considered particularly promising [6].

TPSs (terpene synthases) catalyse the conversion of allylic prenyldiphosphate building blocks, including GPP (geranyl diphosphate; C₁₀), FPP (farnesyl diphosphate; C₁₅) and GGPP (geranylgeranyl diphosphate; C₂₀), into different classes of terpenes [1,7,8]. They are important in directing isoprenoid flux into a specific group of terpenes as they operate at metabolic branch points and often compete with other enzymes for an allylic prenyldiphosphate pool. For example, FPP is the precursor

to many cellular molecules such as squalene, dolichols and the cofactor haem, in addition to various sesquiterpenes formed by sesquiterpene synthases. However, TPSs show lower catalytic efficiency than most of the central metabolic enzymes [9]. The average k_{cat} value of the former is 2.5 s⁻¹, whereas the latter is 79 s⁻¹, approximately 30-fold greater than the former [9], suggesting that TPSs are far from optimized by natural selection in terms of catalytic efficiency. TPSs hold the key to high yield of desired products in metabolic engineering, thus understanding the catalytic mechanisms of these enzymes is of great importance.

TPSs have complicated catalytic mechanisms. They contain large active sites that allow the binding and folding of the substrate, the generation and stabilization of high-energy carbocations, and the acidic/basic catalysis to dictate specific skeletal and metabolic fates [7,10]. Finally, the reaction is ended by product release, which is often a rate-limiting step [11]. It has been proposed that the secondary metabolic enzymes tend to develop catalytic elasticity during evolution [12], which allows emergence of new enzymes with expanded substrate scope and multiple new products. Within and around the active site of a TPS, certain 'special' amino acid residues (plasticity residues) are mainly responsible for the catalytic elasticity [13,14], and alterations to a small number of them can

Abbreviations used: AaADS, *Artemisia annua* amorpho-4,11-diene synthase; AaBOS, *Artemisia annua* α -bisabolol synthase; AaFS, *Artemisia annua* β -farnesene synthase; AaGAS, *Artemisia annua* germacrene A synthase; EST, expressed sequence tag; 2F-FPP, 2-fluorofarnesyl diphosphate; 6F-FPP, 6-fluorofarnesyl diphosphate; FHP, farnesylhydroxyphosphonate; FPP, farnesyl diphosphate; NtEAS, *Nicotiana tobacum* 5-*epi*-aristolochene synthase; qRT-PCR, quantitative real-time PCR; r.m.s.d., root mean square deviation; TPS, terpene synthase.

¹ These authors contributed equally to this work.

² Correspondence may be addressed to either of these authors (email xychen@sibs.ac.cn or pengzhang01@sibs.ac.cn).

Atomic co-ordinates have been deposited in the PDB under accession codes 4FJQ for AaBOS and 4GAX for AaBOS-M2.

profoundly shift the product specificity [8,15,16]. This plasticity allows host organisms to drastically change the spectrum of terpenoid production at the expense of only a few amino acid substitutions in TPSs, thus these enzymes are considered hot spots in evolution [15,17]. Gene duplication and tandem duplication events have provided a genetic reservoir for the biosynthetic diversity of terpenes, as the extra copies of TPS genes are available for mutation and evolution without compromising the original activity [18,19]. The important role of plasticity residues in catalytic specificity suggests that they should exert influence on certain steps of the complex catalytic reaction. This is supported by the reports that changed product profiles could be accompanied by altered reaction velocities as the result of substitutions of a few amino acids of enzymes [15,20,21]; an isoleucine-to-threonine change, for example, switched an *Arabidopsis thaliana* kaurene synthase to a pimaradiene synthase with a 4-fold decrease in catalytic efficiency [16]. However, the involvement of plasticity residues of a TPS in catalytic activity have not been studied in detail, and their potential for improving catalytic efficiency remains unexplored.

Domain swapping between related enzymes and site-directed mutagenesis have been frequently used to identify functional regions and/or residues of TPSs [15–17,19,21,22]. In most cases, domain swapping could lead to defining the minimum segment or residues that mediate product specificity, but without a complete change of products generated, thus suggesting participation of other residues. A structure-based strategy has been developed in recent years, in which comparison of three-dimensional structures of TPSs allowed the identification of plasticity residues directly or indirectly involved in the catalytic process [23]. On the basis of the structure of NtEAS (*Nicotiana tabacum* 5-*epi*-aristolochene synthase) [13], a reciprocal interconversion of catalytic specificities between NtEAS and *Hyoscyamus muticus* premnaspirodiene synthase was achieved with substitutions of nine amino acids [14]. Besides structural information, phylogenetic comparisons are also important in guiding mutagenesis experiments [8].

A. annua harbours a wealth of terpenoids [24–29] and has a long history of use in traditional Chinese medicine [30]. Enzymes catalysing the first step of artemisinin biosynthesis [AaADS (*A. annua* amorpho-4,11-diene synthase)] and subsequent reactions [24,25,28,31–33], as well as regulatory factors [34] of the artemisinin pathway have been reported. Despite the importance of AaADS in artemisinin biosynthesis, at present the molecular basis of its catalytic mechanism remains elusive, partly due to the lack of a functionally similar enzyme in *A. annua* and structural information of AaADS.

In the present paper, we report the characterization of a novel sesquiterpene synthase from AaBOS (*A. annua* α -bisabolol synthase), which displays a high sequence identity with AaADS. Determination of the crystal structure of AaBOS, and domain-swapping and site-directed mutagenesis experiments led to the identification of residues, which we designated the BOS motif, that are essential for product specificity in both enzymes. Importantly, the T399S mutation in the BOS motif of AaADS (AaADS_{T399S}) drastically increased catalytic efficiency of the mutant enzyme.

MATERIALS AND METHODS

Materials

Seeds of *A. annua* cv. Qiute were surface-sterilized and germinated in Murashige and Skoog medium. Seedlings (2 weeks old) were transferred to soil and grown in a greenhouse at 25 °C under a light intensity of 150 μmol of photons $\cdot \text{m}^{-2} \cdot \text{s}^{-1}$

with a 14-h-light/10-h-dark cycle. Tissues of 10-week-old plants were used for gene expression analysis unless otherwise indicated. Leaves close to inflorescences (approximately one-third of the upper stem) were defined as young, and leaves close to the basal part (one-third of the lower stem) were defined as mature. Isolation of glandular trichomes and cDNA library construction were performed according to a previously published protocol [29]. Chemicals were purchased from Sigma–Aldrich, and FHP (farnesylhydroxyphosphonate) was from Enzo Life Sciences. 2F-FPP (2-fluorofarnesyl diphosphate) and 6F-FPP (6-fluorofarnesyl diphosphate) were synthesized as described previously [35,36]. ESTs (expressed sequence tags) were generated at the Shanghai Biochip Company.

Gene isolation, expression and sequence analysis

Total RNA was extracted using TRIzol[®] reagent (Invitrogen). Total RNA (1 μg) was reverse-transcribed using the RNA PCR [AMV (avian myoblastosis virus)] kit (TaKaRa), followed by cDNA synthesis and gene expression analysis. Full-length cDNAs were isolated by 5'- and 3'-rapid amplification of cDNA ends using the Pfu enzyme (Promega). qRT-PCR (quantitative real-time PCR) was performed with SYBR Green PCR Mastermix (TaKaRa) on a Mastercycler[®] ep RealPlex2 (Eppendorf) cyler with *A. annua* actin (EU531837) as a reference. The expression levels of genes were determined as described previously [37], and the relative expression of each gene was compared with the expression level in root which was set to 1. Nucleotide and amino acid sequence alignments were performed using ClustalW [38]. Secondary structure predictions were performed by ESPript [39]. Primers used in this investigation are listed in Supplementary Table S1 (at <http://www.biochemj.org/bj/451/bj4510417add.htm>).

Mutagenesis, bacterial expression and enzyme assay

Mutagenesis was generated by PCR using an overlap extension strategy [40] with the respective primers (Supplementary Table S1). The ORFs (open reading frames) of AaBOS, AaADS and the mutant cDNAs were amplified using Pfu DNA polymerase, followed by insertion into pET32a. *Escherichia coli* BL21(DE3) cells were used for protein production. Recombinant proteins were purified with Ni-NTA (Ni²⁺-nitrilotriacetate) spin columns (Qiagen). The protein concentration was determined using the Bradford method with BSA as a standard.

Assays of sesquiterpene synthases were performed in a volume of 500 μl containing 20 μg of protein, 25 mM Hepes (pH 7.0), 5 mM MgCl₂, 5 mM DTT (dithiothreitol) and 40 μM FPP [41], and conducted at 37 °C for 1 h unless otherwise indicated. The reaction mixture was extracted with 400 μl of hexane and subjected to analysis by GC–MS (Agilent 6890 Series GC System coupled to an Agilent 5973 Network Mass Selective Detector), with the following temperature programme: initial temperature of 80 °C (2-min hold), increase to 250 °C at 10 °C/min, and ramp to 280 °C at 20 °C/min (3-min hold). Products were identified by comparison with authentic standards (Sigma–Aldrich), and NIST (National Institute of Standards and Technology) and Wiley libraries. For quantitative analysis, nerolidol was added as an internal standard during hexane extraction of the enzyme reaction mixture.

The kinetic parameters of AaBOS, AaADS and the mutant enzymes were determined according to previously published methods [8]. Briefly, the purified enzyme of 3 μg was added to each assay mixture containing various concentrations of FPP in the range 3–100 μM . The reaction was allowed to proceed for 5 or

Table 1 Statistics of diffraction data and structure refinement

Numbers in parentheses represent the highest resolution shell.

Parameter	AaBOS	AaBOS-M2
Data collection		
Wavelength (Å)	1.0000	1.0000
Space group	$P2_1$	$P2_1$
Resolution (Å)	50.00–2.00 (2.07–2.00)	50.00–2.00 (2.07–2.00)
Cell parameters		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	52.0, 77.3, 68.8	52.7, 77.1, 68.3
β (°)	102.5	102.3
Observed reflections	131389	134078
Unique reflections [$\ I/I\ > 0$]	35037 (3409)	36270 (3646)
Average redundancy	3.8 (3.7)	3.7 (3.8)
Average I/I	16.0 (2.9)	20.2 (7.4)
Completeness (%)	97.2 (94.7)	99.7 (99.6)
R_{merge} (%)	8.7(52.3)	8.2 (40.8)
Refinement		
Resolution (Å)	50.00–2.00	50.00–2.00
$R_{\text{work}}/R_{\text{free}}$	0.181/0.223	0.175/0.230
Number of protein atoms	4286	4294
Number of water atoms	285	324
Average <i>B</i> factor of all atoms (Å ²)	26.9	38.9
Main chain	24.8	35.7
Side chain	28.6	42.1
Water	29.6	38.7
R.m.s.d.		
Bond lengths (Å)	0.007	0.008
Bond angles (°)	0.998	1.033
Ramachandran plot (%)		
Favoured	96.8	91.5
Allowed	3.2	8.5

10 min at 25 °C, stopped by the addition of 0.5 M EDTA (pH 8.0) and products were extracted with 400 μ l of hexane containing 10 ng/ μ l nerolidol as an internal standard, followed by GC–MS analysis. The assays were performed in TIM (total ion monitoring) mode. One unit of enzyme activity was defined as the amount of enzyme required to produce 1 μ mol of product per min under the conditions stated above. Values were obtained with GraphPad Prism 5 software fitting the Michaelis–Menten curve.

The optimal temperature was determined at a series of temperatures ranging from 25 to 50 °C in 25 mM Hepes (pH 7.0), and the optimal pH was determined at 25 °C in 50 mM Bis-Tris buffer (pH 6.0–7.0), 50 mM Tris/HCl (pH 7.0–8.6) or 50 mM glycine/NaOH (pH 8.6–10.6).

Crystallization, data collection and structure determination

Protein was concentrated to 6 mg/ml for crystal screening, and the protein in 0.8 μ l was mixed with an equal volume of reservoir solution, and crystals were grown at 4 °C by the sitting-drop vapour-diffusion method. Rod-like crystals appeared after 3 days and grew to approximately 0.05 mm \times 0.05 mm \times 0.10 mm in 2 weeks. The crystals used for data collection were grown in 0.1 M Hepes and 29% PEG [poly(ethylene) glycol] 3000 at pH 7.8. The crystals were flash-frozen at 100 K and diffraction data were collected to 2.0 Å (1 Å = 0.1 nm) at beamline BL17U at the SSRF (Shanghai Synchrotron Radiation Facility). Data were processed with HKL2000 [42] and belonged to the space group $P2_1$. Additional processing was carried out using programs from the CCP4 suite [43].

The structure of AaBOS was determined using the molecular replacement method with PDB code 3M01 (NtEAS) [44] as the search model. One protein molecule was found in each asymmetric unit. Model building and refinement were carried

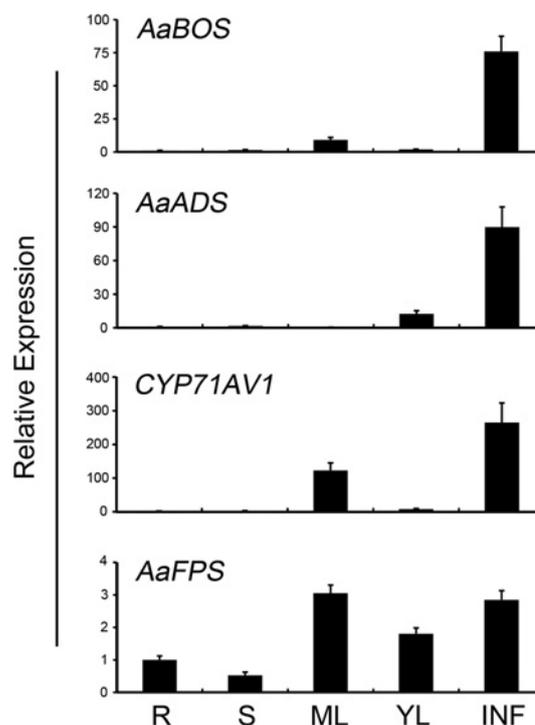


Figure 1 Expression levels of *A. annua* farnesyl diphosphate synthase (AaFPS; AF112881), CYP71AV1 (DQ872632), AaADS and AaBOS in different organs of the 10-week-old *A. annua* plants

The relative expression of each gene was compared with its expression level in roots which was set to 1. Values are means \pm S.D. for three technical replicates. INF, inflorescence; ML, mature leaves; R, root; S, stem; YL, young leaves.

out using Coot [45] and Phenix [46]. The final model contained residues 17–546, with residues 449–457 being disordered. The structure of AaBOS-M2 (a γ -humulene-producing AaBOS mutant) was solved using AaBOS as a model, and refinement was conducted in a similar manner. Data collection and structure refinement statistics are summarized in Table 1.

RESULTS AND DISCUSSION

Isolation and characterization of AaBOS

From a cDNA library of *A. annua* glandular trichomes, one of the ESTs exhibited a high nucleotide sequence identity (92%) with AaADS. The cDNA (GenBank® accession number JQ717161) codes for a protein of 546 amino acids, the same length as AaADS, with a calculated molecular mass of 63.8 kDa. The protein contains a typical DDxxD (DDIYD) motif and an atypical NSE/DTE (NDLAGHKEE) motif common to the sesquiterpene synthase family [7] (Supplementary Figure S1 at <http://www.biochemj.org/bj/451/bj4510417add.htm>). qRT-PCR showed that the highest expression level of this gene was present in inflorescences. In mature leaves, the transcript abundance was low but definite, whereas in young leaves, stems and roots, the expression was barely detectable (Figure 1). Interestingly, the artemisinin pathway genes of *AaADS* and *CYP71AV1* showed a similar enrichment of transcripts in inflorescences, in agreement with previous reports [32,34].

To identify the activity of this putative sesquiterpene synthase, we expressed it in *E. coli* and performed the enzyme assay with FPP. GC-MS analysis of the assay extract revealed a major product, α -bisabolol (Figure 2), which accounted

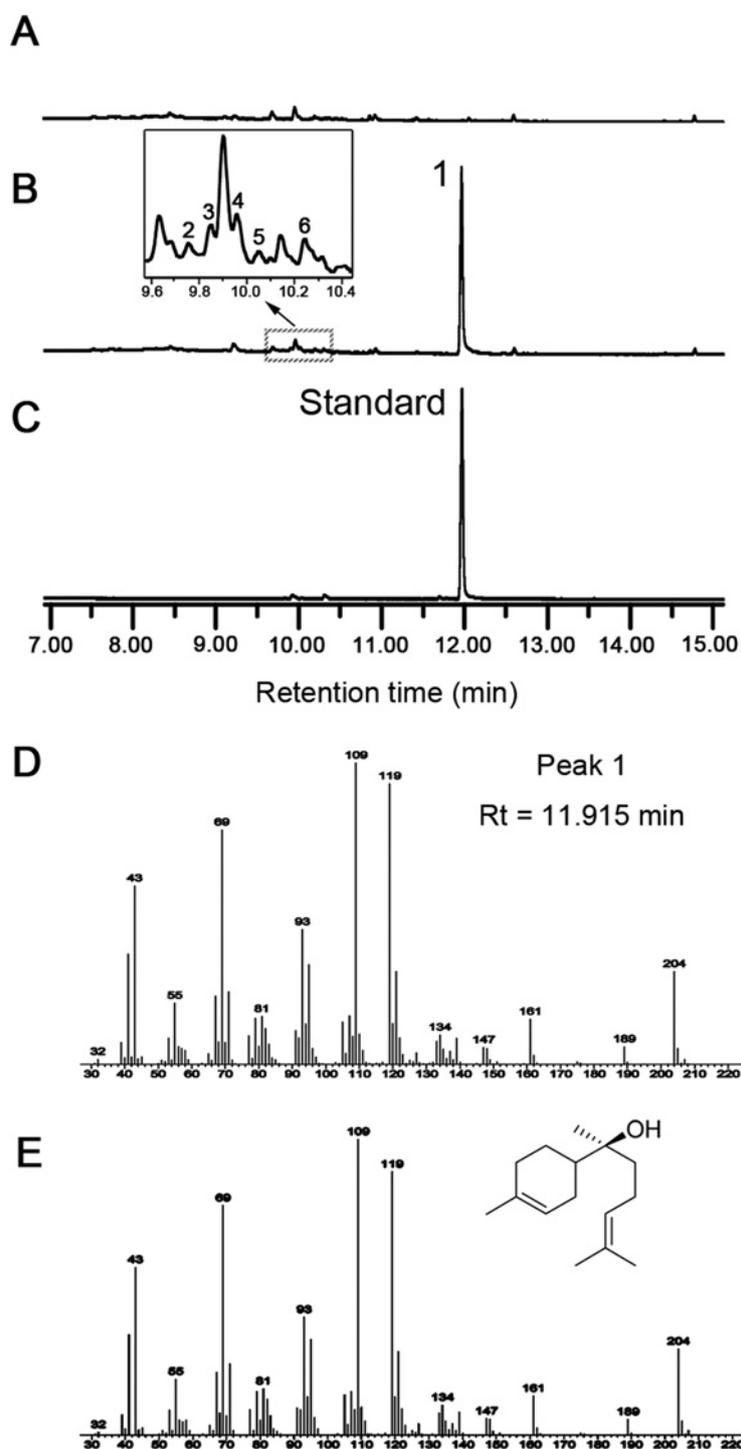


Figure 2 GC-MS analysis of *in vitro* assay products of AaBOS

(A–C) Total ion chromatogram of assay products with proteins of the empty pET32a vector (A), the recombinant AaBOS (B) and the α -bisabolol authentic standard (C). The GC expansion shows peaks of minor products: 2, *trans*- α -bisabolene; 3, β -bisabolene; 4, *cis*- γ -bisabolene; 5, β -sesquiphellandrene and 6, *cis*- α -bisabolene. (D and E) Mass spectra of the major product (peak 1) of AaBOS (D) and of α -bisabolol authentic standard (E).

for 92.8% of the total products, hence the enzyme was named AaBOS. Additionally, five minor sesquiterpenes were detected, they were *trans*- α -bisabolene (1.1%), *cis*- γ -bisabolene (2.4%), *cis*- α -bisabolene (1.2%), β -bisabolene (1.9%) and β -sesquiphellandrene (0.51%) (Supplementary Figures S2 and S3 at <http://www.biochemj.org/bj/451/bj4510417add.htm>).

α -Bisabolene and β -bisabolene are among the reported components of *A. annua* essential oil [30], and α -bisabolene and β -sesquiphellandrene are also the minor products of AaADS [28]. Kinetic analysis with FPP showed that AaBOS has a Michaelis constant (K_m) of 17.11 μ M, k_{cat} of 1.55 s^{-1} , a V_{max} of $4.64 \times 10^{-3} \mu M \cdot s^{-1} \cdot mg^{-1}$ and a specific constant (k_{cat}/K_m)

Table 2 Kinetic parameters of AaBOS, AaADS and the AaADS_{T399S} mutant using FPP as a substrate

Kinetic parameter	AaBOS	AaADS	AaADS _{T399S}
K_m (μM)	17.11	3.56	3.79
k_{cat} (s^{-1})	1.55	1.37	2.51
V_{max} ($\mu\text{M} \cdot \text{s}^{-1} \cdot \text{mg}^{-1}$)	4.64×10^{-3}	4.12×10^{-3}	7.53×10^{-3}
k_{cat}/K_m ($\mu\text{M}^{-1} \cdot \text{s}^{-1}$)	0.0905	0.385	0.662

value of $9.05 \times 10^{-2} \mu\text{M}^{-1} \cdot \text{s}^{-1}$, approximately 4-fold less active than AaADS (Table 2).

The presence of by-products strongly suggests that α -bisabolol is derived from the intermediate bisabolyl cation generated from the *E,Z*-farnesyl cation through electrophilic 1,6-ring closure (Scheme 1). AaBOS promotes water capture of the bisabolyl cation to generate and release α -bisabolol, while deprotonation from C-6, C-8 and C-14 yields *trans*- α -bisabolene (4), *cis*- γ -bisabolene (1), *cis*- α -bisabolene (2) and β -bisabolene (3), and β -sesquiphellandrene is most probably generated through 1,3-hydride shift and subsequent proton-elimination at C-15.

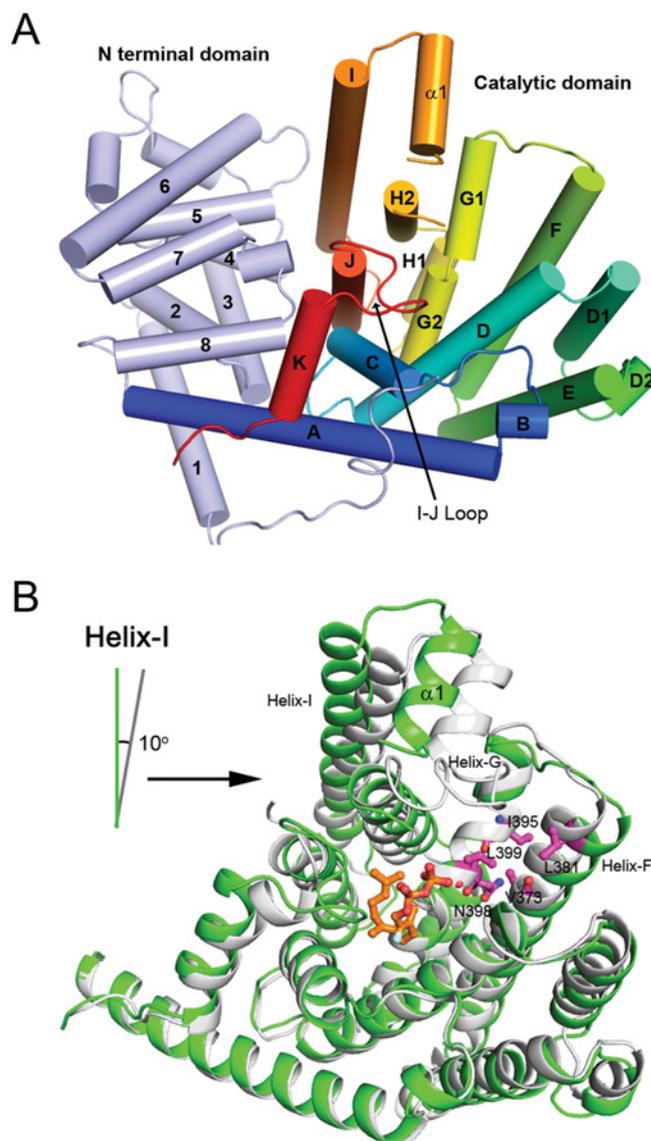
Structure of AaBOS

To disclose the molecular mechanism of product generation, AaBOS was crystallized and its structure solved to 2.0 Å. The final model contained residues 17–546; residues 449–457, which are part of the H3 helix, were not visible in the structure (Table 1). The structure was in apo form, with no metal ion binding despite 2 mM Mg^{2+} being present during the purification and crystallization steps. Co-crystallization of AaBOS with FPP, 2F-FPP, 6F-FPP and FHP, as well as soaking of the crystals of apo-AaBOS with these ligands were attempted, but none of the structures solved contained the above substrates. AaBOS adopts a similar three-dimensional structure to that typically found in plant sesquiterpene synthases: there are two compact α -helical domains, the N-terminal domain and the C-terminal catalytic domain (Figure 3A). Comparison of the AaBOS structure with that of NtEAS showed r.m.s.d. (root mean square deviation) values of 2.0 Å. The major differences are found in the C-terminal domain; superimposition the structures of AaBOS and NtEAS revealed that all of the helices (helix A to helix K) except for helix I and the short $\alpha 1$ helix (connecting helix H and I) fit very well. The axes of helix I and helix $\alpha 1$ rotate approximately 10° away from helix G around a hinge formed by the I–J loop (Figure 3B). In AaBOS and NtEAS, the DDxxD motifs are in helix D, but the NSE/DTE motif of NtEAS is replaced by NGE and the segment is partially disordered in AaBOS, possibly due to the absence of a bound substrate analogue and/or metal ions. This observation suggests that the NGE motif plays a subtle role in binding substrate and metal ions (Supplementary Figure S1).

Identification of AaBOS regions involved in product specificity

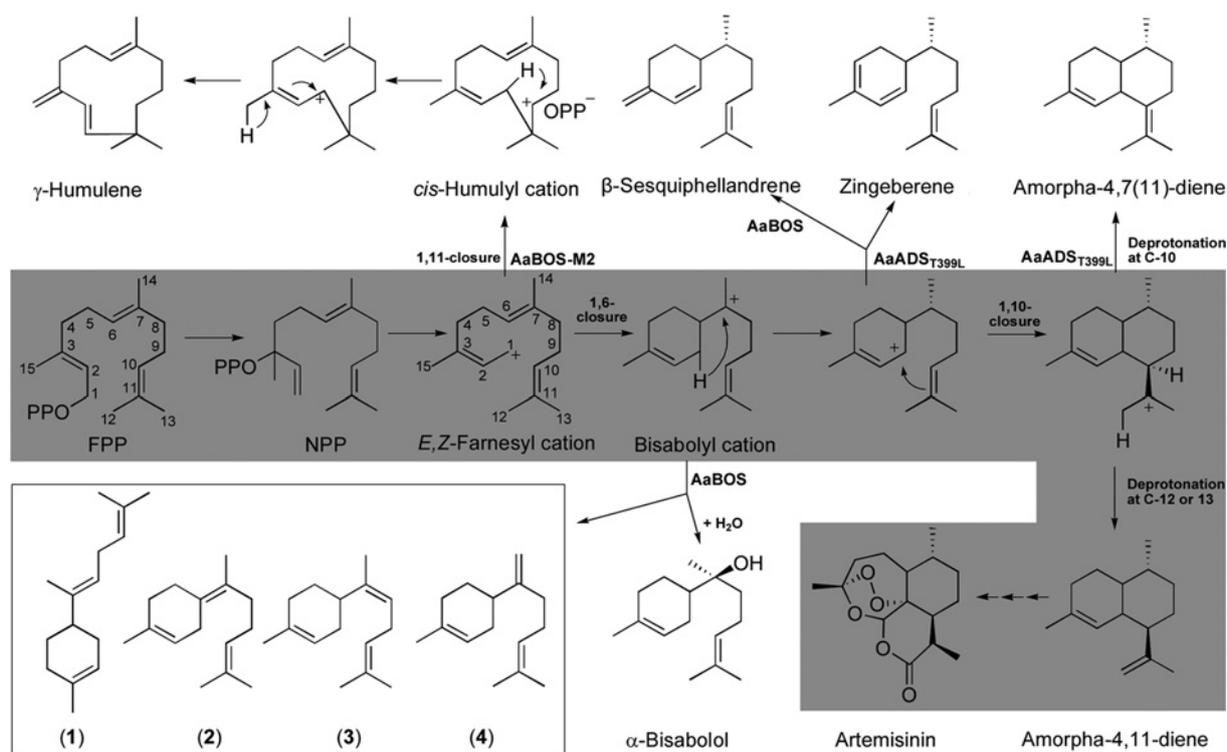
AaBOS has a high amino acid sequence identity (82%) with AaADS, and they may share the bisabolyl cation as common carbocation intermediate (Scheme 1). The similarities in sequence and mechanism suggest that a comparative analysis of the two enzymes could reveal key residues for the activity of AaADS.

We applied a domain-swapping strategy to localize regions related to the distinctive products of AaBOS and AaADS. The proteins were divided into three fragments (I, II and III), an initial set of chimaeric enzymes were constructed with reciprocal

**Figure 3** The structure of AaBOS

(A) The overall structure of AaBOS as a cylinder cartoon. (B) Superimposition of AaBOS (green) and NtEAS (grey) defines the active-site cavity with the substrate analogue 2F-FPP from NtEAS (PDB code 3M01) shown as a golden stick. The side chains of Ile³⁹⁵, Asn³⁹⁸, Leu³⁹⁹, Val³⁷³ and Leu³⁸¹ in AaBOS are shown in magenta.

replacement of the corresponding fragments, and the resultant activities were measured to localize fragments containing the sites essential for activity. The active fragments were further divided into sub-segments and were reciprocally substituted until the minimum segment or residues mediating product specificity were located. After three cycles, in which fragments II and III were divided into sub-fragments II.1, II.2, II.3, III.1 and III.2, and then II.3 into II.3.1 and II.3.2 (Figure 4 and Supplementary Figure S1), we generated mutant AaBOS enzymes that produced an additional product, which was identified as γ -humulene (Figure 4B and Supplementary Figure S4 at <http://www.biochemj.org/bj/451/bj4510417add.htm>). Two continuous fragments, II.3.1 (residues 343–381, located in helix F) and II.3.2 (residues 392–399, located in helix G1) were found to be indispensable (Figure 4C; CH9, CH12 and CH13) for the formation of γ -humulene. Further site-directed mutagenesis



Scheme 1 Proposed mechanisms for the conversion of FPP into sesquiterpene products catalysed by AaBOS, AaADS and their mutants

The scheme is based on previous studies of biosynthesis pathways of terpenes [7] and artemisinin [28,30,31,33] (the grey area denotes the artemisinin biosynthesis pathway). The numbered products are: *cis*- γ -bisabolene (1), *cis*- α -bisabolene (2), β -bisabolene (3), *trans*- α -bisabolene (4).

led to the identification of several residues responsible for the product specificity of AaBOS. A tetra-substituted mutant AaBOS-M1 (V373N, I395V, N398I and L399T) and a penta-substituted mutant AaBOS-M2 (V373N, L381A, I395V, N398I and L399T) yielded 29.3 and 68.8 % of γ -humulene respectively (Figure 4C). Interestingly, γ -humulene is also a minor product of AaADS [28]. We propose that the formation of γ -humulene could be the result of 1,11-ring closure resulting from attack of the C-10, C-11 π -bond by C-1 in an anti-Markovnikov fashion (Scheme 1).

The BOS motif is closely related to the product specificity of AaBOS

Although the five residues identified in AaBOS-M2 may play different roles in the production of γ -humulene, all γ -humulene-producing mutants harboured the L399T substitution, suggesting an important role for this substitution in determining product specificity of AaBOS. In particular, although the CH12-M4 and AaBOS-M2 mutants containing all five substitutions produced γ -humulene, the CH12-M16 which lacked L399T did not (Figure 4C), strongly supporting this notion. However, none of the single-residue substitutions changed the product spectrum of AaBOS (Supplementary Table S2 at <http://www.biochemj.org/bj/451/bj4510417add.htm>), indicating a synergistic effect of these residues. To examine the effect of these residues on product specificity, we crystallized and solved the structure of AaBOS-M2, which produced 68.8 % of γ -humulene. The overall structure of AaBOS-M2 is nearly identical with that of AaBOS, except for a conformational change caused by a slight counter-clockwise rotation of helix G1 and F as a rigid body towards the vertical axis (Figure 5A). In the two structures, four of the above mentioned residues, Val³⁷³, Ile³⁹⁵, Asn³⁹⁸ and Leu³⁹⁹, are proximal to each

other and constitute a motif, which we named the BOS motif. Val³⁷³ from helix F is on one side of the BOS motif and Ile³⁹⁵, Asn³⁹⁸ and Leu³⁹⁹ from helix G1 are on the other side. Residue 381, located at the C-terminus of helix F, protrudes into the substrate-binding pocket and away from the BOS motif (Figures 3B and 5).

Although the conformational change found in the AaBOS-M2 structure resulted in a minor volume change of the substrate-binding pocket, alteration in the shape of the pocket is substantial. This alteration may have caused the variation in folding of the substrate. In combination with the observed indispensable role of Thr³⁹⁹ in affecting product specificity of AaBOS, we postulated that Thr³⁹⁹, probably the side-chain hydroxy group, is involved in the 1,11-ring closure and production of the *cis*-humulyl cation (Scheme 1). However, without the help of other mutations in the BOS motif in altering the pocket shape, the hydroxy group of Thr³⁹⁹ may be away from the functional position and could not exert any effect in directing 1,11-ring closure, which could explain why all of the single-residue substitutions failed to change the product spectrum of AaBOS.

To confirm this finding, the four residues were further replaced by the corresponding ones in AaFS (*A. annua* β -farnesene synthase) (NVVT) [27] and AaGAS (*A. annua* germacrene A synthase) (GNIT) [29] respectively. Both of the mutants (AaBOS-M11 and AaBOS-M12) produced α -bisabolol, together with γ -humulene as a minor product, accounting for approximately 13.7 and 8.3 % of the total products respectively (Supplementary Table S2), again supporting the notion that the BOS motif, composed of plasticity residues, plays an important role in determining the product specificity of AaBOS.

The side chain of Ala³⁸¹/Leu³⁸¹ is located on the rim of exit or entrance of the pocket (Figure 5B) and may exert an effect on the release of product. The large size of the side chain of leucine

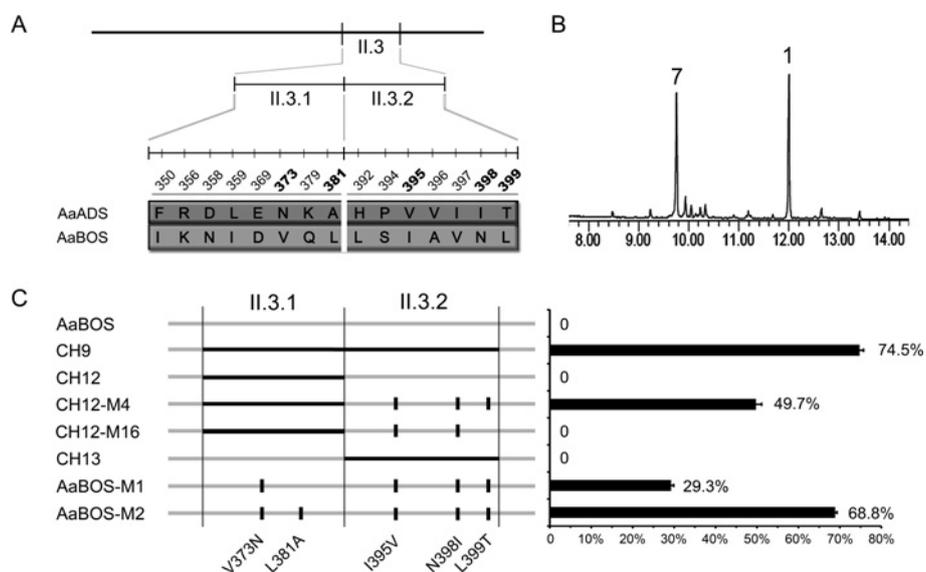


Figure 4 Identification of plasticity residues

(A) Schematic diagram showing amino acid residues that are different between AaBOS and AaADS in fragment II.3. (B) Total ion chromatogram of products of the substituted mutant CH12-M4 of AaBOS. Peaks of products: 1, α -bisabolol; 7, γ -humulene. (C) Schematic view of site-directed substitutions of AaBOS (left-hand panel) and the respective enzyme activities, showing the percentage of γ -humulene in total products (right-hand panel).

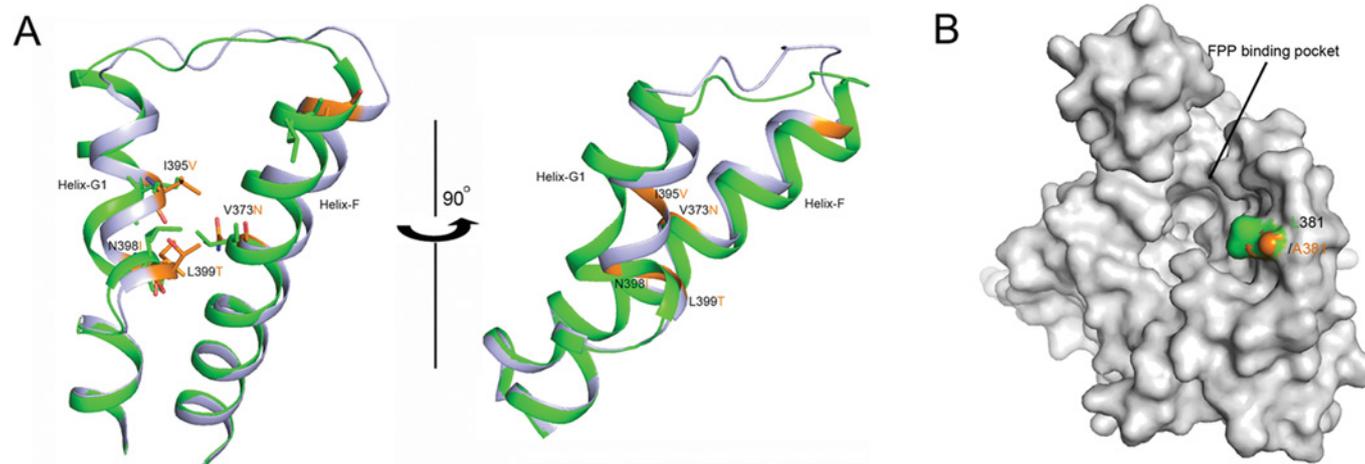


Figure 5 The activity centre structures of AaBOS and AaBOS-M2

(A) Structure comparison of the BOS motif in AaBOS and AaBOS-M2. Helix F and helix G1 of AaBOS are shown in green, whereas those of AaBOS-M2 are in purple. Side chains constituting the BOS motif and residue 381 are shown with green/golden sticks in AaBOS/AaBOS-M2. (B) Molecular surface representation shows the FPP-binding pocket and the side chain of Ala³⁸¹/Leu³⁸¹ located on the rim of the exit or entrance.

may hinder the release of γ -humulene, but not α -bisabolol as α -bisabolol is more linear and flexible than γ -humulene, whereas alanine with a small side chain may have no bias towards these two compounds. Thus the replacement of leucine with alanine may relieve the bias and release more γ -humulene over α -bisabolol. Indeed, AaBOS-M2 with an additional mutation of L381A exhibited an approximate doubling in production of γ -humulene compared with AaBOS-M1 in the assay (Figure 4C).

Role of Thr³⁹⁹ in AaADS catalytic efficiency

With the information on plasticity residues of AaBOS at hand, site-directed mutagenesis was performed on the BOS motif region

of AaADS to explore the catalytic role of the corresponding residues in AaADS. Although most of the mutants were inactive or less active than the wild-type enzyme, one of the single mutants, T399L (AaADS_{T399L}), exhibited an altered product specificity (Figure 6A, and Supplementary Table S3 at <http://www.biochemj.org/bj/451/bj4510417add.htm>). Besides the product amorpha-4,11-diene, AaADS_{T399L} produced a high amount of amorpha-4,7(11)-diene and low amounts of β -sesquiphellandrene and zingiberene (Figure 6A and Supplementary Figure S3). The formation of amorpha-4,7(11)-diene might be the result of proton extraction at C-10, instead of C-12/C-13 as that of amorpha-4,11-diene, and nearly equal yields of amorpha-4,11-diene and amorpha-4,7(11)-diene suggests the loss of regioselectivity of deprotonation. A premature quenching of the C-1 bisabolyl cation

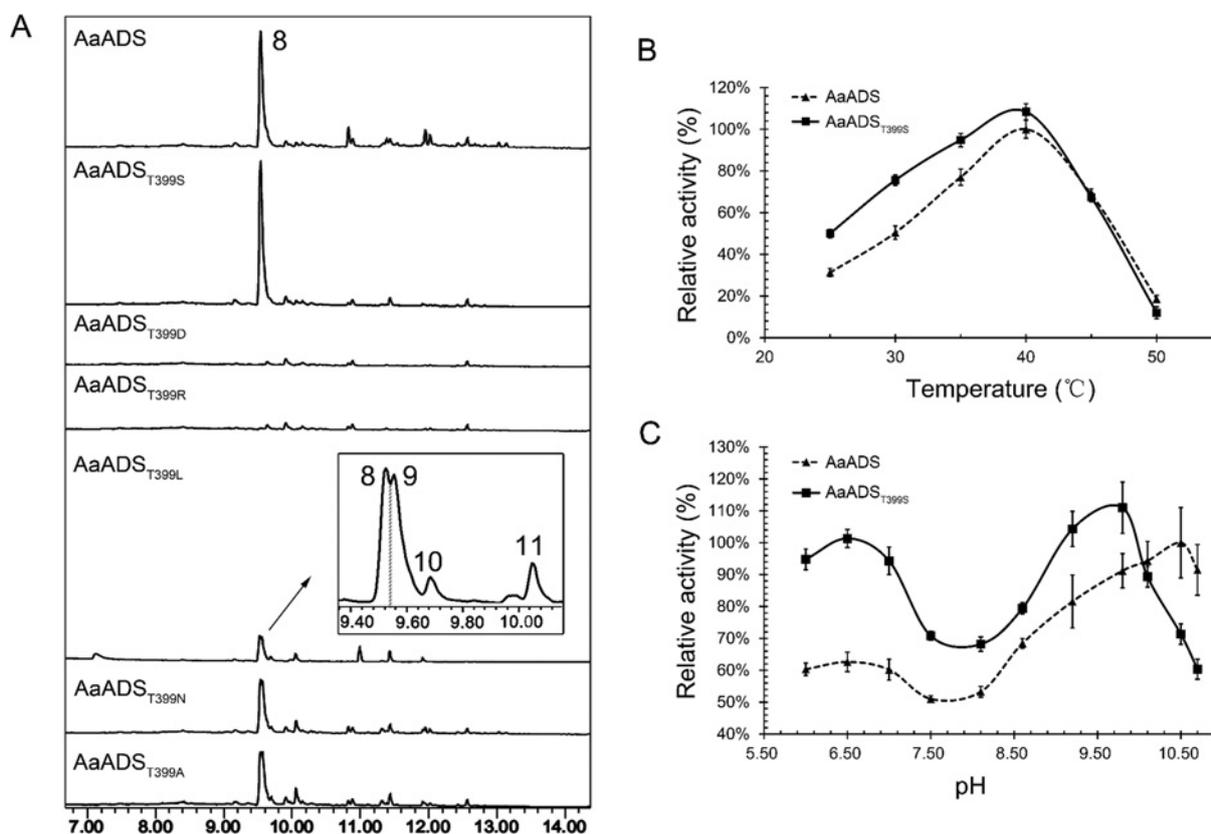


Figure 6 Characterization of AaADS and mutants

GC-MS analysis of replacing of AaADS Thr³⁹⁹ with acidic, alkaline and neutral amino acids (**A**) and comparison of AaADS and AaADS_{T399S} activities under different temperature (**B**) and pH (**C**) conditions. (**A**) Total ion chromatograms of products of AaADS and mutant enzymes as indicated. The GC expansion shows peaks produced by AaADS_{T399L}: 8, amorpho-4,11-diene; 9, amorpho-4,7(11)-diene; 10, zingiberene; 11, β -sesquiphellandrene. (**B** and **C**) The relative activity of the wild-type and the mutant enzymes, shown as a percentage of the maximum activity of the wild-type AaADS for amorpho-4,11-diene.

generates the products of β -sesquiphellandrene and zingiberene (Scheme 1).

Position 399 in AaADS, important for catalytic specificity, was further replaced with acidic, alkaline and neutral amino acids. Replacing with aspartic acid (AaADS_{T399D}) or arginine (AaADS_{T399R}) resulted in loss of catalytic activity, whereas a neutral residue substitution (AaADS_{T399A} or AaADS_{T399N}) led to product distributions similar to that of AaADS_{T399L} (Figure 6A). Strikingly, when the threonine residue was replaced with the similar, but smaller, amino acid serine (AaADS_{T399S}), the yield of amorpho-4,11-diene was increased by 70%, and its proportion also increased from 92.0 to 93.9% of the total products (Figure 6A). The kinetic parameters of AaADS_{T399S} were analysed with steady-state measurements. Compared with the wild-type, AaADS_{T399S} showed higher turnover rate of FPP and higher activity at pH 7.0. k_{cat} is increased by approximately 83%, and the k_{cat}/K_m value is increased by 71.7% (Table 2). Clearly, AaADS_{T399S} is a more efficient enzyme than the wild-type. Under different temperature and pH conditions, AaADS_{T399S} showed varied performances compared with the wild-type enzyme. In the temperature range 25–40°C, AaADS_{T399S} had a higher activity than AaADS; when the temperature was raised to above 45°C, both enzymes rapidly lost activity (Figure 6B). Both AaADS and AaADS_{T399S} have two optimal pH values under acidic (pH 6.5) and alkaline (pH 9.8 and pH 10.5) conditions; mutation of threonine to serine shifted the alkaline optimal pH of the enzyme from 10.5 to 9.8. Under conditions of pH 5.5–9.8, AaADS_{T399S} was generally more active than the wild-type enzyme, but at

exceptionally higher pH values, AaADS_{T399S} lost its activity more rapidly (Figure 6C). These data indicate that the T399S mutation greatly improved the enzymatic activity under a wide range of temperature and pH conditions.

Considering the location of residue 399 in the substrate-binding pocket in AaBOS, it is reasonable to assume that the presence and absence of the hydroxy group in the side chain of amino acid 399 contributes to the catalytic specificity in AaADS. The hydroxy group of threonine might be involved in regioselective deprotonation, as mutations of Thr³⁹⁹ with other amino acid residues (except for serine which also has a hydroxy group) generated an additional product arising from different ways of proton loss at C-10. Thus the yield of amorpho-4,7(11)-diene (Scheme 1) could be the result of the loss of regioselectivity of deprotonation. Furthermore, this change of mechanism may lead to the decrease in the efficiency of proton elimination and the resultant accumulation of intermediates, which is supported by the increased amount of by-products of β -sesquiphellandrene and zingiberene. Notably, AaADS and its functionally normal mutant AaADS_{T399S}, the aforementioned γ -humulene-producing mutants of AaBOS, as well as AaFS and AaGAS, all contain threonine/serine at position 399. Their hydroxy group side chains could take part in catalytic processes. By contrast, AaBOS has a neutral amino acid, leucine, which may not directly participate in catalytic reactions. It is possible that the mutation of T399L of the BOS motif resulted in neofunctionalization of an additional copy of AaADS or its ancestral progenitor and gave rise to AaBOS later during evolution. We searched 140 plant TPSs with known

function, among which 120 have threonine/serine at position 399, suggesting that it is highly, but not invariably, conserved with the two alcoholic amino acids. We postulate that this position is of wide-ranging importance for catalysis by plant TPSs. It is an important finding that a single substitution (T399S) increased the catalytic efficiency of AaADS. As the hydroxy group of threonine/serine is likely to be involved in deprotonation, the final step of the whole catalytic reaction, a change from threonine to serine may favour the deprotonation process. Additionally, it might influence the hydrophilic environment of the pocket from which the final product is released. Compared with threonine, serine has a more polar side chain affording a more hydrophilic environment, thus the hydrophobic product may be released more rapidly. Replacing the threonine residue with a similar, but smaller, amino acid (serine) could maintain the catalytic mechanism without a profound change of the product profile, whereas previous research in plasticity residues has generally led to change in mechanism and shifted product specificity.

Concluding remarks

TPSs tend to develop catalytic elasticity during evolution [12], which is also the case in AaBOS and AaADS. AaBOS has a high nucleotide sequence identity (92%) with and similar catalytic mechanism to AaADS, implying that they are closely related in evolution. This is shown by only a few residue changes for divergence to occur from the ancestral enzyme. The findings of the present study are in agreement with previous reports which suggested that the new function of an existing enzyme is often the result of a relatively low number of mutations [8,12,15,47]. The observation that TPSs tend to evolve new functions rather than improving efficiency for currently recognized substrates supports the proposed role in ecology, i.e. plant terpenoid profiles can rapidly change in response to environmental influence. Moreover, this malleable characteristic has allowed us to change a TPS with a single threonine to serine mutation and profoundly increase its catalytic efficiency.

At present, extraction of *A. annua* plants is still the major source of artemisinin. Many efforts have been made to engineer artemisinin-related production in both microbes and plants [6,48]. By introducing the mevalonate and artemisinin metabolic pathway enzymes into *E. coli* or *Saccharomyces cerevisiae*, artemisinic acid has been produced by fermentation [6,31]. Engineering more efficient enzymes, such as the AaADS_{T399S} reported in the present paper, is expected to further facilitate the synthetic biology approach to artemisinin production. In addition, the natural sesquiterpene alcohol α -bisabolol is widely used in the fragrance and cosmetic industry owing to its anti-inflammatory, analgesic, antibiotic and probably anticancer properties [49]. A recent report demonstrated a novel biosynthetic alternative to D2 diesel fuel, bisabolane, whose precursor bisabolene could be produced by microbial engineering [50]. Through simple chemical reactions, α -bisabolol also can yield the terpene-based biofuel bisabolane. Cloning of α -bisabolol synthase opens up a new avenue to the production of this sesquiterpene alcohol in a large scale via metabolic engineering.

AUTHOR CONTRIBUTION

Jian-Xu Li performed the gene cloning, protein expression and purification, and the enzymatic assay. Xin Fang and Jian-Xu Li did the product mass spectra identification and contributed to writing the paper. Jian-Xu Li, Qin Zhao and Peng Zhang crystallized the proteins and determined the structures. Ju-Xin Ruan, Chang-Qing Yang and Ling-Jian Wang participated in the experiments. David Miller, Juan Faraldos and Rudolf Allemann

contributed to data analysis and discussion. Xiao-Ya Chen and Peng Zhang designed the experiments and wrote the paper.

ACKNOWLEDGEMENTS

We thank Professor Peter E. Brodelius for compound identification, Dr Youli Xiao and Dr Yuan-Heng Cai for discussion, and staff members at the SSRF (Shanghai Synchrotron Radiation Facility) for technical assistance in data collection.

FUNDING

This work was supported by the Ministry of Science and Technology of China [grant number 2007AA021501], the National Natural Science Foundation of China [grant number 30630008], the Shanghai Institutes for Biological Sciences [grant number 2012OHTP], the Shanghai "Pujiang talent" programme [grant number 11PJ1411300], the UK Biotechnology and Biological Sciences Research Council (BBSRC) [grant number BB/G003572/1], the UK Engineering and Physical Sciences Research Council (EPSRC) [grant number EP/D06958/1] and Cardiff University.

REFERENCES

- Tholl, D. (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr. Opin. Plant Biol.* **9**, 297–304
- Schmid, G. and Hofheinz, W. (1983) Total synthesis of qinghaosu. *J. Am. Chem. Soc.* **105**, 624–625
- Nicolaou, K. C., Yang, Z., Liu, J. J., Ueno, H., Nantermet, P. G., Guy, R. K., Claiborne, C. F., Renaud, J., Couladouros, E. A., Paulvannan, K. and Sorensen, E. J. (1994) Total synthesis of taxol. *Nature* **367**, 630–634
- Snow, R. W., Guerra, C. A., Noor, A. M., Myint, H. Y. and Hay, S. I. (2005) The global distribution of clinical episodes of *Plasmodium falciparum* malaria. *Nature* **434**, 214–217
- White, N. J. (2008) Qinghaosu (artemisinin): the price of success. *Science* **320**, 330–334
- Keasling, J. D. (2008) Synthetic biology for synthetic chemistry. *ACS Chem. Biol.* **3**, 64–76
- Christianson, D. W. (2006) Structural biology and chemistry of the terpenoid cyclases. *Chem. Rev.* **106**, 3412–3442
- Kampranis, S. C., Ioannidis, D., Purvis, A., Mahrez, W., Ninga, E., Katerelos, N. A., Anssour, S., Dunwell, J. M., Degenhardt, J., Makris, A. M. et al. (2007) Rational conversion of substrate and product specificity in a *Salvia* monoterpene synthase: structural insights into the evolution of terpene synthase function. *Plant Cell* **19**, 1994–2005
- Bar-Even, A., Noor, E., Savir, Y., Liebermeister, W., Davidi, D., Tawfik, D. S. and Milo, R. (2011) The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410
- Miller, D. J. and Allemann, R. K. (2012) Sesquiterpene synthases: passive catalysts or active players? *Nat. Prod. Rep.* **29**, 60–71
- Rising, K. A., Starks, C. M., Noel, J. P. and Chappell, J. (2000) Demonstration of germacrene A as an intermediate in 5-epi-aristolochene synthase catalysis. *J. Am. Chem. Soc.* **122**, 1861–1866
- Weng, J. K., Philippe, R. N. and Noel, J. P. (2012) The rise of chemodiversity in plants. *Science* **336**, 1667–1670
- Starks, C. M., Back, K., Chappell, J. and Noel, J. P. (1997) Structural basis for cyclic terpene biosynthesis by tobacco 5-epi-aristolochene synthase. *Science* **277**, 1815–1820
- 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. Natl. Acad. Sci. U.S.A.* **103**, 9826–9831
- Yoshikuni, Y., Ferrin, T. E. and Keasling, J. D. (2006) Designed divergent evolution of enzyme function. *Nature* **440**, 1078–1082
- Xu, M., Wilderman, P. R. and Peters, R. J. (2007) Following evolution's lead to a single residue switch for diterpene synthase product outcome. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7397–7401
- Katoh, S., Hyatt, D. and Croteau, R. (2004) Altering product outcome in *Abies grandis* (–)-limonene synthase and (–)-limonene/(–)- α -pinene synthase by domain swapping and directed mutagenesis. *Arch. Biochem. Biophys.* **425**, 65–76
- Aubourg, S., Lechary, A. and Bohlmann, J. (2002) Genomic analysis of the terpenoid synthase (*AtTPS*) gene family of *Arabidopsis thaliana*. *Mol. Genet. Genomics* **267**, 730–745
- Kollner, T. G., Schnee, C., Gershenzon, J. and Degenhardt, J. (2004) The variability of sesquiterpene cultivars is controlled by allelic variation of two terpene synthase genes encoding stereoselective multiple product enzymes. *Plant Cell* **16**, 1115–1131

- 20 Peters, R. J. and Croteau, R. B. (2002) Abietadiene synthase catalysis: mutational analysis of a prenyl diphosphate ionization-initiated cyclization and rearrangement. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 580–584
- 21 Wilderman, P. R. and Peters, R. J. (2007) A single residue switch converts abietadiene synthase into a pimaradiene specific cyclase. *J. Am. Chem. Soc.* **129**, 15736–15737
- 22 Back, K. W. and Chappell, J. (1996) Identifying functional domains within terpene cyclases using a domain-swapping strategy. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 6841–6845
- 23 O'Maille, P. E., Malone, A., Dellas, N., Hess, B. A., Smentek, L., Sheehan, I., Greenhagen, B. T., Chappell, J., Manning, G. and Noel, J. P. (2008) Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat. Chem. Biol.* **4**, 617–623
- 24 Bouwmeester, H. J., Wallaart, T. E., Janssen, M. H., van Loo, B., Jansen, B. J., Posthumus, M. A., Schmidt, C. O., De Kraker, J. W., Konig, W. A. and Franssen, M. C. (1999) Amorpha-4,11-diene synthase catalyses the first probable step in artemisinin biosynthesis. *Phytochemistry* **52**, 843–854
- 25 Chang, Y. J., Song, S. H., Park, S. H. and Kim, S. U. (2000) Amorpha-4,11-diene synthase of *Artemisia annua*: cDNA isolation and bacterial expression of a terpene synthase involved in artemisinin biosynthesis. *Arch. Biochem. Biophys.* **383**, 178–184
- 26 Cai, Y., Jia, J. W., Crock, J., Lin, Z. X., Chen, X. Y. and Croteau, R. (2002) A cDNA clone for β -caryophyllene synthase from *Artemisia annua*. *Phytochemistry* **61**, 523–529
- 27 Picaud, S., Brodelius, M. and Brodelius, P. E. (2005) Expression, purification and characterization of recombinant (*E*)- β -farnesene synthase from *Artemisia annua*. *Phytochemistry* **66**, 961–967
- 28 Picaud, S., Olofsson, L., Brodelius, M. and Brodelius, P. E. (2005) Expression, purification, and characterization of recombinant amorpha-4,11-diene synthase from *Artemisia annua* L. *Arch. Biochem. Biophys.* **436**, 215–226
- 29 Berteau, C. M., Voster, A., Verstappen, F. W., Maffei, M., Beekwilder, J. and Bouwmeester, H. J. (2006) Isoprenoid biosynthesis in *Artemisia annua*: cloning and heterologous expression of a germacrene A synthase from a glandular trichome cDNA library. *Arch. Biochem. Biophys.* **448**, 3–12
- 30 Brown, G. D. (2010) The biosynthesis of artemisinin (Qinghaosu) and the phytochemistry of *Artemisia annua* L. (Qinghao). *Molecules* **15**, 7603–7698
- 31 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
- 32 Teoh, K. H., Polichuk, D. R., Reed, D. W., Nowak, G. and Covello, P. S. (2006) *Artemisia annua* L. (Asteraceae) trichome-specific cDNAs reveal CYP71AV1, a cytochrome P450 with a key role in the biosynthesis of the antimalarial sesquiterpene lactone artemisinin. *FEBS Lett.* **580**, 1411–1416
- 33 Zhang, Y., Teoh, K. H., Reed, D. W., Maes, L., Goossens, A., Olson, D. J. H., Ross, A. R. S. and Covello, P. S. (2008) The molecular cloning of artemisinic aldehyde Δ 11(13) reductase and its role in glandular trichome-dependent biosynthesis of artemisinin in *Artemisia annua*. *J. Biol. Chem.* **283**, 21501–21508
- 34 Yu, Z. X., Li, J. X., Yang, C. Q., Hu, W. L., Wang, L. J. and Chen, X. Y. (2012) The jasmonate-responsive AP2/ERF transcription factors AaERF1 and AaERF2 positively regulate artemisinin biosynthesis in *Artemisia annua* L. *Mol. Plant* **5**, 353–365
- 35 Faraldos, J. A., Zhao, Y., O'Maille, P. E., Noel, J. P. and Coates, R. M. (2007) Interception of the enzymatic conversion of farnesyl diphosphate to 5-epi-aristolochene by using a fluoro substrate analogue: 1-fluorogermacrene A from (2*E*,6*Z*)-6-fluorofarnesyl diphosphate. *ChemBioChem* **8**, 1826–1833
- 36 Miller, D. J., Yu, F. and Allemann, R. K. (2007) Aristolochene synthase-catalyzed cyclization of 2-fluorofarnesyl-diphosphate to 2-fluorogermacrene A. *ChemBioChem* **8**, 1819–1825
- 37 Yu, N., Cai, W. J., Wang, S., Shan, C. M., Wang, L. J. and Chen, X. Y. (2010) Temporal control of trichome distribution by microRNA156-targeted SPL genes in *Arabidopsis thaliana*. *Plant Cell* **22**, 2322–2335
- 38 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) Clustal-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680
- 39 Gouet, P., Courcelle, E., Stuart, D. I. and Metz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305–308
- 40 Aiyar, A., Xiang, Y. and Leis, J. (1996) Site-directed mutagenesis using overlap extension PCR. *Methods Mol. Biol.* **57**, 177–191
- 41 Chen, X. Y., Chen, Y., Heinstein, P. and Davison, V. J. (1995) Cloning, expression, and characterization of (+)- δ -cadinene synthase: a catalyst for cotton phytoalexin biosynthesis. *Arch. Biochem. Biophys.* **324**, 255–266
- 42 Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326
- 43 Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763
- 44 Noel, J. P., Dellas, N., Faraldos, J. A., Zhao, M., Hess, B. A., Smentek, L., Coates, R. M. and O'Maille, P. E. (2010) Structural elucidation of cisoid and transoid cyclization pathways of a sesquiterpene synthase using 2-fluorofarnesyl diphosphates. *ACS Chem. Biol.* **5**, 377–392
- 45 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132
- 46 Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 1948–1954
- 47 Tokuriki, N. and Tawfik, D. S. (2009) Protein dynamism and evolvability. *Science* **324**, 203–207
- 48 Farhi, M., Marhevka, E., Ben-Ari, J., Algamas-Dimantov, A., Liang, Z. B., Zeevi, V., Edelbaum, O., Spitzer-Rimon, B., Abeliovich, H., Schwartz, B. et al. (2011) Generation of the potent anti-malarial drug artemisinin in tobacco. *Nat. Biotechnol.* **29**, 1072–1074
- 49 Kamatou, G. P. P. and Viljoen, A. M. (2010) A review of the application and pharmacological properties of α -bisabolol and α -bisabolol-rich oils. *J. Am. Oil Chem. Soc.* **87**, 1–7
- 50 Peralta-Yahya, P. P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J. D. and Lee, T. S. (2011) Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* **2**, 483

Received 8 January 2013/20 February 2013; accepted 26 February 2013

Published as BJ Immediate Publication 26 February 2013, doi:10.1042/BJ20130041



SUPPLEMENTARY ONLINE DATA

Rational engineering of plasticity residues of sesquiterpene synthases from *Artemisia annua*: product specificity and catalytic efficiency

Jian-Xu LI*^{†1}, Xin FANG*¹, Qin ZHAO*[†], Ju-Xin RUAN*[†], Chang-Qing YANG*, Ling-Jian WANG*, David J. MILLER[‡], Juan A. FARALDOS[‡], Rudolf K. ALLEMANN[‡], Xiao-Ya CHEN*^{§2} and Peng ZHANG*^{||2}

*National Key Laboratory of Plant Molecular Genetics and National Plant Gene Research Center, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China, [†]Graduate School of Chinese Academy of Sciences, Beijing 100049, China, [‡]School of Chemistry and Cardiff Catalysis Institute, Cardiff University, Main Building, Park Place, Cardiff CF10 3AT, U.K., [§]Plant Science Research Center, Shanghai Chenshan Botanical Garden, Shanghai 201602, China, and ^{||}Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, 300 Fenglin Road, Shanghai 200032, China

See the following pages for Supplementary Figures S1–S4 and Supplementary Tables S1–S3.

¹ These authors contributed equally to this work.

² Correspondence may be addressed to either of these authors (email xychen@sibs.ac.cn or pengzhang01@sibs.ac.cn). Atomic co-ordinates have been deposited in the PDB under accession codes 4FJQ for AaBOS and 4GAX for AaBOS-M2.

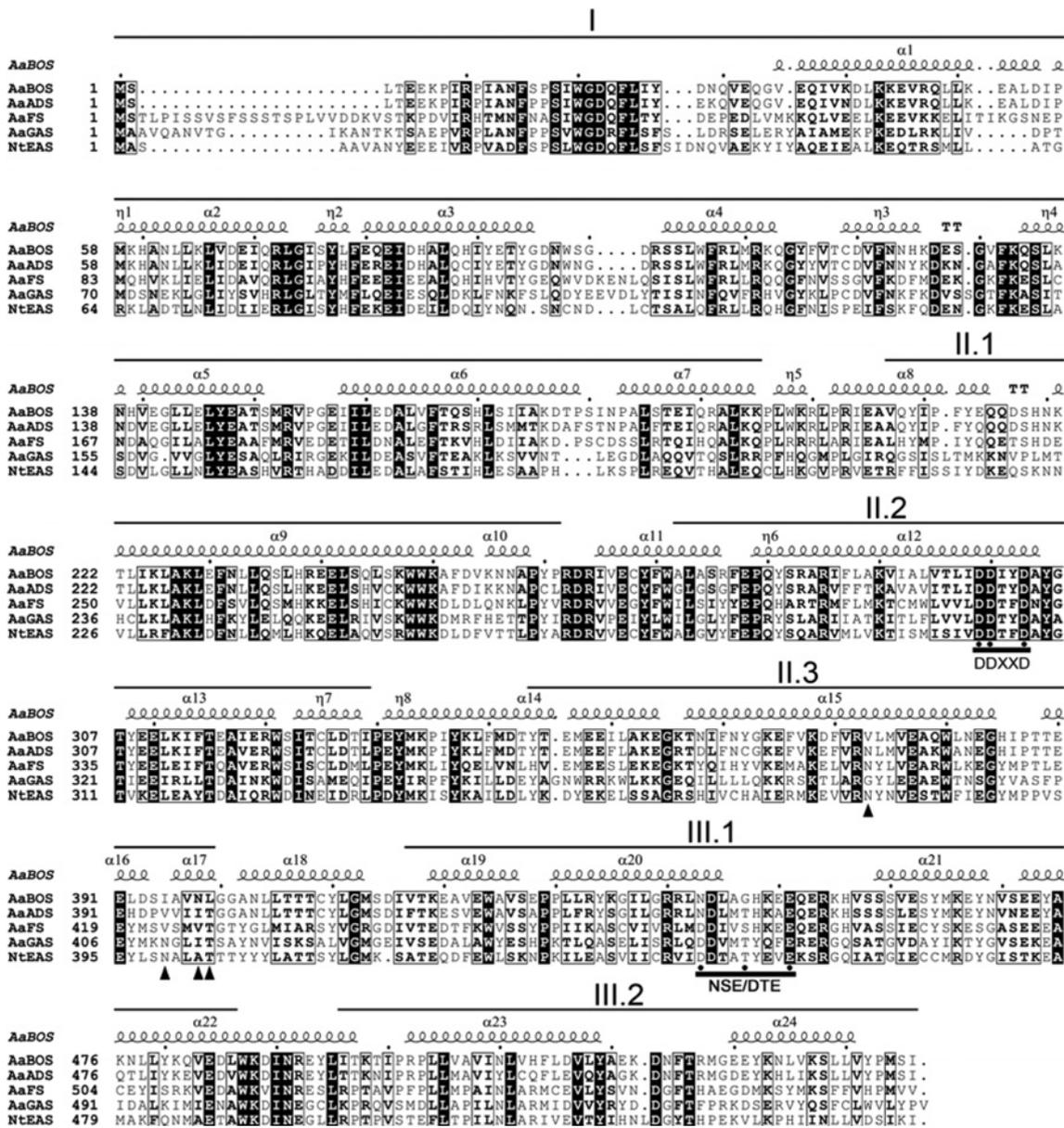


Figure S1 Comparison of amino acid sequences of AaBOS with four plant sesquiterpene synthases

The alignment was generated by the program ClustalW and drawn with ESPrict. Consensus amino acid residues are boxed in black, the secondary structure elements of AaBOS are shown above the sequences, and the highly conserved metal cofactor-binding regions of the DDXD motif and NSE/DTE motif are underlined. The catalytically important mutagenesis residues are indicated by triangles, and the fragments (of AaBOS and AaADS) used in domain swapping are indicated by transverse lines. The sesquiterpene synthase sequences aligned comprise four from *A. annua*, AaBOS (JQ717161), AaADS (CAB94691), AaFS (AAX39387) and AaGAS (ABE03980), and one from *N. tabacum*, NtEAS (Q40577).

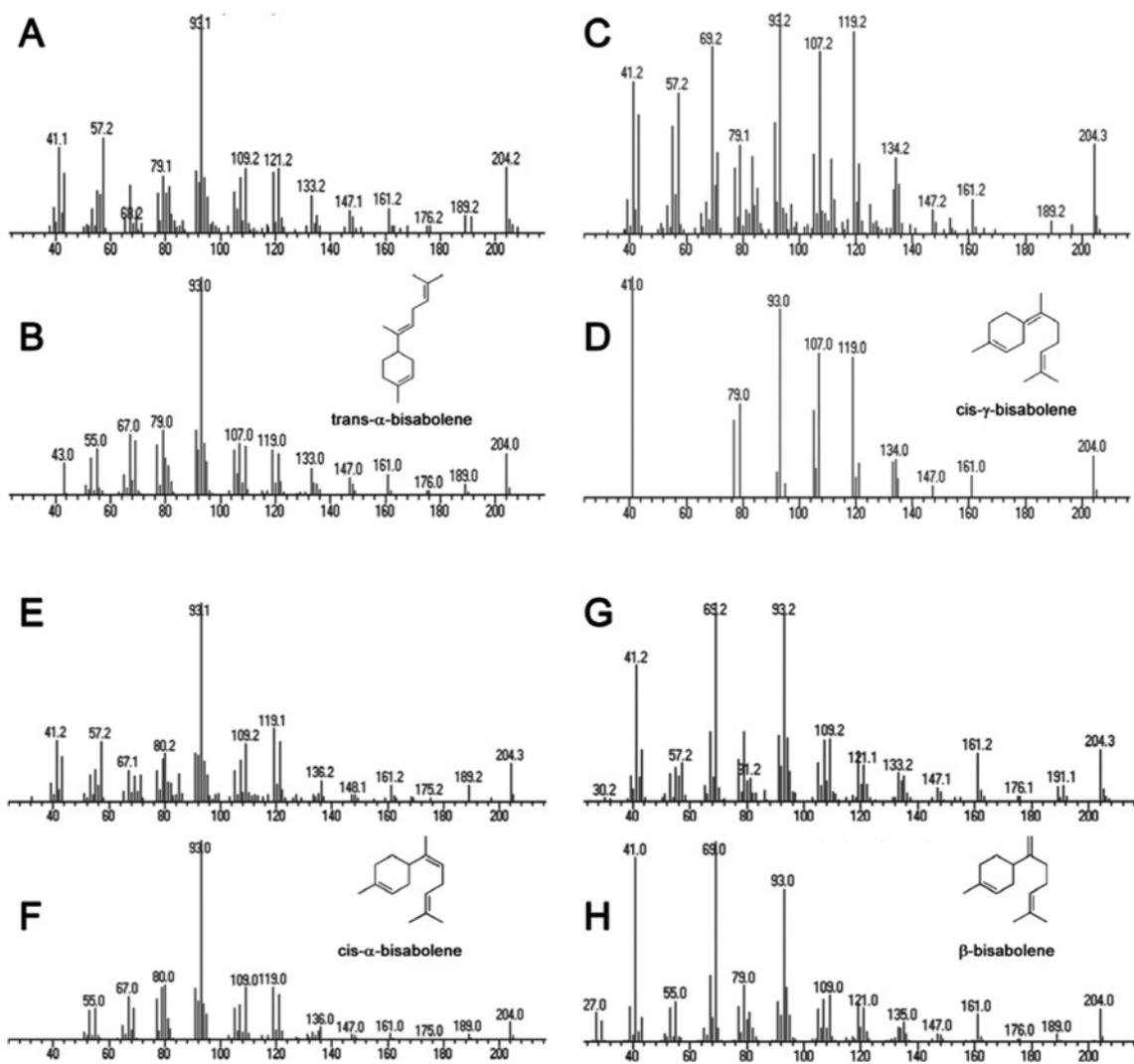


Figure S2 Mass spectra of the by-products of AaBOS

(A) Mass spectrum of peak 2 in Figure 2 of the main text. (B) Mass spectrum of *trans*- α -bisabolene [in the NIST (National Institute of Standards and Technology) library]. (C) Mass spectrum of peak 4 in Figure 2 of the main text. (D) Mass spectrum of *cis*- γ -bisabolene (in the NIST library). (E) Mass spectrum of peak 6 in Figure 2 of the main text. (F) Mass spectrum of *cis*- α -bisabolene (in the NIST library). (G) Mass spectrum of peak 3 in Figure 2 of the main text. (H) Mass spectrum of β -bisabolene (in the NIST library). Mass spectrum of peak 5 in Figure 2 of the main text and β -sesquiphellandrene are shown in Figure S3.

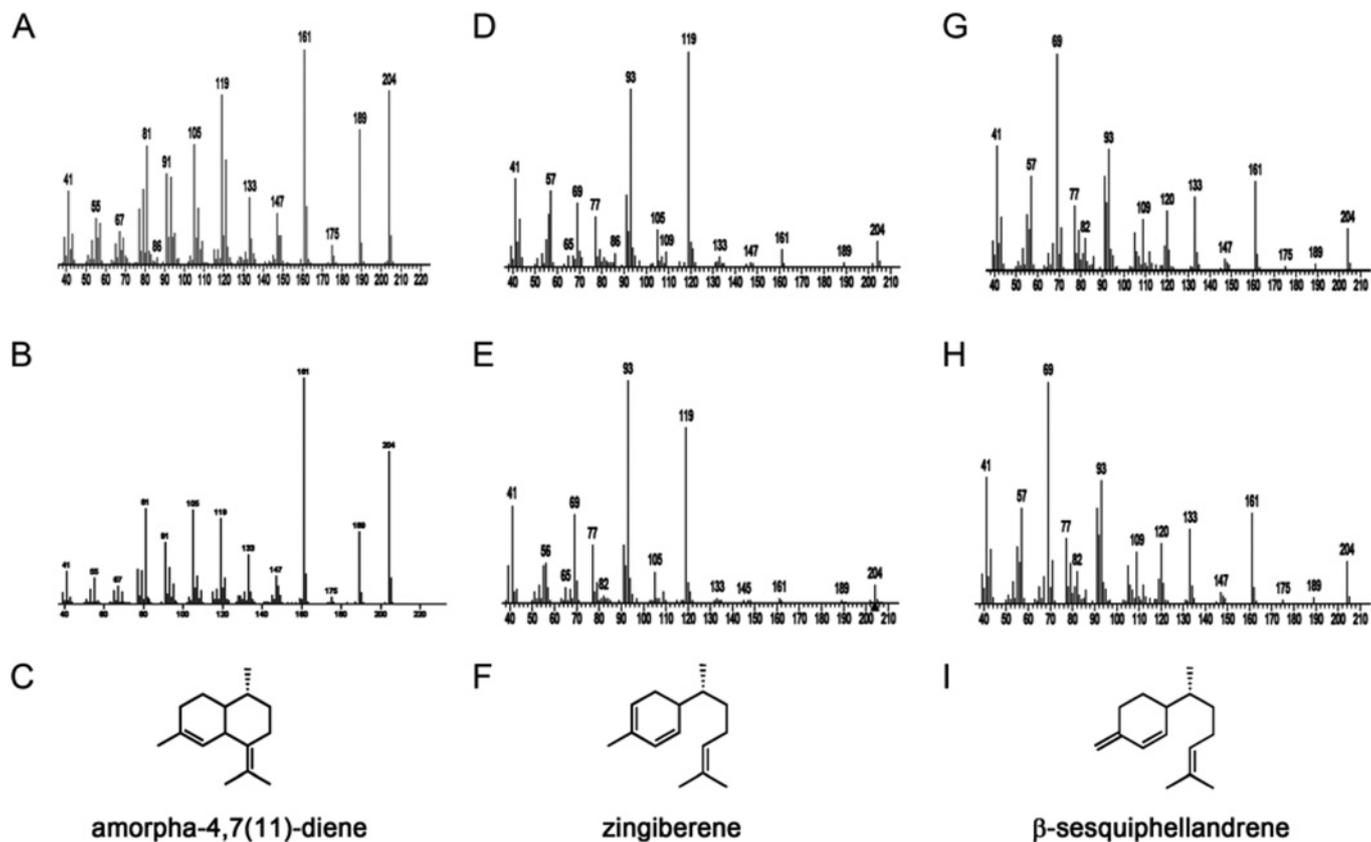


Figure S3 Mass spectra of the additional products formed by AaDS mutants

(A) Mass spectrum of peak 9 in Figure 6 of the main text. (B) Mass spectrum of amorpha-4,7(11)-diene (a gift from Professor Peter E. Brodelius). (C) Structure of amorpha-4,7(11)-diene. (D) Mass spectrum of peak 10 in Figure 6 of the main text. (E) Mass spectrum of zingiberene (in the NIST library). (F) Structure of zingiberene. (G) Mass spectrum of peak 11 in Figure 6 of the main text. (H) Mass spectrum of β -sesquiphellandrene (in the NIST library). (I) Structure of β -sesquiphellandrene (8).

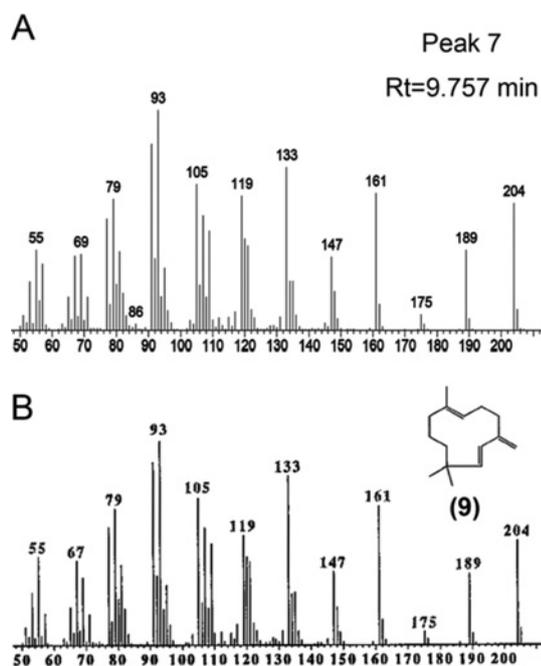


Figure S4 Mass spectra of the additional product formed by AaBOS mutants

(A) Mass spectrum of the additional product (peak 7) in Figure 4 of the main text. Rt, retention time. (B) The reported mass spectrum of γ -humulene [1].

Table S1 Oligonucleotide primers used in the present study

AS, antisense; S, sense.

Primer name	Sequence (5' → 3')	Purpose
AaActin-RT-S	CTTTGAACCCTAAGGCTAATC	qRT-PCR
AaActin-RT-AS	GGGCATACCCCTTCGTAGATAG	qRT-PCR
AaBOS-RT-S3	GATCTTGGCGGCCACAAGGAG	qRT-PCR
AaBOS-RT-AS2	CAGAACATCCAGAAAATGTAC	qRT-PCR
AaADS-SRT-S	GATCTCATGACCACAAGGCC	qRT-PCR
AaADS-SRT-AS	TTGAACCTCAAGAAACTGGCA	qRT-PCR
AaFPS-S	AGAAAGTCCACATGAGAAC	qRT-PCR
AaFPS-AS	TGCCTCTTATCAGGA GAATG	qRT-PCR
CYP71AV1-S	CGGATGTGTTCTTGACCTTAT	qRT-PCR
CYP71AV1-AS	TTCAGCACTGTCTTTGAGCCTTA	qRT-PCR
pETAaBOS-S	TTTCCATGGCTATGCTCTTACAGAAGAAAAACCTA	Protein expression
pETAaBOS-AS-final	TTTGGATCCTCATATACTCATAGGATAAACGAGT	Protein expression
pETAaADS-S	TTTCCATGGCTATGCTACTTACAGAAGAAAAACCTA	Protein expression
pETAaADS-AS-final	TTTGGATCCTCATATACTCATAGGATAAACGAGTAG	Protein expression
AaBOS-DR-S1	CCCCTTTGGAAAAGTTGCCAAG	Domain swapping
AaBOS-DR-AS1	CTTGGCAACCTTTTCCAAGGGG	Domain swapping
AaBOS-DR-S2	TGAGGGACACATACCAACCCTG	Domain swapping
AaBOS-DR-AS2	CAGTGGTTGGTATGTGCCCTCA	Domain swapping
AaBOS-DW-S1	GAGATAGAATTGTTGAATGCTAC	Domain swapping
AaBOS-DW-AS1	GTAGCATTCAACAATTCTATCTC	Domain swapping
AaBOS-DW-S2	CATGAAACCGATATACAAATTATTC	Domain swapping
AaBOS-DW-AS2	GAATAATTTGTATATCGGTTTCATG	Domain swapping
AaBOS-DW-S3	CCTGCTTACAACAACCTGTTATC	Domain swapping
AaBOS-DW-AS3	GATAACAAGTTGTTGTAAGCAGG	Domain swapping
AaBOS-DW-S4	GGAAAGATATAAACCGAGAGTACC	Domain swapping
AaBOS-DW-AS4	GGTACTCTCGGTTTATATCTTTCC	Domain swapping
AaBOS-L392H-S	CCAACCACTGAAGAGCATGATCAATTGCAGTCAAT	Site-directed mutagenesis
AaBOS-L392H-AS	ATTGACTGCAATTGAATCATGCTCTTCAAGTGGTTGG	Site-directed mutagenesis
AaBOS-M4-S1	TGAAGAGCTTGATTGATTGCAATCATTACTGGCGG	Site-directed mutagenesis
AaBOS-M4-AS1	CCGCCAGTAATGATTGCAACTGAATCAAGCTCTTCA	Site-directed mutagenesis
AaBOS-S394P-S	CCAACCACTGAAGAGCTTGATCCAATTGCAGTCAAT	Site-directed mutagenesis
AaBOS-S394P-AS	GATCAAGCTCTTCAAGTGGTTGGTATGTTCC	Site-directed mutagenesis
AaBOS-I395V-S	CCACTGAAGAGCTTGATTGATTGCAATCAATCTTG	Site-directed mutagenesis
AaBOS-I395V-AS	CTGAATCAAGCTCTTCAAGTGGTTGGTATGTTG	Site-directed mutagenesis
AaBOS-A396V-S	TGAAGAGCTTGATTCAATTGATGCAATCTTGGCGG	Site-directed mutagenesis
AaBOS-A396V-AS	ACAATTGAATCAAGCTCTTCAAGTGGTTGGTA	Site-directed mutagenesis
AaBOS-V397I-S	AAGAGCTTGATTCAATTGCAATCAATCTTGGCGGTG	Site-directed mutagenesis
AaBOS-V397I-AS	TTGCAATTGAATCAAGCTCTTCAAGTGGTTGG	Site-directed mutagenesis
AaBOS-N398I-S	GCTTGATTCAATTGCAATCAATCTTGGCGGTGCTAA	Site-directed mutagenesis
AaBOS-N398I-AS	ATGACTGCAATTGAATCAAGCTCTTCAAGTGG	Site-directed mutagenesis
AaBOS-L399T-S	TTGATTCAATTGCAATCAATCTTGGCGGTGCTAACCT	Site-directed mutagenesis
AaBOS-L399T-AS	GTATTGACTGCAATTGAATCAAGCTCTTCAAGT	Site-directed mutagenesis
AaBOS-397/9-AS	GCACCGCCAGTATTGATTGCAATGAATCAAGCTC	Site-directed mutagenesis
AaBOS-398/9-S	GAGCTTGATTCAATTGCAATCAATCTTGGCGGTG	Site-directed mutagenesis
AaBOS-398/9-AS	GCACCGCCAGTAACTGACTGCAATGAATCAAGCTC	Site-directed mutagenesis
AaBOS-395/7/8-S	GAGCTTGATTGCAATGCAATCATTCTTGGCGGTG	Site-directed mutagenesis
AaBOS-395/7/8-AS	GCACCGCCAAGAATGATTGCAACTGAATCAAGCTC	Site-directed mutagenesis
AaBOS-395/7/9-S	GAGCTTGATTGCAATGCAATCAATCTTGGCGGTG	Site-directed mutagenesis
AaBOS-395/7/9-AS	GCACCGCCAGTATTGATTGCAACTGAATCAAGCTC	Site-directed mutagenesis
AaBOS-395/8/9-S	GAGCTTGATTGCAATGCAATCATTACTGGCGGTG	Site-directed mutagenesis
AaBOS-395/8/9-AS	GCACCGCCAGTAACTGACTGCAACTGAATCAAGCTC	Site-directed mutagenesis
AaBOS-397/8/9-S	GAGCTTGATTCAATTGCAATCATTACTGGCGGTG	Site-directed mutagenesis
AaBOS-397/8/9-AS	GCACCGCCAGTAACTGATTGCAATGAATCAAGCTC	Site-directed mutagenesis
AaBOS-350-S	GAAATGGAAAGATTTCTTGCAAAGGAG	Site-directed mutagenesis
AaBOS-350-AS	CTCCTTTGCAAGAAATTTCCATTTTC	Site-directed mutagenesis
AaBOS-356-S	GCAAGAGGGGAAGAACAATAATATT	Site-directed mutagenesis
AaBOS-356-AS	AAATATATTTGTTCTTCCCTCCTTTGC	Site-directed mutagenesis
AaBOS-358-S	GGAGGGAAAAACAGATATATTAACCTG	Site-directed mutagenesis
AaBOS-358-AS	CAGTTAAATATATCTGTTTTCCCTCC	Site-directed mutagenesis
AaBOS-359-S	GGGAAAAACAATCTATTAACTGTGG	Site-directed mutagenesis
AaBOS-359-AS	CCACAGTTAAATAGATTGTTTTCC	Site-directed mutagenesis
AaBOS-369-S	GAATTTGTGAAAGAGTTTGTAGAGTCC	Site-directed mutagenesis
AaBOS-369-AS	GGACTCTAACAACTCTTTCACAAATTC	Site-directed mutagenesis
AaBOS-373-S	GATTTTGTGAAACCTGATGGTTGAAG	Site-directed mutagenesis
AaBOS-373-AS	CTTCAACCATCAGGTTTCAACAAAATC	Site-directed mutagenesis
AaBOS-379-S	GATGGTTGAAGCAAAATGGTTAAATGAG	Site-directed mutagenesis
AaBOS-379-AS	CTCATTAAACATTTGCTTCAACCATC	Site-directed mutagenesis
AaBOS-381-S	GGTTGAAGCACAAATGGGCAAAATGAGGACAC	Site-directed mutagenesis

Table S1 Continued

Primer name	Sequence (5' → 3')	Purpose
AaBOS-381-AS	GTGTCCCTCATTGCCCATTGTGCTTCAACC	Site-directed mutagenesis
AaAaADS-373-AS	TGGTTGGTATGTGTCCCTCATTGCCCATTTTGGCTTCAACCATCAGGACTCTAACAAAC	Site-directed mutagenesis
AaADS-M3-S	GGGACACATACCAACCCTGAAGAGCATGATCCAATTGTAATCAATCTTGGCGGTGCTA	Site-directed mutagenesis
AaADS-373-S1	GAGTTTGTAGAGTCCTGATGGTTG	Site-directed mutagenesis
AaADS-373-AS1	CAACCATCAGGACTCTAACAAACTC	Site-directed mutagenesis
AaADS-395-AS	ATGATTACAATTGGATCATGCTC	Site-directed mutagenesis
AaADS-398-AS	CACCGCCAGTATTGATTACAAC	Site-directed mutagenesis
AaADS-399-AS	TAGCACCGCCAAGAATGATTACAAC	Site-directed mutagenesis
AaADS-373/395-S	GGGACACATACCAACCCTGAAGAGCATGATCCAATTGTAATCATTACTGGCGGTGCTA	Site-directed mutagenesis
AaADS-373/398-S	GGGACACATACCAACCCTGAAGAGCATGATCCAAGTTGTAATCAATCTTGGCGGTGCTA	Site-directed mutagenesis
AaADS-373/399-S	GGGACACATACCAACCCTGAAGAGCATGATCCAAGTTGTAATCAATCTTGGCGGTGCTA	Site-directed mutagenesis
AaADS-395/8-AS	TAGCACCGCCAGTATTGATTACAATTGGATCATGCTC	Site-directed mutagenesis
AaADS-395/9-AS	TAGCACCGCCAAGAATGATTACAATTGGATCATGCTC	Site-directed mutagenesis
AaADS-398/9-AS	TAGCACCGCCAAGATTGATTACAAC	Site-directed mutagenesis
AaADS-373/395/398-S	GGGACACATACCAACCCTGAAGAGCATGATCCAATTGTAATCAATCTTGGCGGTGCTA	Site-directed mutagenesis
AaADS-373/395/399-S	GGGACACATACCAACCCTGAAGAGCATGATCCAATTGTAATCATTCTTGGCGGTGCTA	Site-directed mutagenesis
AaADS-373/398/399-S	GGGACACATACCAACCCTGAAGAGCATGATCCAAGTTGTAATCAATCTTGGCGGTGCTA	Site-directed mutagenesis
AaADS-395/8/9-S	GAGCATGATCCAATTGTAATCAATCTTGGCGGTGCTA	Site-directed mutagenesis
AaADS-395/8/9-AS	TAGCACCGCCAAGATTGATTACAATTGGATCATGCTC	Site-directed mutagenesis
AaADS-T399S-S	GTTGTAATCATTAGTGGCGGTGCTA	Site-directed mutagenesis
AaADS-T399S-AS	TAGCACCGCCACTAATGATTACAAC	Site-directed mutagenesis
AaADS-T399A-S	GTTGTAATCATTGCTGGCGGTGCTA	Site-directed mutagenesis
AaADS-T399A-AS	TAGCACCGCCAGCAATGATTACAAC	Site-directed mutagenesis
AaADS-T399N-S	GTTGTAATCATTAAATGGCGGTGCTA	Site-directed mutagenesis
AaADS-T399N-AS	TAGCACCGCCATTAATGATTACAAC	Site-directed mutagenesis
AaADS-T399D-S	GTTGTAATCATTGATGGCGGTGCTA	Site-directed mutagenesis
AaADS-T399D-AS	TAGCACCGCCATCAATGATTACAAC	Site-directed mutagenesis
AaADS-T399R-S	GTTGTAATCATTCTGGCGGTGCTA	Site-directed mutagenesis
AaADS-T399R-AS	TAGCACCGCCACGAATGATTACAAC	Site-directed mutagenesis

Table S2 Substitutions of AaBOS and products of the respective enzymes

A circle indicates that this product is generated by this mutant. No entry indicates no substituted positions or product not detected.

Substitution name	II.3.1								II.3.2							Products	
	350	356	358	359	369	373	379	381	392	394	395	396	397	398	399	α -Bisabolol	γ -Humulene
CH9	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A	L392H	S394P	I395V	A396V	V397I	N398I	L399T	○	○
CH12	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A								○	
CH13									L392H	S394P	I395V	A396V	V397I	N398I	L399T		
CH12-M8	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A	L392H							○	
CH12-M9	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A		S394P						○	
CH12-M10	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A			I395V					○	
CH12-M11	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A				A396V				○	
CH12-M12	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A					V397I			○	
CH12-M13	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A						N398I		○	
CH12-M14	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A							L399T	○	
CH12-M15	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A			I395V		V397I			○	
CH12-M16	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A			I395V			N398I		○	
CH12-M2	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A			I395V				L399T	○	○
CH12-M17	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A					V397I	N398I		○	
CH12-M18	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A					V397I		L399T	○	
CH12-M3	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A						N398I	L399T	○	○
CH12-M5	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A			I395V		V397I	N398I		○	
CH12-M6	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A			I395V		V397I		L399T	○	○
CH12-M4	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A			I395V			N398I	L399T	○	○
CH12-M7	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A					V397I	N398I	L399T	○	○
CH12-M1	I350F	K356R	N358D	I359L	D369E	V373N	Q379K	L381A			I395V		V397I	N398I	L399T	○	○
AaBOS-M3											I395V			N398I	L399T	○	
AaBOS-M4	I350F										I395V			N398I	L399T	○	
AaBOS-M5		K356R									I395V			N398I	L399T	○	
AaBOS-M6			N358D								I395V			N398I	L399T	○	
AaBOS-M7				I359L							I395V			N398I	L399T	○	
AaBOS-M8					D369E						I395V			N398I	L399T	○	
AaBOS-M1						V373N					I395V			N398I	L399T	○	○
AaBOS-M9							Q379K				I395V			N398I	L399T	○	
AaBOS-M10								L381A			I395V			N398I	L399T	○	○
AaBOS-M2						V373N		L381A			I395V			N398I	L399T	○	○
CH13-M1						V373N			L392H	S394P	I395V	A396V	V397I	N398I	L399T	○	○
CH13-M2						V373N		L381A	L392H	S394P	I395V	A396V	V397I	N398I	L399T	○	○
AaBOS-M11						V373N					I395V			N398V	L399T	○	○
AaBOS-M12						V373G					I395N			N398I	L399T	○	○

Table S3 Substitutions of AaADS and products of the respective enzymes

A circle indicates that this product is generated by this mutant. No entry indicates no substituted positions or product not detected. The numbered products denote: (2) amorpha-4,11-diene, (8) β -sesquiphellandrene, (11) zingiberene and (10) amorpha-4,7(11)-diene.

Substitution name	Substitutions				Products		
	373	395	398	399	2	8, 10, 11	
AaADS-M1	N373V					○	
AaADS-M2		V395I				○	
AaADS-M3			I398N			○	
AaADS-M4				T399L	○	○	
AaADS-M5	N373V	V395I					
AaADS-M6	N373V		I398N		○		
AaADS-M7	N373V			T399L			
AaADS-M8		V395I	I398N		○		
AaADS-M9		V395I		T399L	○	○	
AaADS-M10			I398N	T399L	○	○	
AaADS-M11	N373V	V395I	I398N		○		
AaADS-M12	N373V	V395I		T399L			
AaADS-M13	N373V		I398N	T399L			
AaADS-M14		V395I	I398N	T399L	○	○	
AaADS-M15	N373V	V395I	I398N	T399L			
AaADS _{T399D}				T399D			
AaADS _{T399R}				T399R			
AaADS _{T399A}				T399A	○	○	
AaADS _{T399N}				T399N	○	○	
AaADS _{T399S}				T399S	○		

REFERENCE

- 1 Steele, C. L., Crock, J., Bohlmann, J. and Croteau, R. (1998) Sesquiterpene synthases from grand fir (*Abies grandis*). Comparison of constitutive and wound-induced activities, and cDNA isolation, characterization and bacterial expression of δ -selinene synthase and γ -humulene synthase. *J. Biol. Chem.* **273**, 2078–2089

Received 8 January 2013/20 February 2013; accepted 26 February 2013

Published as BJ Immediate Publication 26 February 2013, doi:10.1042/BJ20130041