Plant Communications
Research Article

CellPress
Partner Journal

# Molecular Basis for Sesterterpene Diversity Produced by Plant Terpene Synthases

Qingwen Chen[1,9], Jianxu Li[2,9], Zhixi Liu[1,6,9], Takaaki Mitsuhashi[3], Yuting Zhang[4], Haili Liu[5], Yihua Ma[1,6], Juan He[1,6], Tetsuro Shinada[7], Tsutomu Sato[8], Yong Wang[5], Hongwei Liu[4], Ikuro Abe[3], Peng Zhang[2,*] and Guodong Wang[1,6,*]

[1]State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China

[2]National Key Laboratory of Plant Molecular Genetics, Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China

[3]Graduate School of Pharmaceutical Sciences, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[4]State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China

[5]Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

[6]University of Chinese Academy of Sciences, Beijing 100039, China

[7]Graduate School of Science, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi, Osaka 558-8585, Japan

[8]Department of Applied Biological Chemistry, Faculty of Agriculture, Graduate School of Science and Technology, Niigata University, 8050 Ikarashi-2, Niigata 950-2181, Japan

[9]These authors contributed equally to this article.

***Correspondence: Peng Zhang (pengzhang01@sibs.ac.cn), Guodong Wang (gdwang@genetics.ac.cn)**

https://doi.org/10.1016/j.xplc.2020.100051

## ABSTRACT

**Class I terpene synthase (TPS) generates bioactive terpenoids with diverse backbones. Sesterterpene synthase (sester-TPS, C25), a branch of class I TPSs, was recently identified in Brassicaceae. However, the catalytic mechanisms of sester-TPSs are not fully understood. Here, we first identified three nonclustered functional sester-TPSs (AtTPS06, AtTPS22, and AtTPS29) in *Arabidopsis thaliana*. AtTPS06 utilizes a type-B cyclization mechanism, whereas most other sester-TPSs produce various sesterterpene backbones via a type-A cyclization mechanism. We then determined the crystal structure of the AtTPS18–FSPP complex to explore the cyclization mechanism of plant sester-TPSs. We used structural comparisons and site-directed mutagenesis to further elucidate the mechanism: (1) mainly due to the outward shift of helix G, plant sester-TPSs have a larger catalytic pocket than do mono-, sesqui-, and di-TPSs to accommodate GFPP; (2) type-A sester-TPSs have more aromatic residues (five or six) in their catalytic pocket than classic TPSs (two or three), which also determines whether the type-A or type-B cyclization mechanism is active; and (3) the other residues responsible for product fidelity are determined by interconversion of AtTPS18 and its close homologs. Altogether, this study improves our understanding of the catalytic mechanism of plant sester-TPS, which ultimately enables the rational engineering of sesterterpenoids for future applications.**

**Keywords:** terpenoid, terpene synthase, sesterterpene, cyclization mechanism, crystal structure

**Chen Q., Li J., Liu Z., Mitsuhashi T., Zhang Y., Liu H., Ma Y., He J., Shinada T., Sato T., Wang Y., Liu H., Abe I., Zhang P., and Wang G.** (2020). Molecular Basis for Sesterterpene Diversity Produced by Plant Terpene Synthases. Plant Comm. **1**, 100051.

## INTRODUCTION

Plants synthesize many terpenoids (including phytohormones) with diverse ecological functions to help them survive in their local environments. The structural diversity of plant terpenoids is largely determined by two consecutive, but separate, enzymes, namely, short-chain prenyltransferases and terpene synthases (TPSs) (Chen et al., 2011; Wang et al., 2016a). Plant TPSs can be divided

Published by the Plant Communications Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, CAS.

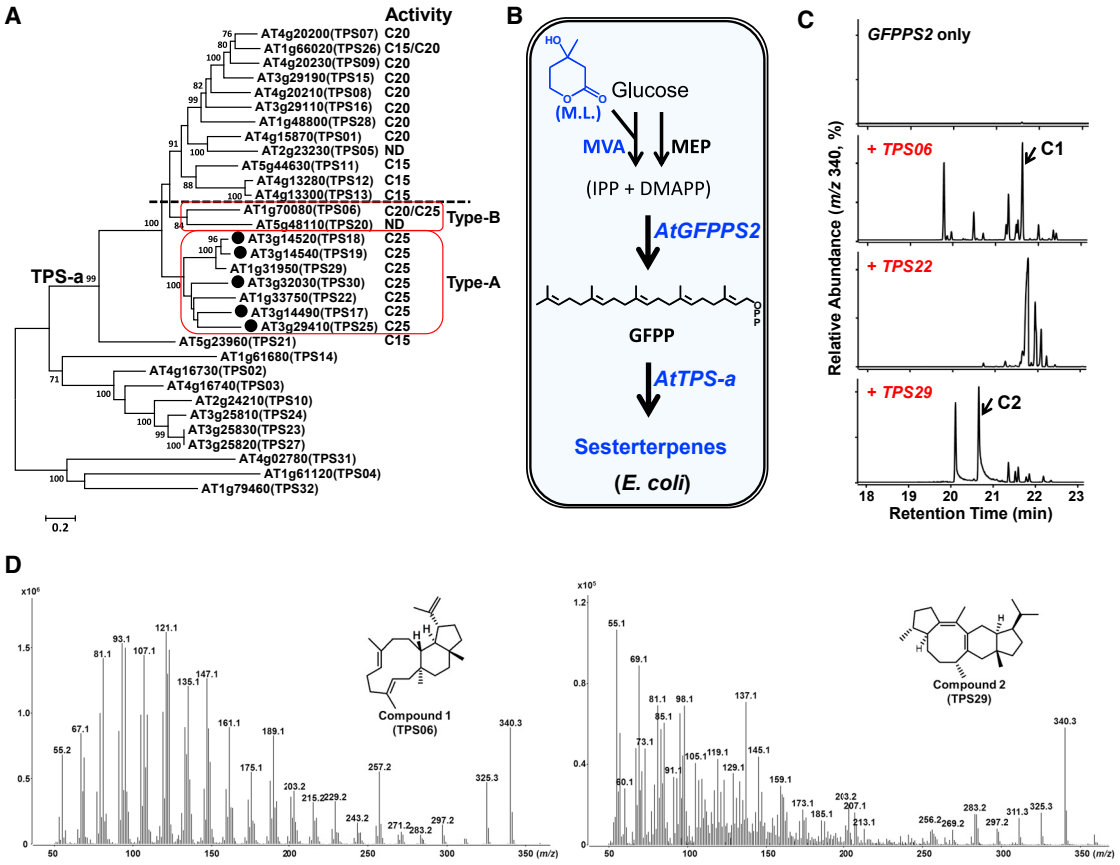



**Figure 1. Sesterterpene Synthases in Arabidopsis and Structures of the Main Products of TPS06 and TPS29.**

**(A)** Phylogenetic analyses of 32 TPS proteins from *A. thaliana* (Columbia-0 ecotype) using the maximum-likelihood method. Bootstrap values (based on 1000 replicates) >75% are shown for the corresponding nodes. The five known Arabidopsis sester-TPSs are marked with black dots, and the functionally identified sester-TPSs are boxed in red (type A and type B). ND, not determined.

**(B)** Scheme for screening sester-TPS genes using the *E. coli* system. MEP, 2-*C*-methyl-D-erythritol 4-phosphate pathway; ML, mevalonolactone; MVA, mevalonate pathway.

**(C)** GC–MS analysis (SIM mode, *m/z* 340 for $C_{25}H_{40}$) of the sesterterpenes produced in *E. coli* harboring different TPS-a genes from Arabidopsis: upper panel, AtTPS06; middle panel, AtTPS22; bottom panel, AtTPS29. AtTPS06, a multi-product enzyme, produces more than ten different sesterterpenes from GFPP.

**(D)** Mass spectra and chemical structures of the main sesterterpene products of TPS06 (**C1,** left) and TPS29 (**C2,** right).

into two classes based on the initiation mechanisms of prenyl pyrophosphate cyclization. Class I TPSs usually catalyze ionization-initiated cyclization reactions, whereas class II TPSs catalyze protonation-initiated reactions (Gao et al., 2012; Christianson, 2017). Overall, class I TPSs contain two $Mg^{2+}$-binding motifs, namely, one conserved motif (DDXXD; X is any amino acid) on helix D of the active α domain and another nonconserved motif ((N,D)D(L,I,V)X(S,T)XXXE, designated as the NSE or DTE motif) on helix H of the α domain (Christianson, 2017). After the ionization of prenyl pyrophosphate, the cationic prenyl chain moves into the hydrophobic and aromatic-group-rich pocket of the active site, leading to the formation of diverse terpenoid backbones through carbocation rearrangements and neutralization (Christianson, 2017; Starks et al., 1997 and literature therein). Recently, Srividya et al. (2015) performed alanine-scanning mutagenesis of 48 amino acids that potentially surround the active site of (*4S*)-limonene synthase from spearmint (*Mentha spicata* L.). Combined with product analysis, the authors determined the importance of W324 (tryptophan) and H579 (histidine) in the stabilization of the carbocation intermediates (PDB: 20NH).

Phylogenetic analyses show that eight subfamilies of plant TPSs are well established (TPS-a to TPS-h). Among them, the TPS-a subfamily (class I TPSs with an N-terminal inactive β domain and a C-terminal active α domain), which is common in flowering plants, provides the basis for the sub- and/or neofunctionalization of plant TPSs in this clade (Chen et al., 2011). There are 22 TPS-a members in the Arabidopsis genome (a total of 32 *TPS* genes), some of which have been demonstrated to be mono- (C10), sesqui- (C15), and di-TPS (C20) derivatives (Tholl and Lee, 2011; Wang et al., 2016b). Recently, our group and Osbourn's group functionally characterized novel *PT-TPS-P450* gene clusters in *Arabidopsis thaliana*, and these genes showed unprecedented sester-TPS activity by utilizing geranylfarnesyl pyrophosphate (GFPP) as a substrate (Wang et al., 2016a; Huang et al., 2017, 2018; Shao et al., 2017; He et al., 2019). Although only approximately 140 plant sesterterpenoids have been isolated and characterized to date, these compounds are structurally diverse and widely distributed in the plant kingdom (Liu et al., 2016; Huang et al., 2017). Compared with the extensive studies on mono-, sesqui-, and di-TPSs, there are

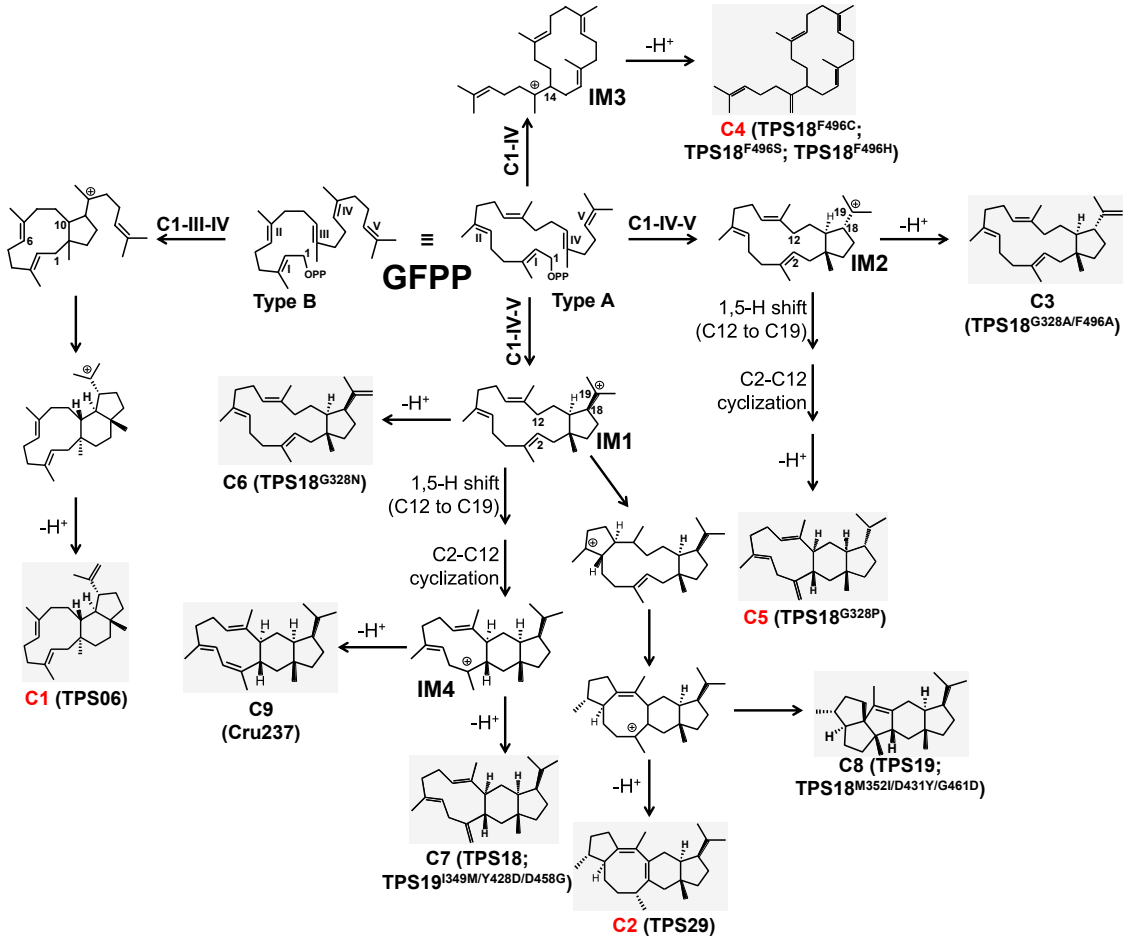



**Figure 2. Proposed Cyclization Scheme for the Synthesis of Sesterterpene Backbones by Arabidopsis sester-TPSs.**
The chemical structures elucidated in this study are indicated by red numbers. The products of AtTPS18 and its mutants are numbered **C3–C7** according to their retention time in the GC–MS experiment carried out in this study (see below). All chemical structure elucidation data in this study are shown in Supplemental Tables 1 and 2 and Supplemental Figures 1–5. IM, intermediate. **C3** ((+)-brassitetraene B**)**, **C6** ((+)-brassitetraene A**)**, and **C9** ((−)-caprutriene) were previously determined by Huang et al. (2017, 2018), whereas **C7** ((+)-thalianatriene) and **C8** ((−)-retigeranin B**)** were elucidated by Shao et al., (2017).

only a few studies on sester-TPSs from plants and fungi (Chiba et al., 2013; Matsuda et al., 2015, 2016; Ye et al., 2015; Chai et al., 2016; Bian et al., 2017; Huang et al., 2017, 2018; Shao et al., 2017). Two early-stage cyclization mechanisms of GFPP in the catalytic pocket of sester-TPSs have been proposed, namely, type A, $C_1$-IV-V (cyclization between the $C_1$ cation, the $C_{14}$–$C_{15}$ olefin [the fourth double bond, IV] and the $C_{18}$–$C_{19}$ olefin [the fifth double bond, V]) and type B, $C_1$-III-IV (cyclization between the $C_1$ cation, the $C_{10}$–$C_{11}$ olefin [the third double bond, III], and the $C_{14}$–$C_{15}$ olefin [the fourth double bond, IV]) (Minami et al., 2018; Sato et al., 2018). However, the amino acid(s) determining the difference between type-A sester-TPS and type-B sester-TPS remains largely unknown. Moreover, no sester-TPS from plants utilizing the type-B cyclization mechanism has been characterized to date. Although several routes to the formation of polycyclic sesterterpene backbones from GFPP have been proposed, all identified plant sester-TPSs share the 5/15 bicyclic intermediate (type-A route) (Huang et al., 2017, 2018; Shao et al., 2017). Recently, we demonstrated that single amino acids ($Gly^{328}$ in AtTPS18/AtTPS25, $Gly^{325}$ in AtTPS19, and $Pro^{328}$ in AtTPS30) are critical for the substrate specificity of sester-TPSs in Arabidopsis. Specifically, $AtTPS18^{G328W}$, $AtTPS19^{G325W}$, $AtTPS25^{G328W}$, and $AtTPS30^{P328W}$ lost their sester-TPS activity, while mutants ($AtTPS18^{G328W}$, $AtTPS19^{G325W}$, and $AtTPS25^{G328W}$) utilized geranyl pyrophosphate, farnesyl pyrophosphate (FPP), or geranylgeranyl pyrophosphate (GGPP) as substrates (Chen et al., 2019). However, it is unclear how sester-TPSs accommodate GFPP conformers and direct the rearrangements and neutralization of C25 prenyl carbocation intermediates to produce diverse sesterterpenes.

In this study, we identified three additional sester-TPS encoding genes (*AtTPS06*, *AtTPS22*, and *AtTPS29*) that are not located in the *GFPPS-TPS-P450* cluster in Arabidopsis. Interestingly, AtTPS06 catalyzed reactions through the 5/11 bicyclic intermediate (type-B route) based on the structure of the resulting sesterterpene. We further discovered the amino acids that control the product specificity/fidelity of plant sester-TPSs by determining the crystal structure of AtTPS18, and the key amino acids were

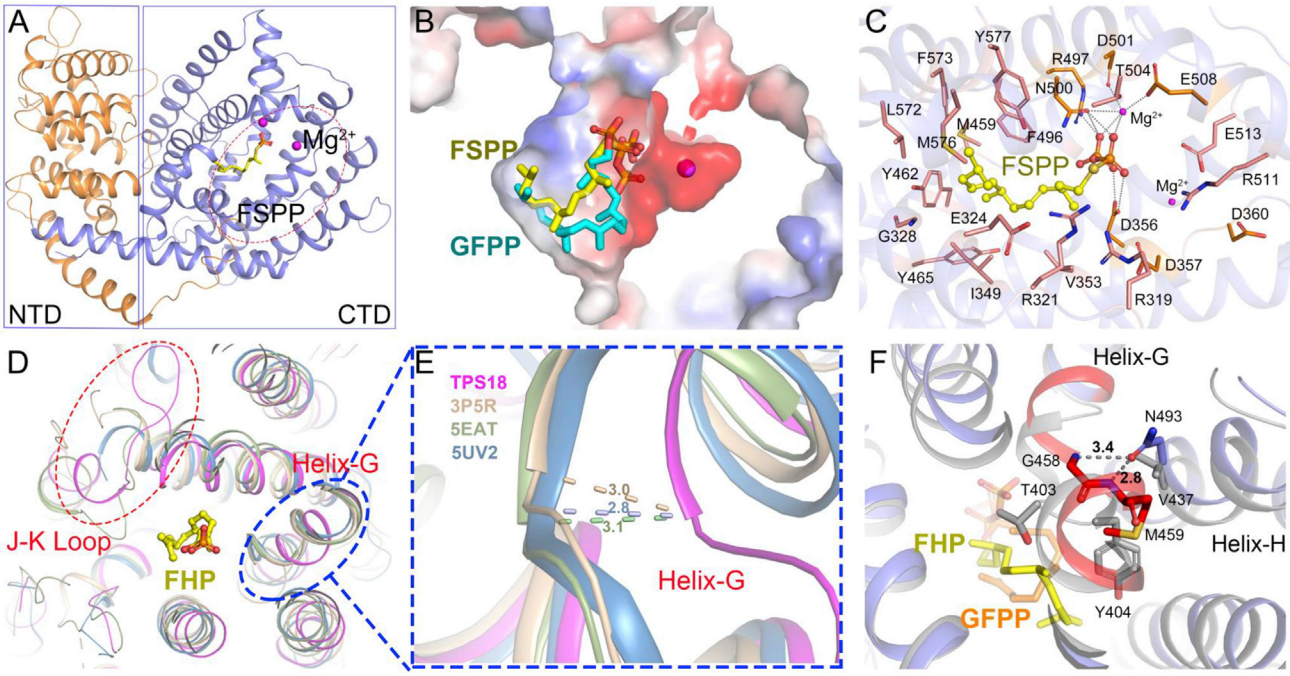



**Figure 3. Structure of the AtTPS18–FSPP Complex.**

**(A)** The overall structure of the AtTPS18–FSPP complex as a ribbon cartoon (NTD and CTD are shown in orange and cyan, respectively). The substrate analog, FSPP, is shown as a yellow stick model, and the two $Mg^{2+}$ atoms are shown as magenta balls.

**(B)** Expansion of the substrate-binding pocket and the docking results showing the substrate binding of GFPP (electron static surface). FSPP and GFPP are shown as yellow and cyan stick models, respectively.

**(C)** Interaction of amino acid residues with FSPP in the binding site. The bound FSPP is depicted as a ball-and-stick model. The 24 residues in the active site of AtTPS18 are shown. Among them, $Arg^{319}$, $Arg^{321}$, $Asp^{356}$, $Asp^{357}$, $Asp^{360}$, $Arg^{497}$, $Asn^{500}$, $Asp^{510}$, $Thr^{504}$, $Glu^{508}$, $Arg^{511}$, and $Glu^{513}$ were not considered for mutagenesis because they are conserved in the ionization of prenyl diphosphate substrates.

**(D)** Superimposition of the cartoon models of AtTPS18 (magenta), NtEAS (yellow, PDB: 5EAT), taxadiene synthase from *Taxus brevifolia* (tan, PDB: 3P5R), and limonene synthase from *Citrus sinensis* (slate blue, PDB: 5UV2) from a top–down perspective. Farnesyl hydroxyphosphonate (FHP) in NtEAS and GFPP in AtTPS18 are shown as yellow and cyan sticks, respectively. The positions of the I–J loop and helix G (G1 and G2) are indicated by red and blue circles, respectively.

**(E)** Close-up view of the helix G region of AtTPS18, 5EAT, 3P5R, and 5UV2 (for details, see the legend of **D**). The shift distance (Å) of helix G of AtTPS18, relative to that of other plant TPSs, was measured and labeled.

**(F)** Possible interactions between $Asn^{493}$ (in helix H) and helix G in the AtTPS18–FSPP complex. Dashed lines indicate possible hydrogen bonds.

confirmed by site-directed mutagenesis. Our results provide a feasible pipeline for producing high-value or unnatural sesterterpenoids on a large scale using a synthetic biology strategy.

## RESULTS AND DISCUSSION

### Type-B Sester-TPSs, Possible Progenitors of Type-A Sester-TPSs, Are Discovered in Brassicaceae

Due to the huge functional divergence of plant TPSs (Chen et al., 2011; Zhou and Pichersky, 2020), a sufficient number of characterized sester-TPSs, together with the chemical structures of their products, is a prerequisite for understanding the catalytic mechanism of plant sester-TPSs. In an effort to discover new sester-TPSs from *A. thaliana*, we screened ten uncharacterized members (*AtTPS01*, *AtTPS06*, *AtTPS07*, *AtTPS09*, *AtTPS15*, *AtTPS16*, *AtTPS22*, *AtTPS26*, *AtTPS28*, and *AtTPS29*; we did not obtain the full-length cDNA of *AtTPS05*) in the TPS-a subfamily using an optimized *Escherichia coli* (*E. coli*) system (Figure 1A). A partial mevalonate (MVA) pathway (pMBIS construct [Martin et al., 2003]) was introduced into the *E. coli* system, and mevalonolactone was added to the culture medium to increase the production of the terpenoid of interest (Figure 1B). When coexpressed with *AtGFPPS2* (*At3g14550*), sesterterpenes (selective ion monitoring [SIM] mode for gas chromatography–mass spectrometry [GC–MS], *m/z* 340 for $C_{25}H_{40}$) were detected in transgenic *E. coli* strains harboring *AtTPS06*, *AtTPS22*, or *AtTPS29* (Figure 1C). No sester-TPS activity was observed for the other tested TPSs. Notably, *AtTPS01*, *AtTPS06*, *AtTPS07*, *AtTPS09*, *AtTPS15*, *AtTPS16*, and *AtTPS26* encoded di-TPSs (Supplemental Figures 6 and 7), and the chemical structures of these diterpenes merit further investigation. To determine their structures, we purified compound **1** from AtTPS06 (0.26 mg) and compound **2** from AtTPS29 (0.44 mg). One-dimensional and two-dimensional nuclear magnetic resonance (NMR) analysis of these compounds (**C1** and **C2**, respectively) revealed their planar structures and relative stereochemistries, and both compounds were found to be novel sesterterpenes (all chemical structure elucidation data in this study are shown in Supplemental Tables 1 and 2, and Supplemental Figures 1–5). **C2** possessed a 5-8-6-5 fused ring system, which appeared to be formed by the type-A route

| Mutations | Product distribution (%) | | | | | | | | | Relative yield (-fold)[c] |
|---|---|---|---|---|---|---|---|---|---|---|
| | a[a] | b | C | C3[b] | C4 | C5 | C6 | d | C7 | |
| AtTPS18 (WT) | 0.15 | 0.01 | 2.8 | 0.6 | 1.3 | 3.2 | 0.8 | 6.1 | 85.1 | 1.0 |
| N493F | – | – | 2.6 | 0.65 | 2.4 | 4.2 | 1.2 | 7.8 | 81.2 | 0.4 |
| Y465A | 0.5 | 0.15 | 2.4 | 3.8 | 4.7 | 4.8 | 1.8 | 16.1 | 65.7 | 1.01 |
| G328P | 0.34 | – | 1.5 | 3.1 | 1.7 | 88.6 | 0.05 | 0.69 | 3.95 | 1.62 |
| G328A | 0.72 | – | 2.1 | 3.34 | 4.94 | 21.8 | 0.04 | 7.1 | 59.9 | 1.35 |
| G328S | 0.04 | – | 4.11 | 1.77 | 1.47 | 2.2 | 0.92 | 9.4 | 80.05 | 1.15 |
| G328C | 0.3 | – | 2.4 | 1.7 | 1.4 | 15.1 | 0.06 | 7.2 | 71.9 | 0.67 |
| G328L | – | – | – | 10.1 | 6.5 | 77.7 | – | – | 5.75 | 0.16 |
| G328N | 2.3 | – | – | 16.9 | 0.19 | – | 79.2 | – | 1.35 | 0.08 |
| F496S | 8.2 | 3.7 | – | 9.5 | 40.1 | – | 17.3 | 0.45 | 20.8 | 0.11 |
| F496M | 1.6 | 0.25 | 1.7 | 8.1 | 10 | 5.2 | 4.5 | 5.2 | 63.6 | 0.49 |
| F496H | 1.5 | – | 0.2 | 13.2 | 45.9 | – | 12.6 | 4.8 | 21.7 | 0.22 |
| F496A | – | 10.2 | – | 9.3 | 28.3 | – | 26.7 | 1.2 | 24.3 | 0.17 |
| F496C | 7.9 | 6.8 | 0.2 | 9.7 | 37.0 | 0.55 | 10.7 | 2.7 | 24.3 | 0.14 |
| G328A/F496A | 1.4 | 5.7 | 0.3 | 10.2 | 23.0 | 14.4 | 16.3 | 3.1 | 25.4 | 0.2 |

**Table 1. Product Distribution of AtTPS18 and Its Mutants.**

All product distributions were calculated from at least three independent experiments, whereas the results of F496S and F496A mutants were obtained from two replicates.

[a]The structures of sesterterpenes a–d have not yet been determined.

[b]The chemical structures of **C3–C7** are shown in Figure 2.

[c]The sum of the areas of the chromatogram peaks (GC–MS with SIM mode, *m/z* 340) of the main sesterterpenes produced by AtTPS18 and its mutants were calculated as the yield after normalization with an internal standard and culture density ($OD_{600}$ value). The yield of AtTPS18 (wild type) was set to 1.0.

(Figures 1D and 2). Although 5-8-6-5 fused ring sesterterpenoids have been reported in plants and fungi, they vary in terms of their stereochemistry and position of their double bonds (Ye et al., 2015; Huang et al., 2017). **C1** was revealed to be an *E/Z* isomer of flocerene, a sesterterpene with an 11-6-5 fused ring system, which was initially isolated from the secretions of the insect *Ceroplastes floridensis* (Figure 1D and Supplemental Figure 2) (Naya et al., 1981). The 11-6-5 fused ring system of **C1** appeared to be generated by the type-B route (Figure 2). We did not obtain sufficient purified AtTPS22 for NMR analysis due to the low yield and poor resolution of the signals by liquid chromatography.

Notably, AtTPS06 is the first type-B sester-TPS identified in plants. Moreover, AtTPS06 is more similar to sesqui-TPSs than to sester-TPSs in the phylogenetic analysis (Figure 1A), which suggests that TPS06 and its close homologs are possible intermediate enzymes between mono-/sesqui-/di-TPSs and sester-TPSs in the evolution of plants. AtTPS06 was initially characterized as a di-terpene synthase that converted GGPP to dolabelladienol and other unidentified diterpenes (Supplemental Figures 6 and 7; Wang et al., 2016b). The catalytic promiscuity of AtTPS06 (using GGPP and GFPP as substrates) suggests that it served as a progenitor of type-A sester-TPSs during the evolution of TPSs, at least in Brassicaceae (Weng, 2014).

qRT–PCR analyses revealed that both *AtTPS22* and *AtTPS29* were highly expressed in roots and siliques, whereas *AtTPS06* expression was low in seedling roots and flowers (Supplemental Figure 8). We did not detect **C1** and **C2**, the main C25 terpene products of AtTPS06 and AtTPS29, respectively, in any tested Arabidopsis tissues (100 mg [fresh weight] of roots and flowers) using GC–MS with the full-scan mode. However, we cannot simply conclude that **C1** and **C2** are not produced in Arabidopsis, given that the amount of **C1** and **C2** might be below the detection limit of GC–MS. Another possibility is that **C1** and **C2** might be further metabolized *in planta*.

## Plant Sester-TPSs Have a Larger Catalytic Pocket than Classic TPSs

To investigate the mechanism underlying the divergence of these Arabidopsis sester-TPSs (as they generated profiles different from that of GFPP), we recombinantly expressed truncated AtTPS18 (residues 52–605; the first 51 amino acids were predicted to be the plastidial signal peptides) in *E. coli* and purified the protein. The AtTPS18 complex with FSPP (farnesyl-*S*-thiolodiphosphate, an inert analog of FPP; Figure 3 and Supplemental Table 3) was crystallized, and its structure was determined at a resolution of 2.3 Å. The final model contained residues 71–605, and it was comprised of two compact $\alpha$-helical domains, namely, the N-terminal domain (NTD) and the C-terminal catalytic domain (CTD) (Figure 3A). FSPP adopts an extension configuration when binding to the CTD. Two $Mg^{2+}$ centers located near the DDXXD and NSE/DTE motifs are responsible for binding to the pyrophosphate head of FSPP, and the carbon tail of FSPP binds to the hydrophobic pocket that is mainly formed by hydrophobic residues from helices C to J (Figure 3A–3C). The cocrystallization of AtTPS18 with

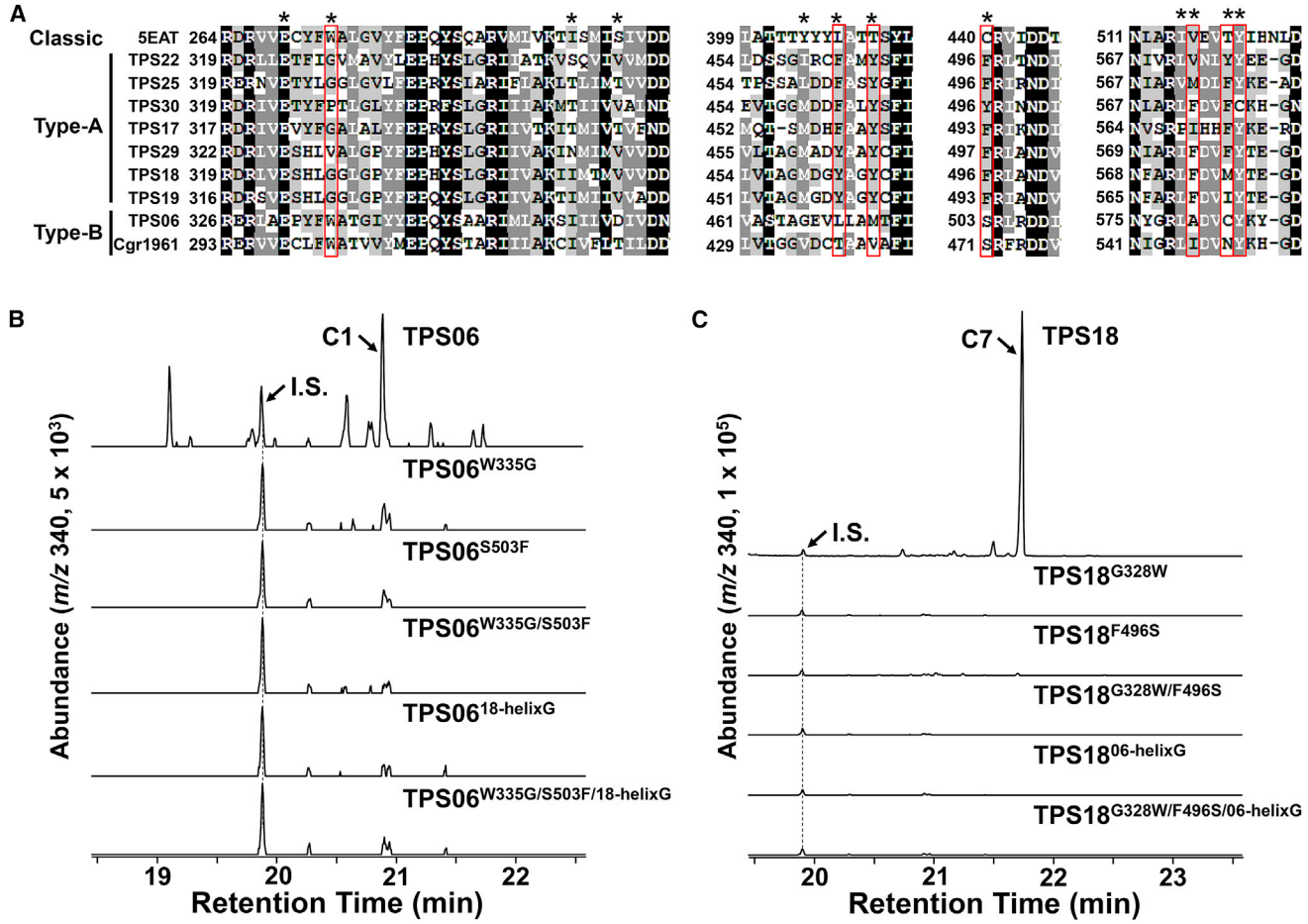



**Figure 4. Effect of Two Amino Acids on the Activity of Two Types of Sester-TPSs.**

**(A)** Sequence alignment of nine characterized sester-TPSs and 5EAT (a sesquiterpene synthase). The amino acids marked with stars are located in the catalytic pocket of AtTPS18 (see Figure 3C). Among them, the positions occupied by aromatic residues in sester-TPSs and $Gly^{328}$ are boxed in red.

**(B)** GC–MS analysis (SIM mode, *m/z* 340 for $C_{25}H_{40}$) of the sesterterpenes produced in *E. coli* harboring AtTPS06 and its mutants. (−)-Variculatriene, the main product of AtTPS25, was used as the internal standard (I.S.). Notably, the same amount of (−)-variculatriene was used in the experiments shown here and in **(C)**.

**(C)** GC–MS analysis (SIM mode, *m/z* 340 for $C_{25}H_{40}$) of the sesterterpenes produced in *E. coli* harboring AtTPS18 and its mutants.

synthesized GFPP and the soaking of apo-AtTPS18 crystals in this ligand were attempted, but none of the solved structures contained the GFPP substrate. We therefore used AutoDock to model the GFPP substrate in the AtTPS18 structure and found that GFPP fit well in the substrate-binding pocket and generally overlapped with the FSPP-binding site (Figure 3B). These results indicate that the substrate-binding pocket of AtTPS18 is very suitable for the binding of GFPP.

A comparison of the AtTPS18 structure with that of NtEAS (5-epi-aristolochene synthase, PDB: 5EAT [Starks et al., 1997]) showed a root-mean-square deviation value of 1.67 Å. AtTPS18 adopted a three-dimensional structure similar to that typically found in plant sesquiterpene synthases (Figure 3D). The major differences were found at the C-terminus. A superimposition of the structures of AtTPS18 and NtEAS revealed that all helices (helices A to K), except for helix G, H2, and the I–J loop (connecting helices I and J), fit well (Figure 3D). A comparison of the structures of other plant terpene synthases with the structure of AtTPS18 revealed that the I–J loop served as a lid, which was slightly separated from the top of the substrate-binding pocket (Figure 3D). In addition, the most obvious conformational change was the outward shift of helix G in AtTPS18, which created a larger catalytic pocket compared with those of other plant TPSs, thus allowing it to accommodate GFPP (Figure 3D and 3E). Compared with mono-TPS (limonene synthase, PDB: 5UV2 [Kumar et al., 2017]), sesqui-TPS (PDB: 5EAT [Starks et al., 1997]), di-TPS (taxadiene synthase, PDB: 3P5R [Koksal et al., 2011]), and AtTPS18 (obtained in this study), helix G in AtTPS18 shifted outward by 2.8 Å, 3.1 Å, and 3.0 Å, respectively (Figure 3E). We further measured the pocket volume of the three representative plant TPSs and AtTPS18 using CAVER Analyst 2.0 (Jurcik et al., 2018). The results showed that the volume of the pockets in terpene synthases (5UV2: 593.0 $Å^3$; 5EAT: 801.9 $Å^3$; 3P5R: 1285.5 $Å^3$; AtTPS18: 1513.8 $Å^3$) increased gradually with the chain length of the substrate (every isoprene unit occupied approximately 300 $Å^3$, Supplemental Figure 9A–9D). Notably, AtTPS06 (type B; 1536.3 $Å^3$) and AtTPS18 (type A) had similar pocket volumes, although they used different cyclization mechanisms (Supplemental Figure 9E).

Previous studies have clearly demonstrated that the amino acid residues in helix G affect the product specificity and capacity of TPS enzymes (Starks et al., 1997; Li et al., 2013). To further investigate the mechanism of the outward shift in helix G (the sequence alignment is shown in Supplemental Figure 10A and 10B), we carefully checked the possible interactions between the amino acids in the kink of helix G in the AtTPS18–FSPP complex. As shown in Figure 3F, we found that the side chain of $Asn^{493}$ in helix H formed two hydrogen bonds with the main chain of residues $Gly^{458}$ and $Met^{459}$ in helix G, which probably resulted in the outward shift of helix G. Notably, position 493 was occupied by Asn in approximately 50% of active sester-TPSs ($n$ = 18), whereas this position was always occupied by Ala in most mono-/sesqui-TPSs ($n$ = 59; Supplemental Figure 10C). Because synthetic GFPP was in short supply, we used the *E. coli* system harboring *AtGFPPS2* to test the product profiles of the mutated enzymes in this study (as shown in Figure 1B and 1C; for detailed calculations, seeMethods). Western blotting (anti-myelin basic protein [MBP] tag) was used to ensure no significant difference in the protein solubility among the mutated sester-TPSs (Supplemental Figure 11). Site-directed mutagenesis and biochemical assays showed that mutation of the polar residue $Asn^{493}$ to the nonpolar residue Phe decreased sester-TPS activity by approximately 60% (Table 1). This might be due to the disruption of the hydrogen bonds contributed by $Asn^{493}$, thus resulting in an inward shift of the elastic helix G and shrinkage of the pocket volume. We also replaced the kink sequence in helix G of AtTPS18 ($^{456}$TAGMDGY$^{462}$, the kink sequence is underlined; Supplemental Figure 10A) with the kink sequence of 5EAS enzyme ($^{401}$TTTYYYL$^{407}$). This replacement might also reduce the volume of the catalytic pocket, thus inactivating the enzyme (Supplemental Data 1). These data confirm the importance of the enlarged catalytic pocket created by the outward shift of helix G in sester-TPS activity.

## Type-A Sester-TPSs, Not Type-B Sester-TPSs, Have More Aromatic Residues in the Catalytic Pocket than Classic TPSs

In general, only two or three aromatic residues, which possibly stabilize reactive carbocation intermediates, are found in the classic TPS catalytic pocket (Christianson, 2017). However, the AtTPS18 structure clearly revealed that there were five aromatic residues ($Tyr^{462}$, $Tyr^{465}$, $Phe^{496}$, $Phe^{573}$, and $Tyr^{577}$) in the catalytic pocket (Figure 3C). Moreover, three positions ($Tyr^{462}$, $Tyr^{465}$, and $Phe^{496}$ in AtTPS18) were conserved and unique to type-A sester-TPSs (Figure 4A and Supplemental Figure 12), whereas these three positions were occupied by nonaromatic residues in most classic TPSs. We had previously demonstrated that position 328 (in AtTPS18) was occupied by Gly or Pro in plant sester-TPSs, whereas this position was occupied by an aromatic residue (always Trp) in most classic TPSs (Chen et al., 2019). Therefore, these residue features ($Gly^{328}$, $Tyr^{462}$, $Tyr^{465}$, and $Phe^{496}$) could be used to distinguish type-A sester-TPSs from classic plant TPSs (Supplemental Figure 12).

To test the effects of these aromatic residues on sester-TPS activity, we replaced the aromatic residues with Ala one by one. The results showed that most mutants (Y462A, F573A, and Y577A) completely lost their sester-TPS activity (Supplemental Data 1), implying their critical roles in stabilizing carbocation intermediates in catalysis. However, no change in total activity and product specificity was noted between the Y465A mutant and AtTPS18 (Table 1), although $Tyr^{465}$ was conserved across all characterized type-A sester-TPSs. We cannot presently arrive at a role for $Tyr^{465}$ in the catalysis of AtTPS18. The F496A mutant showed approximately 17% of the activity of AtTPS18 (Table 1). The effects of $Phe^{496}$ on sester-TPS product specificity were further investigated, as described in another section.

The discovery of type-B sester-TPS (AtTPS06) has made it is possible to compare residues in the catalytic pocket of both types of sester-TPSs. Both residue features of type-A sester-TPSs could not be applied to type-B sester-TPSs, which are similar to classic TPSs, in that only two aromatic residues, $Trp^{335}$ and $Tyr^{584}$, were found in AtTPS06, and the two positions were obviously different (occupied by aromatic or nonaromatic) in AtTPS18 ($Gly^{328}$ and $Phe^{496}$) and AtTPS06 ($Trp^{335}$ and $Ser^{503}$; Figure 4A).

It has been proposed that the initial conformation of GFPP, which is fixed in the catalytic pocket of different sester-TPSs, determines the sesterterpene cyclization mechanism (type A or type B; shown in Figure 2 [Sato et al., 2018]). To test the effects of $Trp^{335}$ and $Ser^{503}$ on GFPP accommodation in type-B sester-TPSs, we replaced these two amino acids in AtTPS06 with Gly and Phe, both individually and simultaneously. All mutants (AtTPS06$^{W335G}$, AtTPS06$^{S503F}$, and AtTPS06$^{W335G/S503F}$) completely lost their sester-TPS activity (Figure 4B), which suggested that both residues are required for correct GFPP accommodation in the catalytic pocket of AtTPS06, if not all plant type-B sester-TPSs. In addition, we exchanged the helix G region between AtTPS18 ($^{456}$TAGMDGY$^{462}$, including $Tyr^{462}$) and AtTPS06 ($^{463}$STAGEVL$^{469}$, including $Leu^{469}$), and this replacement also resulted in enzyme inactivity (Figure 4B). Similarly, all corresponding AtTPS18 mutants (with the replacement of the corresponding amino acid in AtTPS06), except AtTPS18$^{F496S}$ (no type-B sesterterpenes were found), were enzymatically inactive (Figure 4C and Table 1). These results indicate that, in addition to these two amino acids, there must be other amino acids controlling GFPP accommodation in type-B sester-TPSs. Cgr1961, isolated from *Capsella grandiflora*, was reported to be a nonclustered sester-TPS (Chen et al., 2019). Based on its sequence features (only two aromatic residues, $Trp^{302}$ and $Ser^{471}$, are found in Cgr1961; Figure 4A) and phylogenetic analyses, we propose that Cgr1961 is also a type-B sester-TPS; however, our findings will need to be validated in future structural elucidation studies of Cgr1961 products.

## Controlling Product Specificity in AtTPS18 by $Gly^{328}$ and $Phe^{496}$

Based on the importance of $Gly^{328}$ and $Phe^{496}$ in type-A sester-TPS activity, as discussed above, we further investigated the effects of other residues at these two positions on the enzymatic properties of AtTPS18. We generated 12 mutants for the $Gly^{328}$ position of AtTPS18 from $Gly^{328}$ to amino acids with hydrophobic side chains (Ala, Phe, Leu, and Trp), amino acids with positively charged side chains (Lys, His, and Arg), amino acids with negatively charged side chains (Asp), amino acids with polar

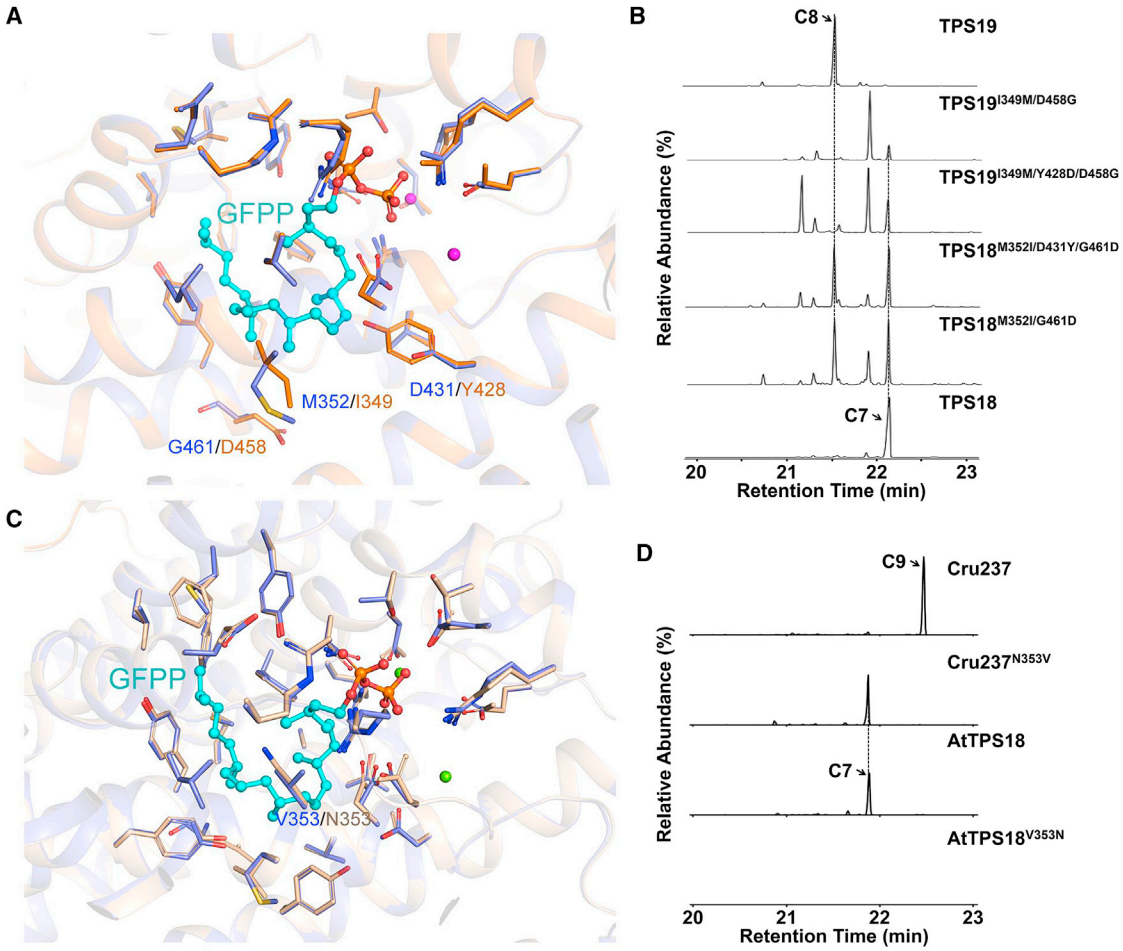



**Figure 5. Conversion between AtTPS18 and Its Close Homologs.**

**C7** ((+)-thalianatriene), **C8** ((−)-retigeranin), and **C9** ((−)-caprutriene) are the main products of AtTPS18, AtTPS19, and Cru237, respectively.

**(A)** Comparison of the models of AtTPS18 (blue) and AtTPS19 (orange). The different amino acids in the catalytic pockets of AtTPS18 and AtTPS19 are highlighted.

**(B)** GC–MS (SIM mode, *m/z* 340 for $C_{25}H_{40}$) analysis of AtTPS18, AtTPS19, and their mutants. The peaks of **C7** (the main product of AtTPS18) and **C8** (the main product of AtTPS19) are marked with arrows.

**(C)** Comparison of the models of AtTPS18 (blue) and Cru237 (orange). The amino acid at position 353 is highlighted.

**(D)** GC–MS (SIM mode, *m/z* 340 for $C_{25}H_{40}$) analysis of AtTPS18, Cru237, and their mutants.

uncharged side chains (Ser and Asn), or other amino acids (Cys and Pro). The findings showed that exchanging $Gly^{328}$ with a charged residue (Asp, Arg, His, or Lys) resulted in an enzyme that was inactive for the production of sesterterpenes, which suggested that a charged residue probably destabilizes the formation of the carbocation intermediate in the catalytic pocket (IM1 and IM2 in Figure 2). Moreover, G328F and G328W substitutions (both Phe and Trp have bulky side chains) also resulted in enzyme inactivity (Supplemental Data 1). By contrast, the G328P substitution increased the total production of sesterterpenes by at least 60% ($n$ = 3; Table 1 and Supplemental Data 1).

Using the same strategy, we generated 11 mutants for the $Phe^{496}$ position of AtTPS18 (replacing $Phe^{496}$ with Ala, Ile, Leu, Met, Val, Lys, His, Asp, Ser, Cys, or Pro). Compared to AtTPS18, all mutants showed decreased sester-TPS activity. Interestingly, the F496A mutant retained 17.3% ($n$ = 2) of the sester-TPS activity of AtTPS18; however, replacing the $Phe^{496}$ position with Ala in other Arabidopsis sester-TPSs ($AtTPS19^{F493A}$ and $AtTPS30^{F496A}$) resulted in the loss of sester-TPS activity (Supplemental Figure 13). Additionally, all $Phe^{496}$ mutants showed decreased product specificity compared to AtTPS18, as no main product accounted for >50% of the total sesterterpene production (Table 1). These findings provide a starting point for creating new sester-TPSs using directed evolution strategies. Because multiple novel sesterterpene peaks were detected in the chromatograms of the reaction mixtures obtained with AtTPS18 mutants (Supplemental Figure 14), we scaled up the production of $AtTPS18^{G328P}$ and $AtTPS18^{G328A/F496A}$ using the *E. coli* system (Shao et al., 2017). Four compounds (**C3** to **C6**) were purified and structurally elucidated by NMR spectroscopy (Supplemental Table 2; Supplemental Figures 4 and 5). Among them, **C3** ((+)-brassitetraene B) and **C6** ((+)-brassitetraene A) have been previously reported by Huang et al. (2018). **C4** is an isomer of (2*E*)-α-cericerene with a different double-bond position (Qin

et al., 2016), whereas **C5** (18*S*) is an epimer of **C7** (18*R*, (+)-thalianatriene) (Shao et al., 2017). Notably, the ratio of **C5**/**C7** isomers increased with the increase in the side-chain size of the amino acid at position 328 (Gly and Ser < Ala and Cys < Leu and Pro; Table 1). Moreover, AtTPS18$^{G328N}$ produced **C6** as the main product, although with a lower yield than AtTPS18. It is possible that the side chain of the amino acid at position 328 is close to $C_{18}$/$C_{19}$ of the carbocation intermediate, allowing it to determine the conformation of the final product. For example, Asn$^{328}$ might stabilize the formation of IM1 and facilitate the deprotonation of IM1 to produce **C6** (Figure 2).

Thus far, there have been no reports on the catalytic parameters of sester-TPSs, probably because GFPP is not commercially available. We therefore compared the catalytic efficiency of plant sester-TPSs using GFPP with their ancestor enzyme, a mono-/sesqui-TPS from the TPS-a subfamily. Here, we only determined the catalytic efficiency of purified AtTPS18 and AtTPS18$^{G328P}$ due to insufficient synthetic GFPP. AtTPS18$^{G328P}$ showed much higher substrate affinity ($K_M$ = 1.28 ± 0.49 μM, $k_{cat}$ = 0.047 ± 0.05 $s^{-1}$, $n$ = 3) than AtTPS18 ($K_M$ = 4.79 ± 1.82 μM, $k_{cat}$ = 0.11 ± 0.017 $s^{-1}$, $n$ = 3). The catalytic efficiency of AtTPS18 fell within the range previously reported for mono-TPSs or sesqui-TPSs from plants (Chen et al., 2004; Tholl et al., 2005; Nagegowda et al., 2008; Wang et al., 2008).

### Interconversion between AtTPS18 and Its Close Homologs Using Structure-Based Comparisons

Previous studies have reported that AtTPS18, AtTPS19, and Cru237 (a characterized sester-TPS from *Capsella rubella*) are similar to each other, sharing >80% identity at the protein level (Supplemental Figure 15; Shao et al., 2017; Huang et al., 2018; Chen et al., 2019). Moreover, the three sester-TPSs share common carbocation intermediates in the production of their own products (Figure 2) (Huang et al., 2017). Next, we attempted to interconvert their sester-TPS activities using the AtTPS18 structural information obtained in this study. Structural comparisons showed that there were three key amino acid pairs (Met$^{352}$ in AtTPS18 versus Ile$^{349}$ in AtTPS19; Asp$^{431}$ in AtTPS18 versus Tyr$^{428}$ in AtTPS19; Gly$^{461}$ in AtTPS18 versus Asp$^{458}$ in AtTPS19; Figure 5A). The results of site-directed mutagenesis and biochemical assays showed that a single mutant (M352I, D431Y, or G461D) did not notably change the product profile relative to wild-type AtTPS18. One of the double mutants (M352I/G461D) produced 34.5% **C8** and 36.6% **C7** (the yield was 0.89 when the yield of AtTPS18 was set to 1.0, $n$ = 2). A triple mutant (M352I/D431Y/G461D) produced 33.3% **C8** and 30.8% **C7** (the yield was 0.32, $n$ = 2; Supplemental Data 3). Moreover, one double mutant (I349M/D458G) produced 1.9% **C8** and 14.9% **C7** (the yield was 0.12 when the yield of AtTPS19 was set to 1.0, $n$ = 2), and a triple mutant (I349M/Y428D/D458G) produced 1.7% **C8** and 19.3% **C7** (the yield was 0.36, $n$ = 2).

In addition, we found that one amino acid (Val$^{353}$ [hydrophobic side chain] in AtTPS18 versus Asn$^{353}$ [polar side chain] in Cru237) is probably responsible for the product specificity between AtTPS18 and Cru237 (Figure 5C). The results of biochemical assays showed that the Cru237 mutant (N353V) showed complete AtTPS18 activity, whereas the corresponding AtTPS18 mutant (V353N) displayed inactivity (Figure 5D).

Equally important, there are still several open questions with regard to the catalytic mechanism of plant sester-TPSs, including the mutants generated in this study. For example, how are the carbocation intermediates controlled in the catalytic pocket of sester-TPSs to generate different products? Which residues are involved in the final deprotonation reaction? Computational-guided methods (e.g., quantum mechanical/molecular mechanical–molecular dynamic simulations), together with the structural features of AtTPS18, will help address these interesting questions (Zhang et al., 2020).

## Concluding Remarks

Sester-TPSs, an evolutionarily new clade of the TPS-a subfamily, are not Brassicaceae-specific; instead, they are widely distributed in the plant kingdom (Liu et al., 2016). To date, approximately 140 sesterterpenoids have been identified; however, the chemical diversity of the plant sesterterpenes is expected to be much larger than that initially assumed due to the possible rearrangements of the C25 carbocation intermediate by sester-TPSs. More importantly, further studies are needed to investigate the physiological functions of sesterterpenoids from both plants and microbes (Wei et al., 2004; Luo et al., 2010). In this study, we revealed the general catalytic mechanism of plant sester-TPSs by combining chemical structure elucidation, enzyme structure analysis, and biochemical assays. The novel sesterterpenoids generated by the newly mutated sester-TPSs will significantly expand the diversity of terpenoids, which is important for further applications. As sester-TPSs have also been identified in microbes (Matsuda et al., 2015; Ye et al., 2015; Tian et al., 2017), whether microbe sester-TPSs (PT-TPS fusion proteins) utilize a catalytic mechanism similar to that of plant sester-TPSs is an interesting topic that should be addressed in future studies.

## METHODS

### Plant Materials, RNA Analysis, and Chemicals

*A. thaliana* plants (Columbia-0 ecotype) were grown in soil in a growth room or on half-strength solid Murashige–Skoog medium plates in a growth chamber at 22°C under a 16-h light/8-h dark photoperiod. RNA extraction, reverse transcription reactions, and gene cloning were performed as described previously (Chen et al., 2015).

All chemicals used in this study were purchased from Sigma-Aldrich (USA) except FSPP, which was purchased from Echelon Biosciences (USA). GFPP (1 mg) was chemically synthesized following a published protocol (Sato et al., 2013). All DNA constructs generated in this study were verified by sequencing.

### Gene Cloning, Expression in *E. coli*, and Biochemical Assays

All truncated Arabidopsis TPS-a or mutated genes (signal peptide sequence was removed; for detailed primer information, see Supplemental Data 2) were cloned into the pMAL-c2x vector (New England Biolabs). An MBP tag was added at the C-terminus of each truncated *TPS* gene. *At*GFPPS2 (*At3g14550*) was ligated into the pET-28a vector. To increase sesterterpene production, we cotransformed three plasmids, *pMBIS-Tc$^r$* containing five MVA genes (mevalonate kinase, phosphomevalonate kinase, mevalonate pyrophosphate decarboxylase, IPP isomerase, and FPP synthase [Martin et al., 2003]), *pET-Kan$^r$-GFPPS2*, and *pMAL-Amp$^r$-TPSa* into the *E. coli* JM109 (DE3) strain. The transformed *E. coli* harboring *pMBIS-Tc$^r$*, *pET-Kan$^r$-AtGFPPS2*, and *pMAL-Amp$^r$* (empty vector) were used as negative controls. The *E. coli* system (200 ml in Tris–borate [TB] buffer with the appropriate

antibiotics) was induced by the addition of 0.25 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 10 mM mevalonolactone to an $OD_{600}$ (optical density at a wavelength of 600 nm) of 0.6. After 48 h of induction at 16°C, the cells were harvested by centrifugation at 5000 *g* for 20 min. The harvested cells were resuspended in 12 ml of Tris–HCl buffer (pH 8.0) in a 50-ml conical tube and lysed with a probe sonicator. The resulting solution was then extracted with 20 ml of EtOAc by vortexing for 1 h. The EtOAc phase was collected and concentrated to 200 μl under flowing nitrogen. Sesterterpene analyses by triple-quadrupole GC–MS (GC-QQQ-MS; Agilent 7890B/7000C, SIM mode *m/z* 340) were performed as described previously (Chen et al., 2019). The sum of the areas of the chromatographic peaks of the main sesterterpene compounds produced by AtTPS18 and its mutants was calculated as the total yield after normalization to the internal standard and culture density ($OD_{600}$ value).

The soluble recombinant sester-TPSs (AtTPS06, AtTPS18, AtTPS19, and Cru237) and their mutants in *E. coli* were quantitatively validated using western blotting with an anti-MBP monoclonal antibody (New England Biolabs). In brief, soluble crude proteins (20 ng) from each *E. coli* strain were separated by 8% SDS–PAGE, transferred onto a polyvinylidene fluoride membrane, and blotted with an anti-MBP antibody (1:50 000 dilution) following a published protocol (Chen et al., 2015).

To determine the catalytic efficiency of AtTPS18 and At $TPS18^{G328P}$, we used 2 μg of purified recombinant protein in each biochemical assay (200-μl volume, buffer contained 10 mM $Mg^{2+}$, 2 mM dithiothreitol, and 50 mM Tris–HCl [pH 7.5]) and a range of GFPP concentrations (0, 1, 2, 5, 10, 20, 50, and 100 ng/μl). After 1 h of incubation at 30°C, the products were extracted twice with 200 μl of EtOAc. The extracted sesterterpenes were combined and analyzed by GC-QQQ-MS (MRM, 340.0 > 119.0) using the aforementioned program. The target sesterterpenes were quantified by comparing with standard curves that were generated with **C5** ($AtTPS18^{G328P}$ main product) and **C7** (AtTPS18 main product). The $K_M$ and $k_{cat}$ of each protein were determined by extrapolating the values from Lineweaver–Burk plots. The enzymatic assays were carried out in triplicate.

### Purification of Sesterterpenes and Elucidation of Chemical Structures

The *E. coli* JM109 (DE3) strain harboring *pMAL-c2x-AtTPS06*, *AtTPS18* (and its mutants, $TPS18^{G328P}$ for **C5** preparation and $TPS18^{G328A/F496A}$ for **C3**/**C4**/**C6** preparation), or *AtTPS29*, *pET-28a-AtGFPPS2*, and *pMBIS-Tc*$^r$ plasmids (as mentioned above) was incubated in 1 l of TB medium containing 100 μg/ml of ampicillin (sodium salt), 20 μg/ml of tetracycline, and 50 μg/ml of kanamycin sulfate at 37°C and agitated at 130 rpm in a 2-l Erlenmeyer flask. When the cultures reached an $OD_{600}$ of 0.4–0.5, 250 μl of 1 M IPTG and 8 ml of 50 mg/ml mevalonolactone were added in two allotments, followed by incubation for 48 h at 16°C/130 rpm. The cultured cells were harvested by centrifugation and lyophilized. The lyophilized cells from 8 l of culture media were extracted with acetone. The extract was concentrated, re-extracted with *n*-hexane, and repeatedly subjected to silica-gel column chromatography. The sesterterpenes were eluted with *n*-hexane to yield different amounts of pure **C1** to **C6** (all were obtained as colorless oils; for the NMR data of **C1**, **C2**, **C4**, and **C5**, see Supplemental Tables 1 and 2 and Supplemental Figures 2–5).

### Large-Scale AtTPS18 Production, Purification, and Crystallization

*AtTPS18* was inserted into the pET-Duet vector with a 6×His tag at the N-terminus. The plasmid was transformed into the *E. coli* BL21 (DE3) strain. The transformed bacterial cells were grown in Luria–Bertani medium containing ampicillin at 37°C and induced by 0.25 mM IPTG for 12 h at 18°C. The cells were harvested and resuspended in buffer A (20 mM Tris–HCl [pH 8.0] and 100 mM NaCl) supplemented with 1 mM phenylmethylsulfonyl fluoride. Subsequently, the cells were lysed with a high-pressure cell disruptor at 18 000 pounds per square inch (psi), and the lysate was centrifuged at 20 000 *g* for 45 min. The supernatant was loaded onto a $Ni^{2+}$–nitrilotriacetic acid affinity column (Qiagen) and washed with buffer A containing 20 mM imidazole. The proteins were eluted with buffer A containing 250 mM imidazole and purified by gel filtration using a Superdex 200 column (GE Healthcare) with buffer A. The peak fractions were collected and concentrated for subsequent structural and biochemical studies.

To obtain crystals of AtTPS18 in a complex with its substrates, we incubated the protein with a 10-fold molar excess of GFPP or FSPP on ice for 30 min before crystallization. Crystals of the AtTPS18 complex were grown at 20°C using the sitting-drop vapor-diffusion method by mixing 0.5 μl of the protein suspension with 0.5 μl of a solution containing 16% (w/v) polyethylene glycol 8000, 0.04 M potassium phosphate (monobasic), and 20% (v/v) glycerol. Crystals for data collection were directly flash-frozen under a nitrogen stream at 100 K. The data for the AtTPS18 complex were collected using the beamline BL19U1/17U1 at the Shanghai Synchrotron Radiation Facility and processed using the HKL3000 package. The structures of the AtTPS18 complex were solved by molecular replacement using Phenix with tobacco 5-epi-aristolochene synthase (PDB: 5EAT) as the initial structure. All models were refined with Phenix (Adams et al., 2010) and manually built with Coot (Emsley and Cowtan, 2004). The data collection and refinement statistics are summarized in Supplemental Table 3. The protein structures generated in this study have been deposited in the world-wide protein data bank (AtTPS18-PP (PDB code 7BZB) and AtTPS18-FSPP (PDB code 7BZC).

### Homology Modeling

Homology models of Arabidopsis sester-TPSs were generated using the modeling server SWISS-MODEL (https://swissmodel.expasy.org/) with the AtTPS18–FSPP structure serving as a template (Waterhouse et al., 2018). GFPP docking and C25 prenyl intermediate carbocation were carried out using AutoDock4.2.6 with AutoDockTools (Morris et al., 2009).

## PHYLOGENETIC ANALYSES

A total of 32 Arabidopsis TPS protein sequences were extracted from The Arabidopsis Information Resource (http://www.arabidopsis.org). A maximum-likelihood tree was constructed using MEGA 6.0 (Tamura et al., 2013).

## SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Plant Communications Online*.

## FUNDING

This work was financially supported by the National Key R&D Program of China (grant no. 2018YFA0900600), the National Natural Science Foundation of China (grant nos. 31970315 and 31700263), the "Priority Research Program" of the Chinese Academy of Sciences (grant nos. ZDRW-ZS-2019-2 and XDB27020103), the Grant-in-Aid Program for Scientific Research from the MEXT, Japan (JSPS KAKENHI grant no. JP16H06443), and the State Key Laboratory of Plant Genomics of China (grant no. SKLPG2016A-13). J.L. was supported by the Foundation of Youth Innovation Promotion Association of the Chinese Academy of Sciences.

## AUTHOR CONTRIBUTIONS

G.W. and P.Z. designed the research. Z.L., Q.C., and Y.M. performed the gene cloning, site-directed mutagenesis, protein expression, and enzymatic assays. J.L. performed the crystallization, diffraction, data collection, and refinement. Q.C. prepared the sesterterpenes on a large scale using the *E. coli* system. T.M. and I.A. carried out the purification and structural elucidation of **C1** and **C2**. Haili Liu and Y.W. carried out the

purification and structural elucidation of **C3**. Y.Z. and Hongwei Liu carried out the purification and structural elucidation of **C4**, **C5**, and **C6**. T. Shinada and T. Sato synthesized the GFPP. Q.C., J.L., Z.L., P.Z., and G.W. analyzed the data. G.W. wrote the manuscript with input from all authors.

## ACKNOWLEDGMENTS

No conflict of interest declared.

Received: March 20, 2020
Revised: April 10, 2020
Accepted: April 26, 2020
Published: April 29, 2020

## REFERENCES

**Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al.** (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. **66**:213–221.

**Bian, G.K., Han, Y.C., Hou, A.W., Yuan, Y.J., Liu, X.H., Deng, Z.X., and Liu, T.G.** (2017). Releasing the potential power of terpene synthases by a robust precursor supply platform. Metab. Eng. **42**:1–8.

**Chai, H.Z., Yin, R., Liu, Y.F., Meng, H.Y., Zhou, X.Q., Zhou, G.L., Bi, X.P., Yang, X., Zhu, T.H., Zhu, W.M., et al.** (2016). Sesterterpene ophiobolin biosynthesis involving multiple gene clusters in *Aspergillus ustus*. Sci. Rep. **6**:27181.

**Chen, F., Ro, D.K., Petri, J., Gershenzon, J., Bohlmann, J., Pichersky, E., and Tholl, D.** (2004). Characterization of a root-specific *Arabidopsis* terpene synthase responsible for the formation of the volatile monoterpene 1,8-cineole. Plant Physiol. **135**:1956–1966.

**Chen, F., Tholl, D., Bohlmann, J., and Pichersky, E.** (2011). The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. Plant J. **66**:212–229.

**Chen, Q., Fan, D., and Wang, G.** (2015). Heteromeric geranyl(geranyl) diphosphate synthase is involved in monoterpene biosynthesis in *Arabidopsis* flowers. Mol. Plant **8**:1434–1437.

**Chen, Q., Jiang, T., Liu, Y.X., Liu, H.L., Zhao, T., Liu, Z.X., Gan, X.C., Hallab, A., Wang, X.M., He, J., et al.** (2019). Recently duplicated sesterterpene (C25) gene clusters in *Arabidopsis thaliana* modulate root microbiota. Sci. China Life Sci. **62**:947–958.

**Chiba, R., Minami, A., Gomi, K., and Oikawa, H.** (2013). Identification of ophiobolin F synthase by a genome mining approach: a sesterterpene synthase from *Aspergillus clavatus*. Org. Lett. **15**:594–597.

**Christianson, D.W.** (2017). Structural and chemical biology of terpenoid cyclases. Chem. Rev. **117**:11570–11648.

**Emsley, P., and Cowtan, K.** (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. **60**:2126–2132.

**Gao, Y., Honzatko, R.B., and Peters, R.J.** (2012). Terpenoid synthase structures: a so far incomplete view of complex catalysis. Nat. Prod. Rep. **29**:1153–1175.

**He, J., Chen, Q.W., Xin, P.Y., Yuan, J., Ma, Y.H., Wang, X.M., Xu, M.M., Chu, J.F., Peters, R.J., and Wang, G.D.** (2019). CYP72A enzymes catalyse 13-hydroxylation of gibberellins. Nat. Plants **5**:1057–1065.

**Huang, A.C., Hong, Y.J., Bond, A.D., Tantillo, D.J., and Osbourn, A.** (2018). Diverged plant terpene synthases reroute the carbocation cyclization path towards the formation of unprecedented 6/11/5 and 6/6/7/5 sesterterpene scaffolds. Angew. Chem. Int. Ed. **57**:1291–1295.

**Huang, A.C., Kautsar, S.A., Hong, Y.J., Medema, M.H., Bond, A.D., Tantillo, D.J., and Osbourn, A.** (2017). Unearthing a sesterterpene biosynthetic repertoire in the Brassicaceae through genome mining reveals convergent evolution. Proc. Natl. Acad. Sci. U S A **114**:E6005–E6014.

**Jurcik, A., Bednar, D., Byska, J., Marques, S.M., Furmanova, K., Daniel, L., Kokkonen, P., Brezovsky, J., Strnad, O., Stourac, J., et al.** (2018). CAVER Analyst 2.0: analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. Bioinformatics **34**:3586–3588.

**Koksal, M., Jin, Y.H., Coates, R.M., Croteau, R., and Christianson, D.W.** (2011). Taxadiene synthase structure and evolution of modular architecture in terpene biosynthesis. Nature **469**:116–U138.

**Kumar, R.P., Morehouse, B.R., Matos, J.O., Malik, K., Lin, H.K., Krauss, I.J., and Oprian, D.D.** (2017). Structural characterization of early Michaelis complexes in the reaction catalyzed by (+)-limonene synthase from *Citrus sinensis* using fluorinated substrate analogues. Biochemistry **56**:1716–1725.

**Li, J.X., Fang, X., Zhao, Q., Ruan, J.X., Yang, C.Q., Wang, L.J., Miller, D.J., Faraldos, J.A., Allemann, R.K., Chen, X.Y., et al.** (2013). Rational engineering of plasticity residues of sesquiterpene synthases from *Artemisia annua*: product specificity and catalytic efficiency. Biochem. J. **451**:417–426.

**Liu, Y., Luo, S.H., Schmidt, A., Wang, G.D., Sun, G.L., Grant, M., Kuang, C., Yang, M.J., Jing, S.X., Li, C.H., et al.** (2016). A geranylfarnesyl diphosphate synthase provides the precursor for sesterterpenoid (C-25) formation in the glandular trichomes of the mint species *Leucosceptrum canum*. Plant Cell **28**:804–822.

**Luo, S.H., Luo, Q., Niu, X.M., Xie, M.J., Zhao, X., Schneider, B., Gershenzon, J., and Li, S.H.** (2010). Glandular trichomes of *Leucosceptrum canum* harbor defensive sesterterpenoids. Angew. Chem. Int. Ed. **49**:4471–4475.

**Martin, V.J.J., Pitera, D.J., Withers, S.T., Newman, J.D., and Keasling, J.D.** (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. Nat. Biotechol. **21**:796–802.

**Matsuda, Y., Mitsuhashi, T., Lee, S., Hoshino, M., Mori, T., Okada, M., Zhang, H.P., Hayashi, F., Fujita, M., and Abe, I.** (2016). Astellifadiene: structure determination by NMR spectroscopy and crystalline sponge method, and elucidation of its biosynthesis. Angew. Chem. Int. Ed. **55**:5785–5788.

**Matsuda, Y., Mitsuhashi, T., Quan, Z.Y., and Abe, I.** (2015). Molecular basis for stellatic acid biosynthesis: a genome mining approach for discovery of sesterterpene synthases. Org. Lett. **17**:4644–4647.

**Minami, A., Ozaki, T., Liu, C.W., and Oikawa, H.** (2018). Cyclopentane-forming di/sesterterpene synthases: widely distributed enzymes in bacteria, fungi, and plants. Nat. Prod. Rep. **35**:1330–1346.

**Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S., and Olson, A.J.** (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J. Comput. Chem. **30**:2785–2791.

**Nagegowda, D.A., Gutensohn, M., Wilkerson, C.G., and Dudareva, N.** (2008). Two nearly identical terpene synthases catalyze the formation of nerolidol and linalool in snapdragon flowers. Plant J. **55**:224–239.

**Naya, Y., Yoshihara, K., Iwashita, T., Komura, H., Nakanishi, K., and Hata, Y.** (1981). Unusual sesterterpenoids from the secretion of *Ceroplastes floridensis* (Coccidae), an orchard pest. Application of the allylic benzoate method for determination of absolute configuration. J. Am. Chem. Soc. **103**:7009–7011.

**Qin, B., Matsuda, Y., Mori, T., Okada, M., Quan, Z.Y., Mitsuhashi, T., Wakimoto, T., and Abe, I.** (2016). An unusual chimeric diterpene synthase from *Emericella variecolor* and its functional conversion into a sesterterpene synthase by domain swapping. Angew. Chem. Int. Ed. **55**:1658–1661.

**Sato, H., Mitsuhashi, T., Yamazaki, M., Abe, I., and Uchiyama, M.** (2018). Computational studies on biosynthetic carbocation

rearrangements leading to quiannulatene: initial conformation regulates biosynthetic route, stereochemistry, and skeleton type. Angew. Chem. Int. Ed. **57**:14752–14757.

**Sato, T., Yamaga, H., Kashima, S., Murata, Y., Shinada, T., Nakano, C., and Hoshino, T.** (2013). Identification of novel sesterterpene/triterpene synthase from *Bacillus clausii*. Chembiochem **14**:822–825.

**Shao, J., Chen, Q.W., Lv, H.J., He, J., Liu, Z.F., Lu, Y.N., Liu, H.L., Wang, G.D., and Wang, Y.** (2017). (+)-Thalianatriene and (-)-retigeranin B catalyzed by sesterterpene synthases from *Arabidopsis thaliana*. Org. Lett. **19**:1816–1819.

**Srividya, N., Davis, E.M., Croteau, R.B., and Lange, B.M.** (2015). Functional analysis of (4S)-limonene synthase mutants reveals determinants of catalytic outcome in a model monoterpene synthase. Proc. Natl. Acad. Sci. U S A **112**:3332–3337.

**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.

**Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S.** (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. **30**:2725–2729.

**Tholl, D., Chen, F., Petri, J., Gershenzon, J., and Pichersky, E.** (2005). Two sesquiterpene synthases are responsible for the complex mixture of sesquiterpenes emitted from *Arabidopsis* flowers. Plant J. **42**:757–771.

**Tholl, D., and Lee, S.** (2011). Terpene specialized metabolism in *Arabidopsis thaliana*. *Arabidopsis* book **9**:e0143.

**Tian, W., Deng, Z.X., and Hong, K.** (2017). The biological activities of sesterterpenoid-type ophiobolins. Mar. Drugs **15**:229.

**Wang, C., Chen, Q., Fan, D., Li, J., Wang, G., and Zhang, P.** (2016a). Structural analyses of short-chain prenyltransferases identify an evolutionarily conserved GFPPS clade in Brassicaceae plants. Mol. Plant **9**:195–204.

**Wang, G., Tian, L., Aziz, N., Broun, P., Dai, X., He, J., King, A., Zhao, P.X., and Dixon, R.A.** (2008). Terpene biosynthesis in glandular trichomes of hop. Plant Physiol. **148**:1254–1266.

**Wang, Q., Jia, M.R., Huh, J.H., Muchlinski, A., Peters, R.J., and Tholl, D.** (2016b). Identification of a dolabellane type diterpene synthase and other root-expressed diterpene synthases in *Arabidopsis*. Front. Plant Sci. **7**:1761.

**Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al.** (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. **46**:W296–W303.

**Wei, H., Itoh, T., Kinoshita, M., Nakai, Y., Kurotaki, M., and Kobayashi, M.** (2004). Cytotoxic sesterterpenes, 6-epi-ophiobolin G and 6-epi-ophiobolin N, from marine derived fungus *Emericella variecolor* GF10. Tetrahedron **60**:6015–6019.

**Weng, J.K.** (2014). The evolutionary paths towards complexity: a metabolic perspective. New Phytol. **201**:1141–1149.

**Ye, Y., Minami, A., Mandi, A., Liu, C.W., Taniguchi, T., Kuzuyama, T., Monde, K., Gomi, K., and Oikawa, H.** (2015). Genome mining for sesterterpenes using bifunctional terpene synthases reveals a unified intermediate of di/sesterterpenes. J. Am. Chem. Soc. **137**:11846–11853.

**Zhang, F., An, T.Y., Tang, X.W., Zi, J.C., Luo, H.B., and Wu, R.B.** (2020). Enzyme promiscuity versus fidelity in two sesquiterpene cyclases (TEAS versus ATAS). ACS Catal. **10**:1470–1484.

**Zhou, F., and Pichersky, E.** (2020). More is better: the diversity of terpene metabolism in plants. Curr. Opin. Plant Biol. **55**:1–10.

**Plant Communications, Volume *1***

## Supplemental Information

# Molecular Basis for Sesterterpene Diversity Produced by Plant Terpene Synthases

**Qingwen Chen, Jianxu Li, Zhixi Liu, Takaaki Mitsuhashi, Yuting Zhang, Haili Liu, Yihua Ma, Juan He, Tetsuro Shinada, Tsutomu Sato, Yong Wang, Hongwei Liu, Ikuro Abe, Peng Zhang, and Guodong Wang**

## Supplementary Information for

# Molecular Basis for Sesterterpene (C25) Diversity Produced by Plant Terpene Synthases

Qingwen Chen [a,1], Jianxu Li [b,1], Zhixi Liu [a,f,1], Takaaki Mitsuhashi [c], Yuting Zhang [d], Haili Liu [e], Yihua Ma [a,f], Juan He [a,f], Tetsuro Shinada [g], Tsutomu Sato [h], Yong Wang [e], Hongwei Liu [d], Ikuro Abe [c], Peng Zhang [b,2] and Guodong Wang [a,f, 2]

[a] State Key Laboratory of Plant Genomics and National Center for Plant Gene Research, Institute of Genetics and Developmental Biology, The Innovative Academy of Seed Design, Chinese Academy of Sciences, Beijing 100101, China

[b] National Key Laboratory of Plant Molecular Genetics, Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China

[c] Graduate School of Pharmaceutical Sciences, University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

[d] State Key Laboratory of Mycology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, 100101, China

[e] Key Laboratory of Synthetic Biology, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China

[f] University of Chinese Academy of Sciences, Beijing 100039, China

[g] Graduate School of Science, Osaka City University, 3-3-138 Sugimoto, Sumiyoshi, Osaka 558-8585, Japan

[h] Department of Applied Biological Chemistry, Faculty of Agriculture, Graduate School of Science and Technology, Niigata University, 8050 Ikarashi-2, Niigata 950-2181, Japan

[1] These authors contributed equally to this work.

[2] To whom correspondence should be addressed:

E-mail: gdwang@genetics.ac.cn or pengzhang01@sibs.ac.cn

**This PDF file includes:**

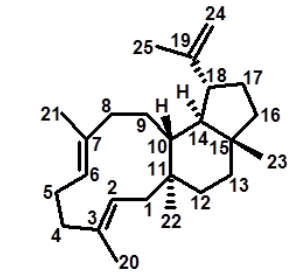
Supplemental Figures. S1 to S15

Supplemental Tables. S1 to S3

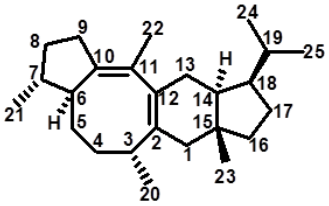
**Other supplementary materials for this manuscript include the following:**

Supplemental Datasets. S1 to S3

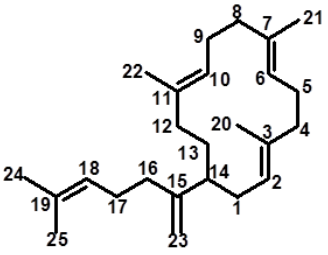
## Supplemental Figures



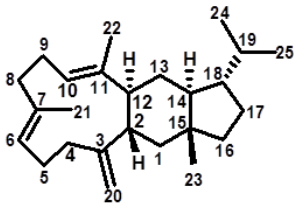
Compound 1



Compound 2



Compound 4



Compound 5

**Figure S1. Four novel sesterterpenes (C1, C2, C4, and C5) elucidated in this study (detailed NMR data see Figures S2-S5, Tables S1-S2).**

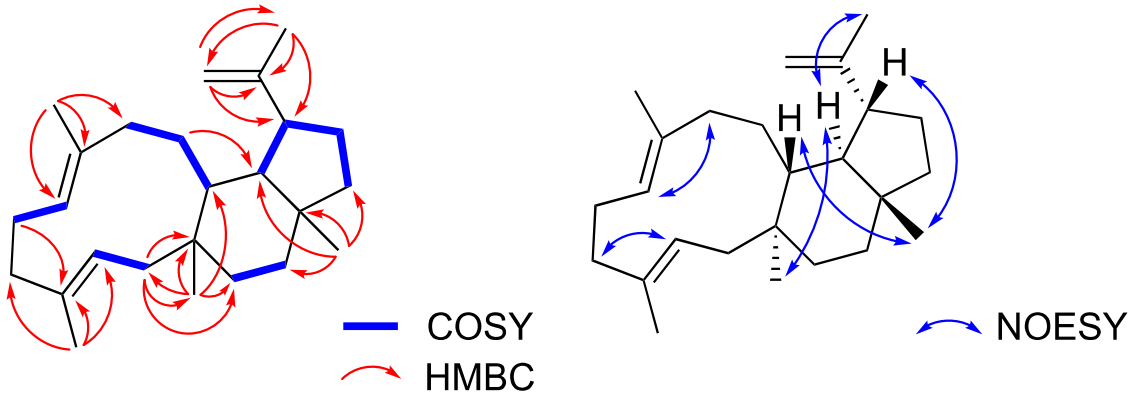



**Figure S2. $^{1}$H-$^{1}$H COSY, key HMBC, and Key NOESY correlations in C1.**

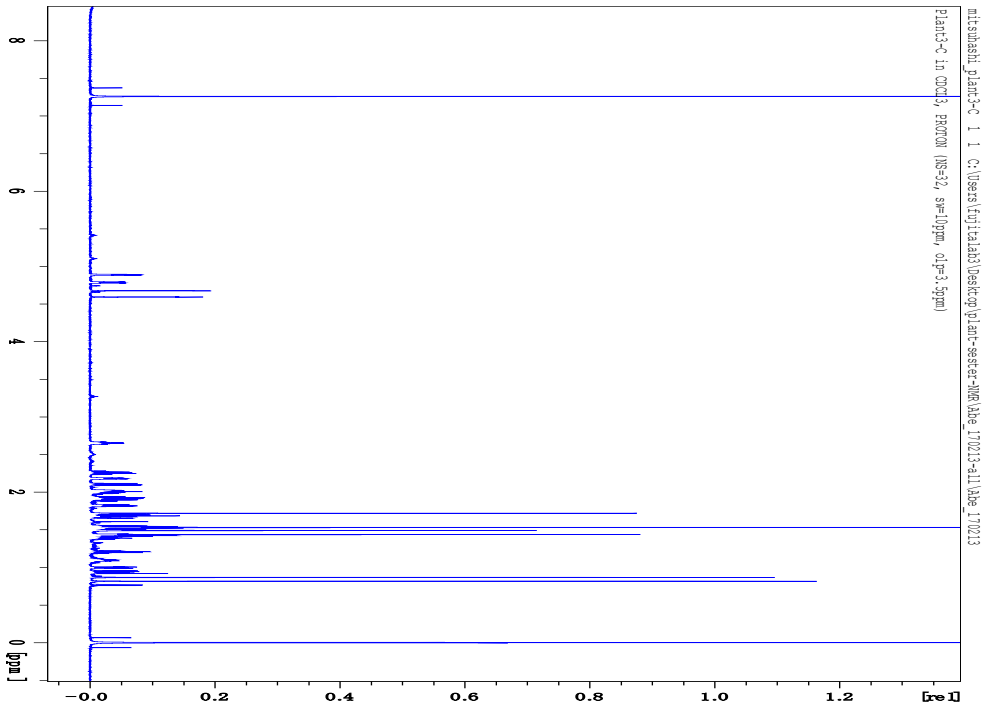



**Figure S2-1. $^{1}$H NMR spectrum of C1 in $CDCl_3$ (900 MHz).**

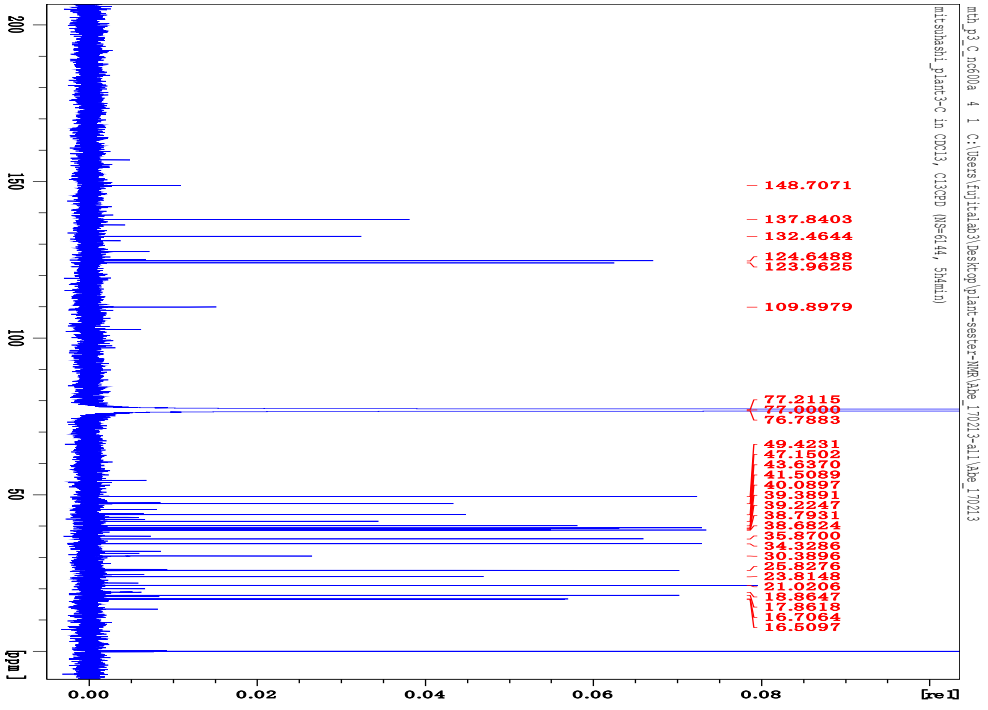



**Figure S2-2. $^{13}$C NMR spectrum of C1 in $CDCl_3$ (200 MHz).**

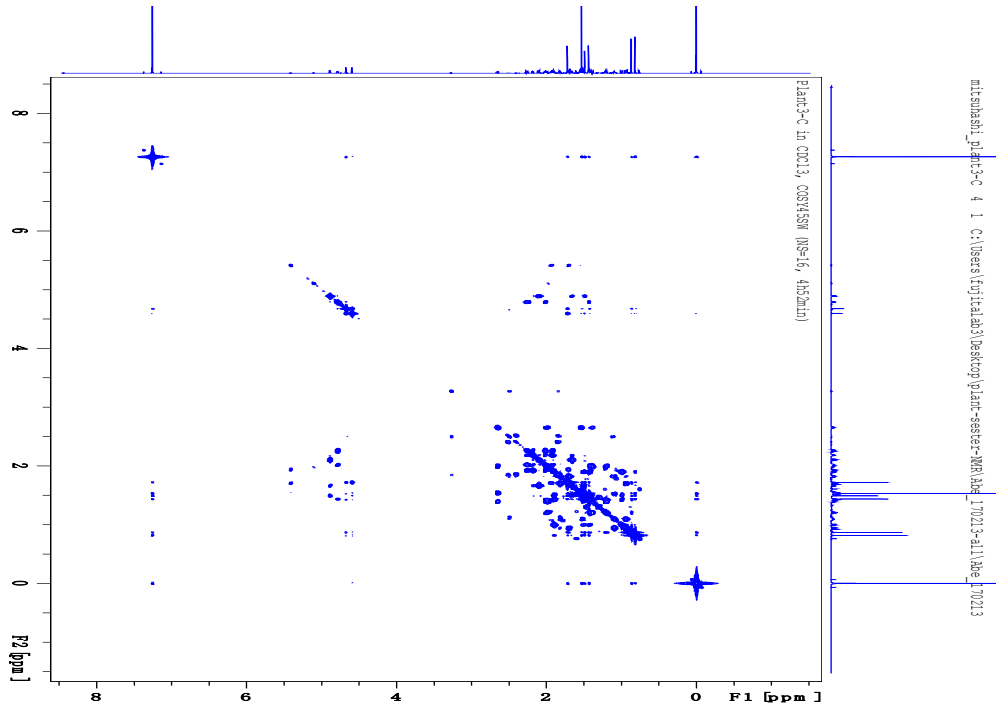



**Figure S2-3. $^{1}$H-$^{1}$H COSY spectrum of C1 in $CDCl_3$ (900 MHz).**

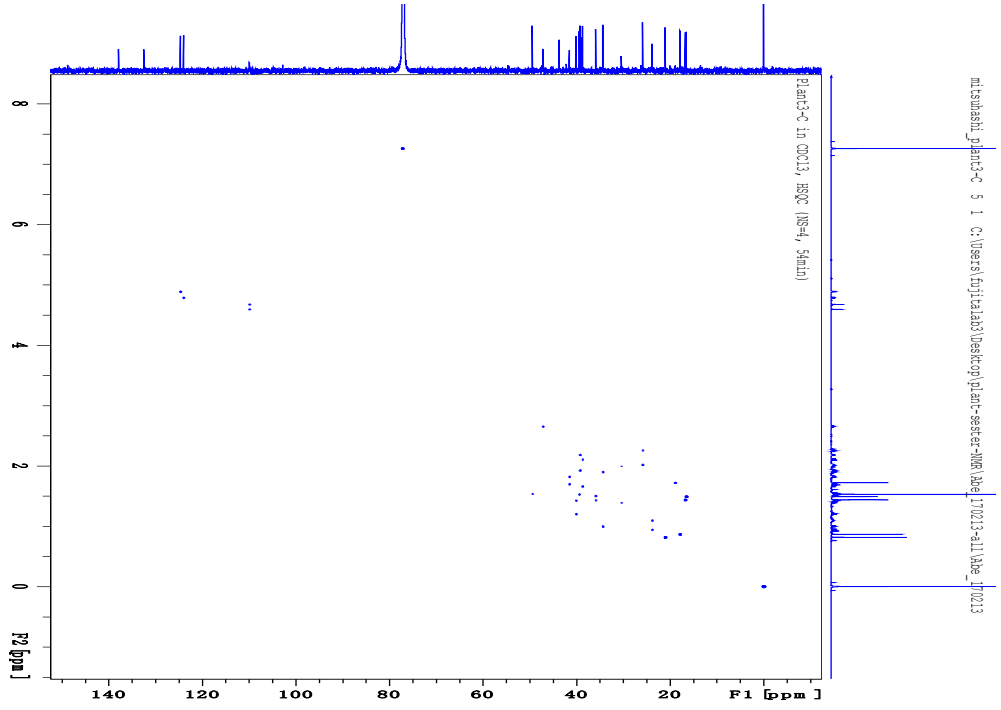



**Figure S2-4. HSQC spectrum of C1 in $CDCl_3$ (900 MHz).**

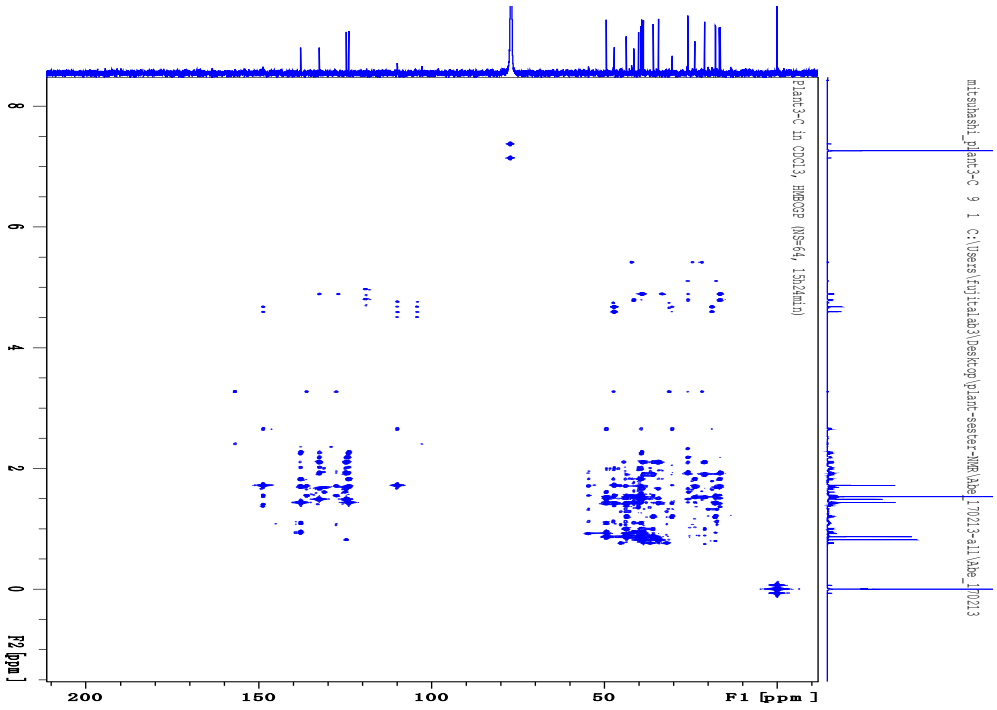



**Figure S2-5. HMBC spectrum of C1 in $CDCl_3$ (900 MHz).**

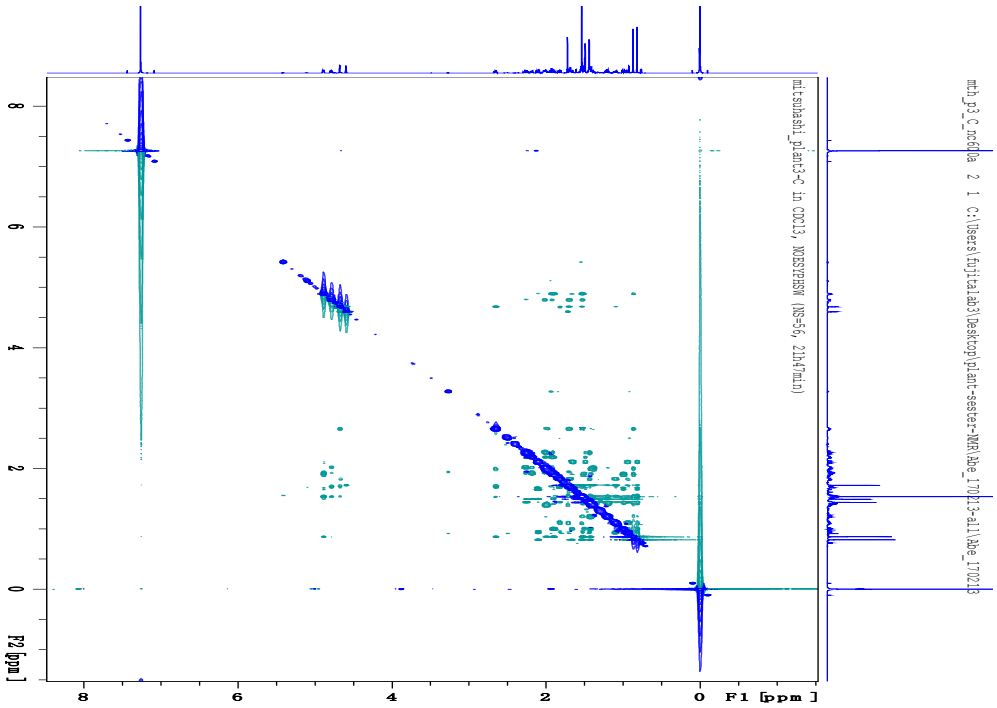



**Figure S2-6. NOESY spectrum of C1 in $CDCl_3$ (600 MHz).**

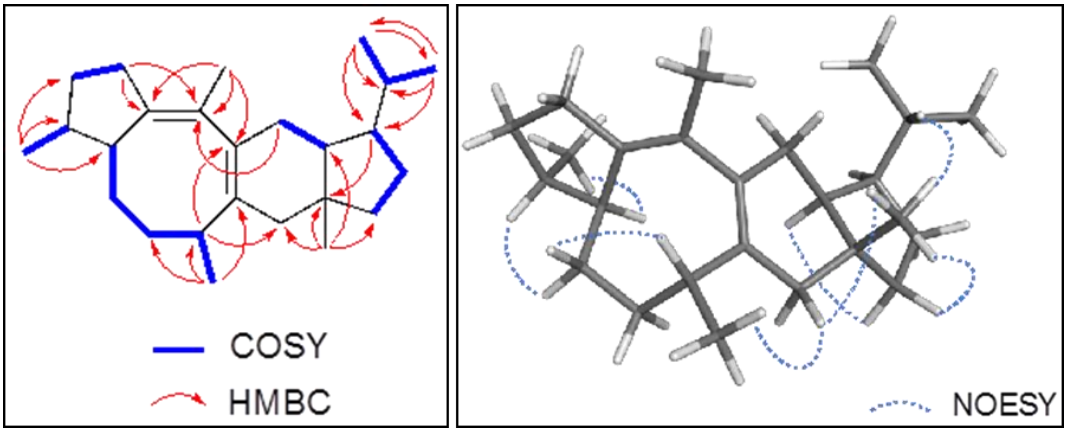



**Figure S3. $^{1}$H-$^{1}$H COSY, key HMBC, and Key NOESY correlations in C2.**

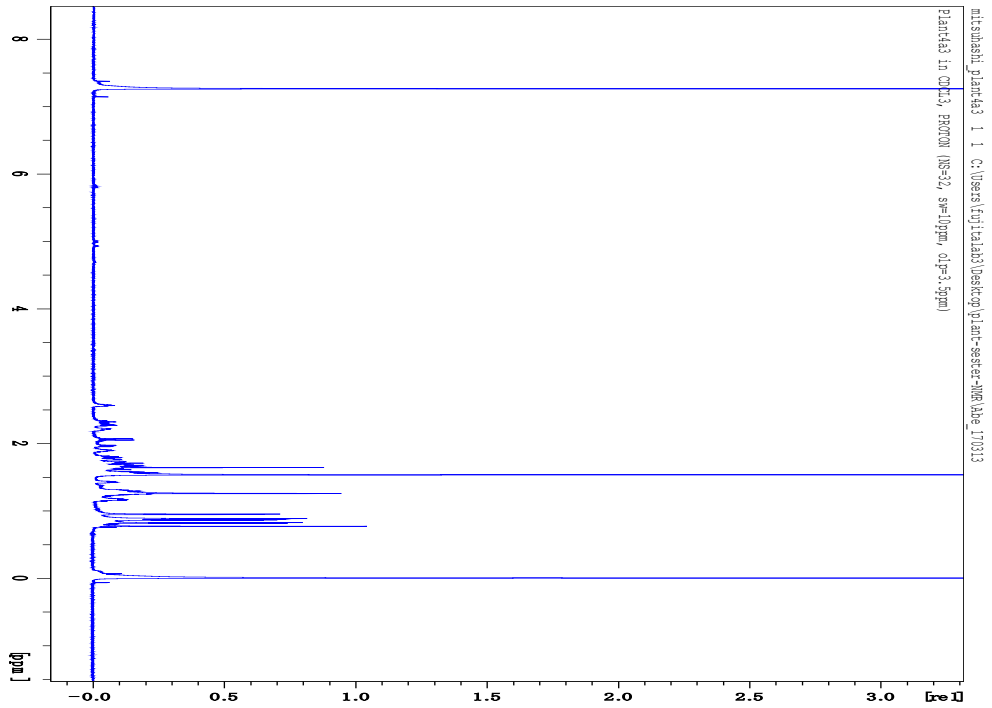



**Figure S3-1. $^{1}$H NMR spectrum of C2 in $CDCl_3$ (900 MHz).**

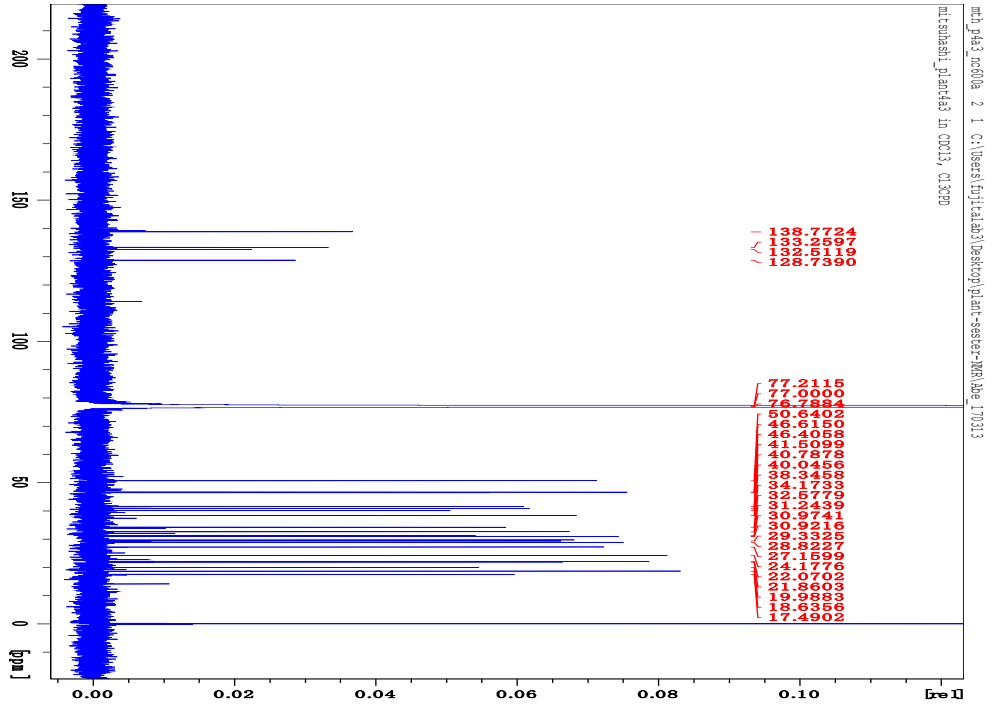



**Figure S3-2. $^{13}$C NMR spectrum of C2 in $CDCl_3$ (200 MHz).**

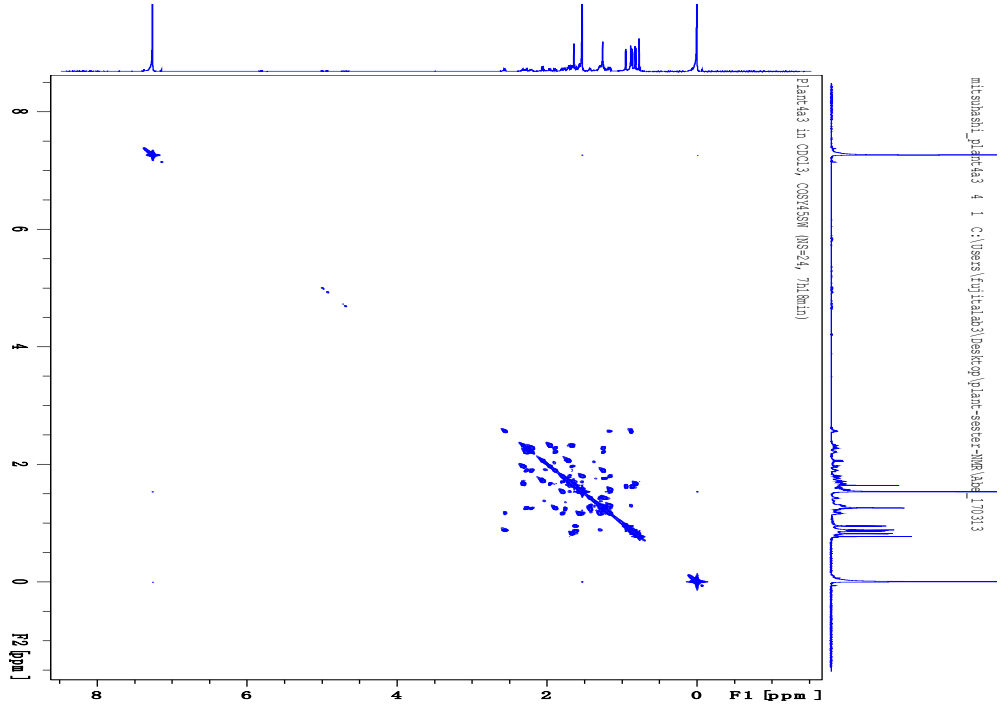



**Figure S3-3. $^{1}$H-$^{1}$H COSY spectrum of C2 in $CDCl_3$ (900 MHz).**

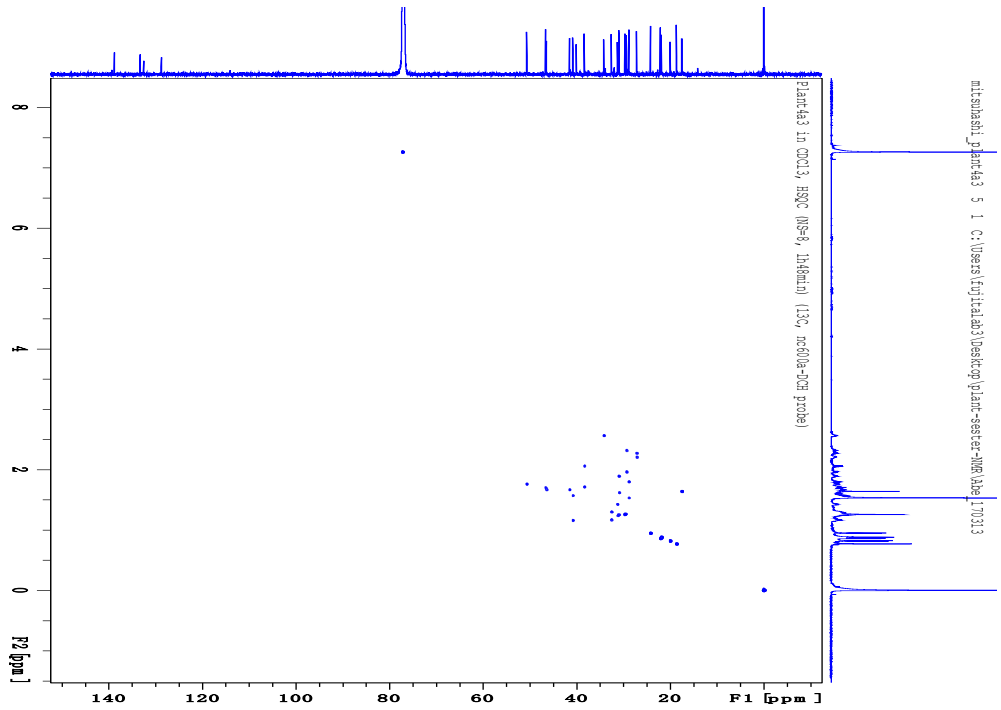



**Figure S3-4. HSQC spectrum of C2 in $CDCl_3$ (900 MHz).**

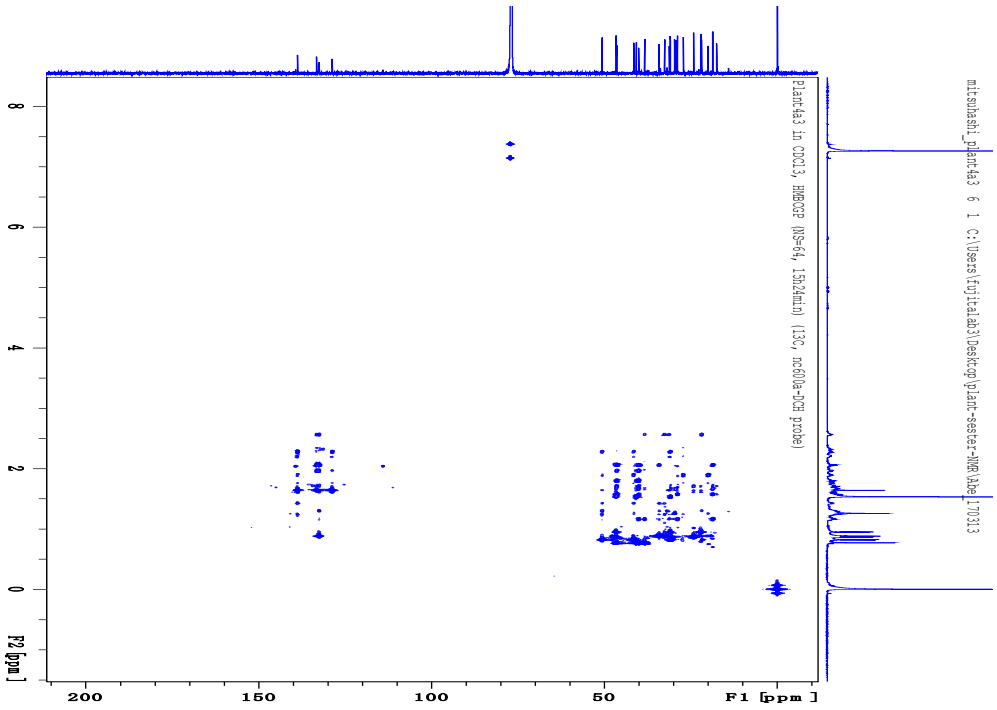



**Figure S3-5. HMBC spectrum of C2 in $CDCl_3$ (900 MHz).**

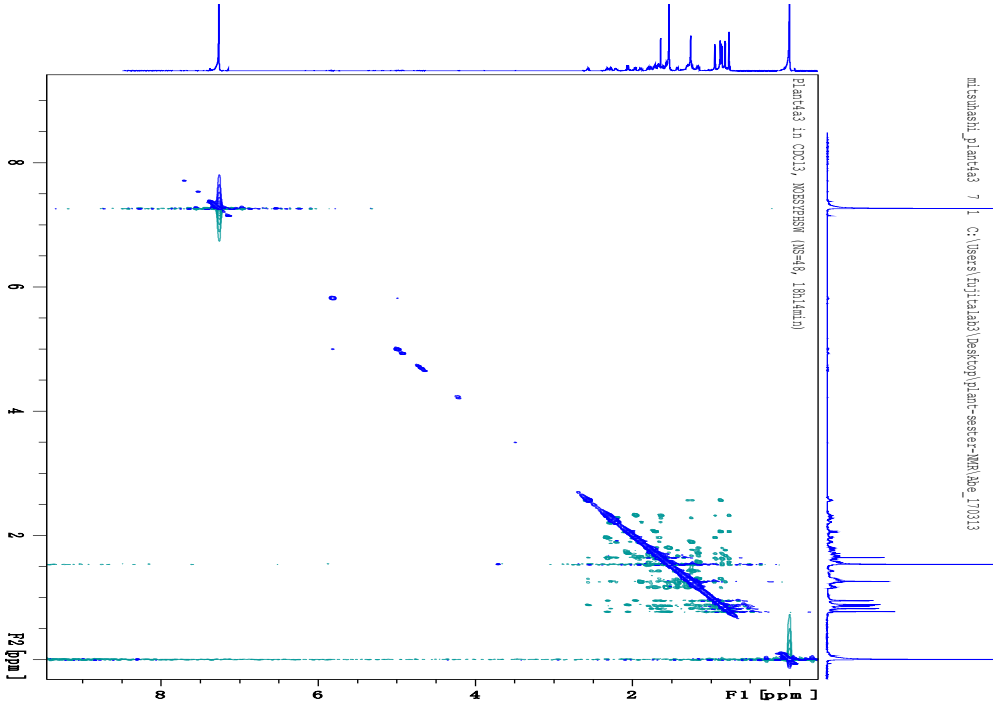



**Figure S3-6. NOESY spectrum of C2 in $CDCl_3$ (900 MHz).**

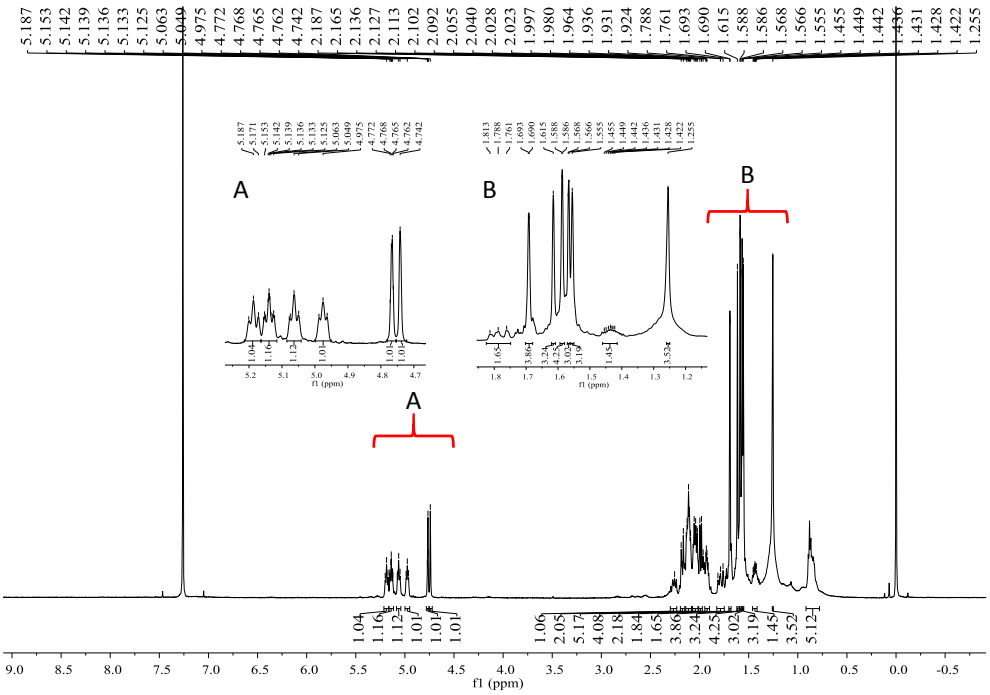



**Figure S4-1. $^{1}$H NMR spectrum of C4 in $CDCl_3$ (500 MHz).**

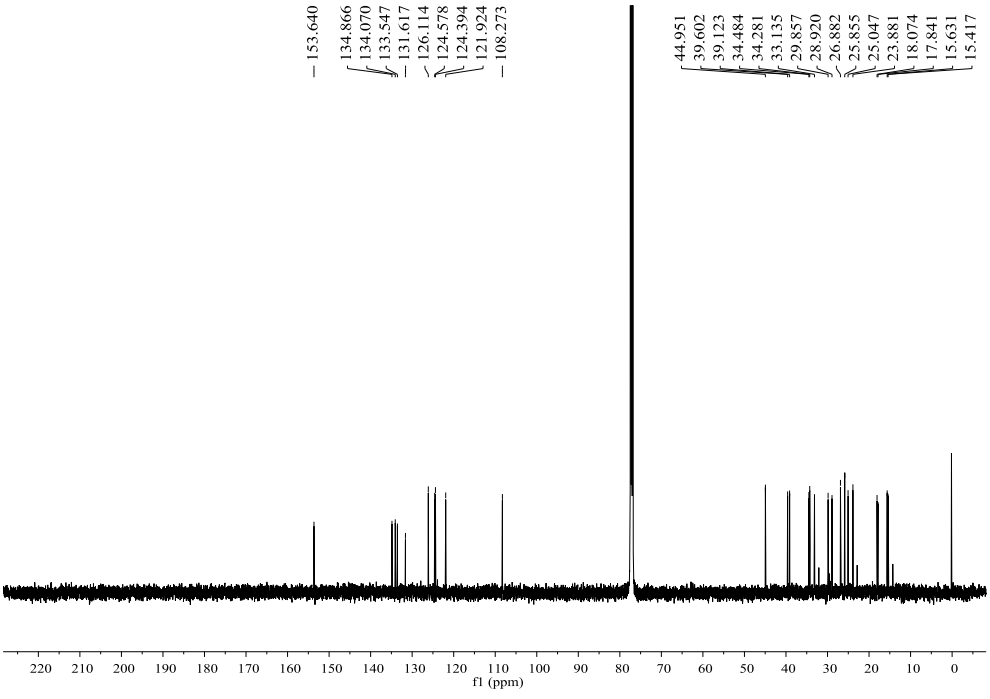



**Figure S4-2. $^{13}$C NMR spectrum of C4 in $CDCl_3$ (125 MHz).**

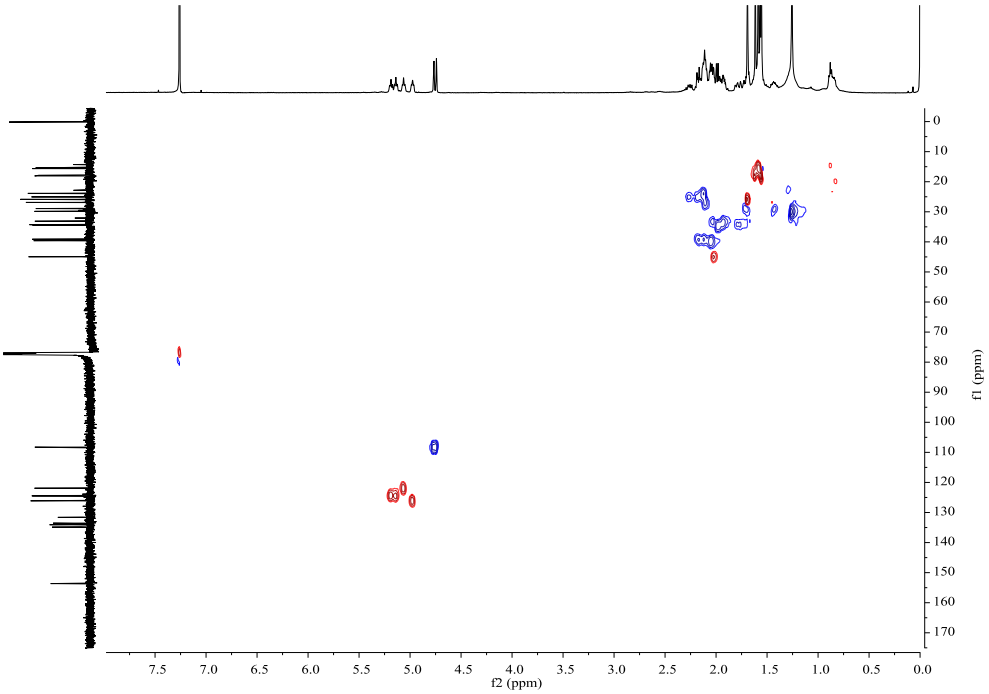



**Figure S4-3. HSQC spectrum of C4 in $CDCl_3$ (500 MHz).**

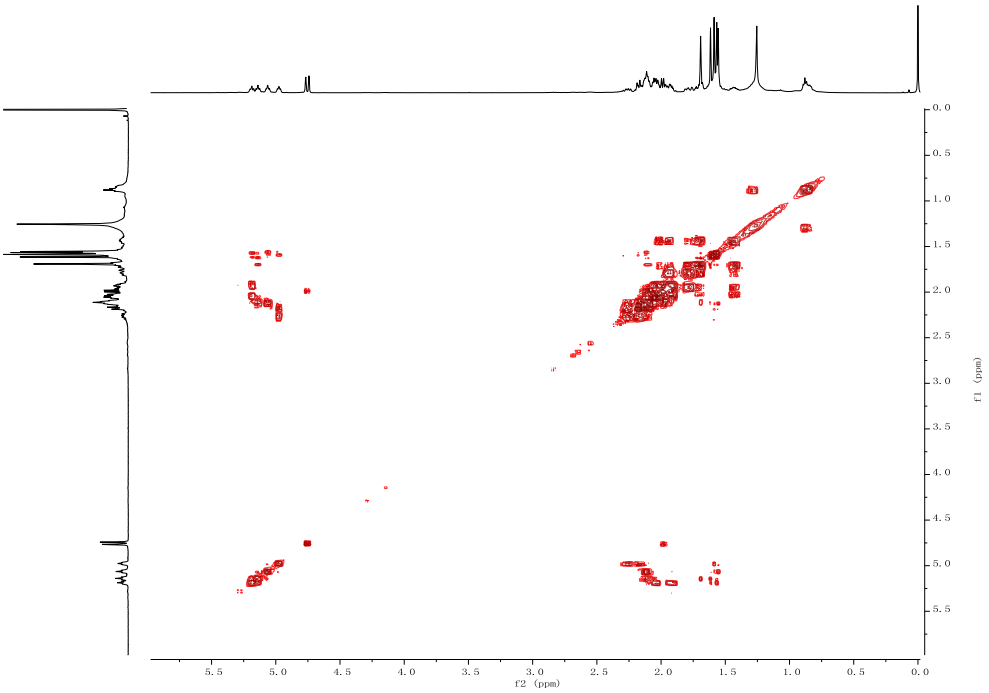



**Figure S4-4. $^{1}H$-$^{1}H$ COSY spectrum of C4 in $CDCl_3$ (500 MHz).**

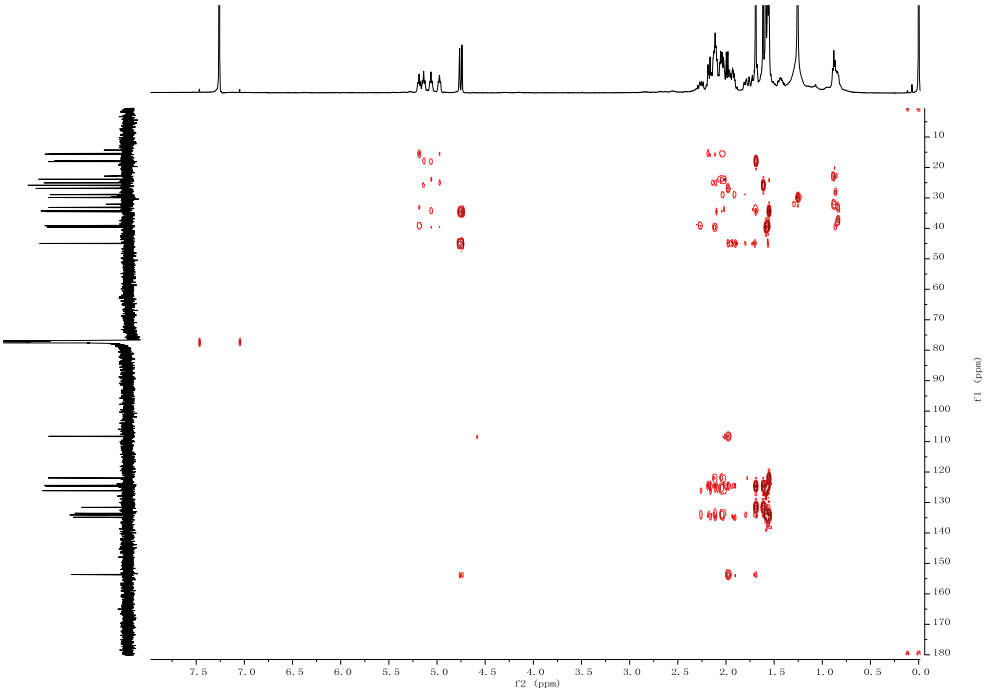
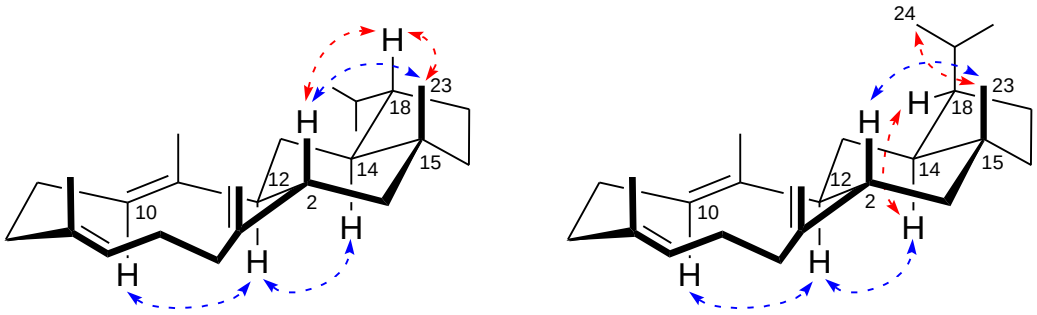



**Figure S4-5. HMBC spectrum of C4 in $CDCl_3$ (500 MHz).**




18-*epi*-thalianatrie (**5**) (+)-thalianatrie (**7**)

**Figure S5. Key NOESY correlations for C5 and C7.**

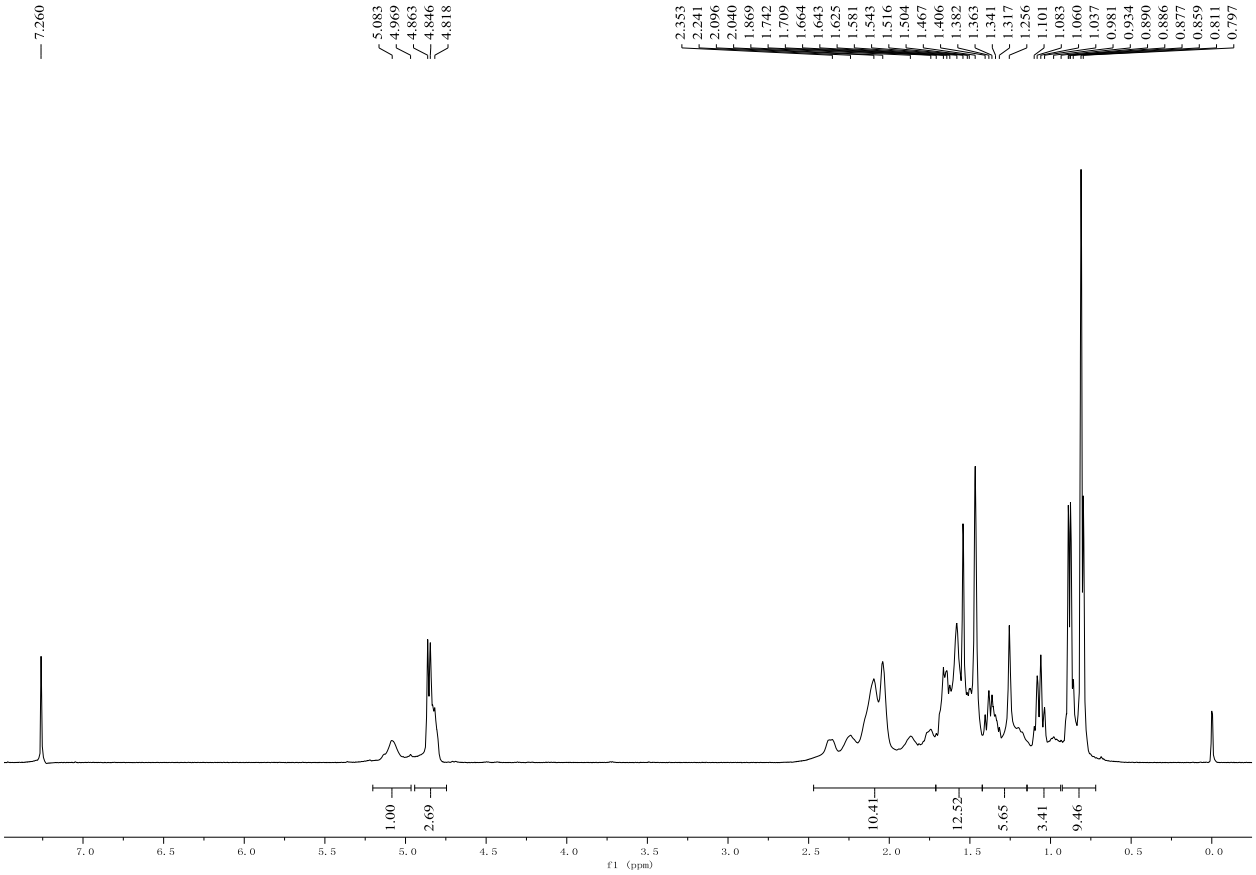



**Figure S5-1. $^{1}$H NMR spectrum of C5 in $CDCl_3$ (500 Hz).**

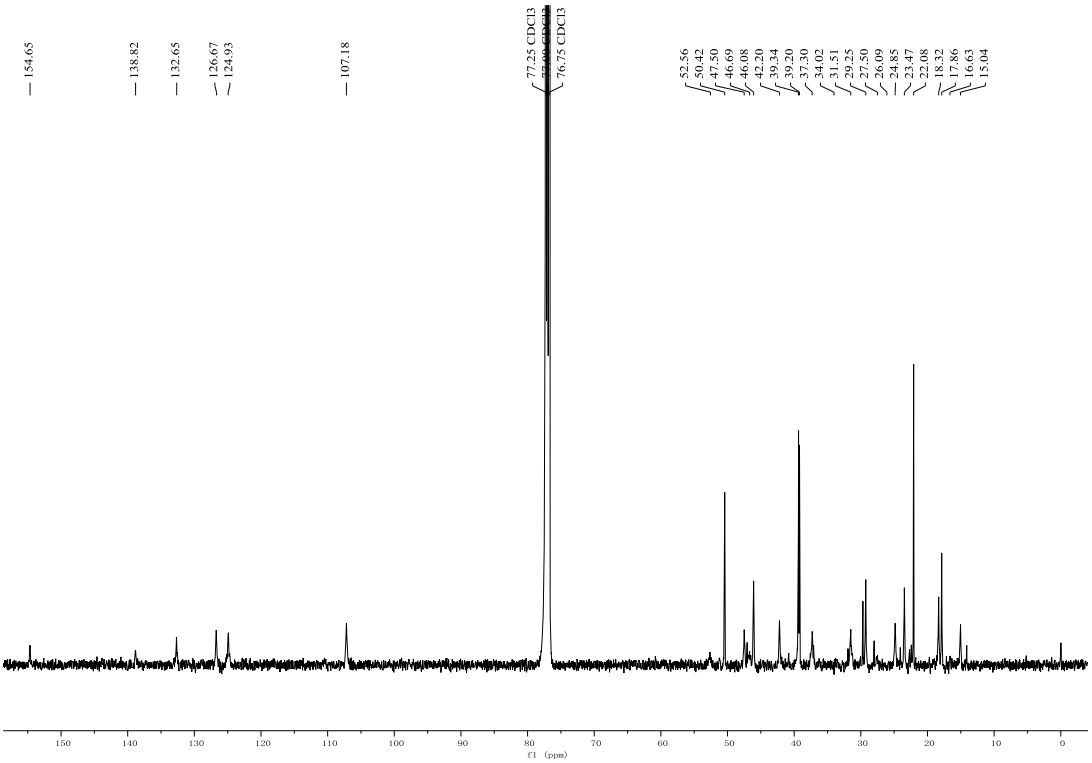



**Figure S5-2. $^{13}$C NMR spectrum of C5 in $CDCl_3$ (125 MHz).**

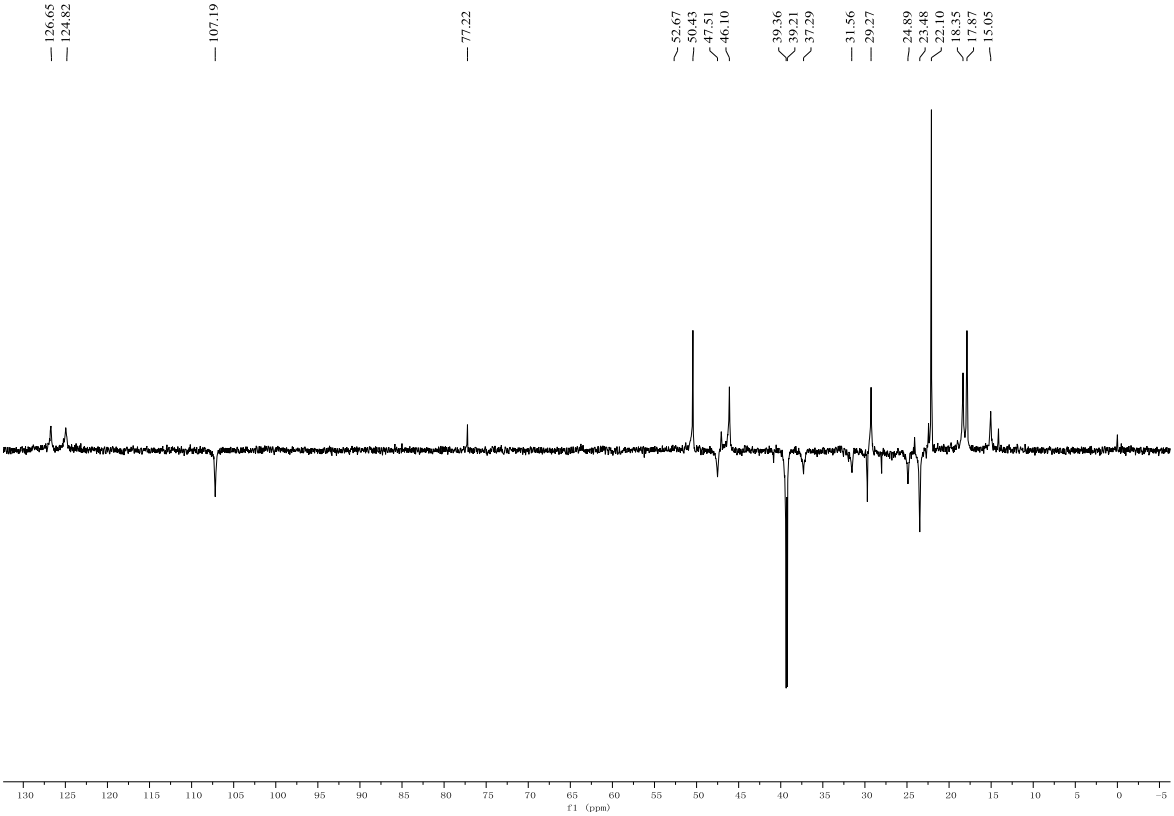



**Figure S5-3. DEPT spectrum of C5 in $CDCl_3$ (125 MHz).**

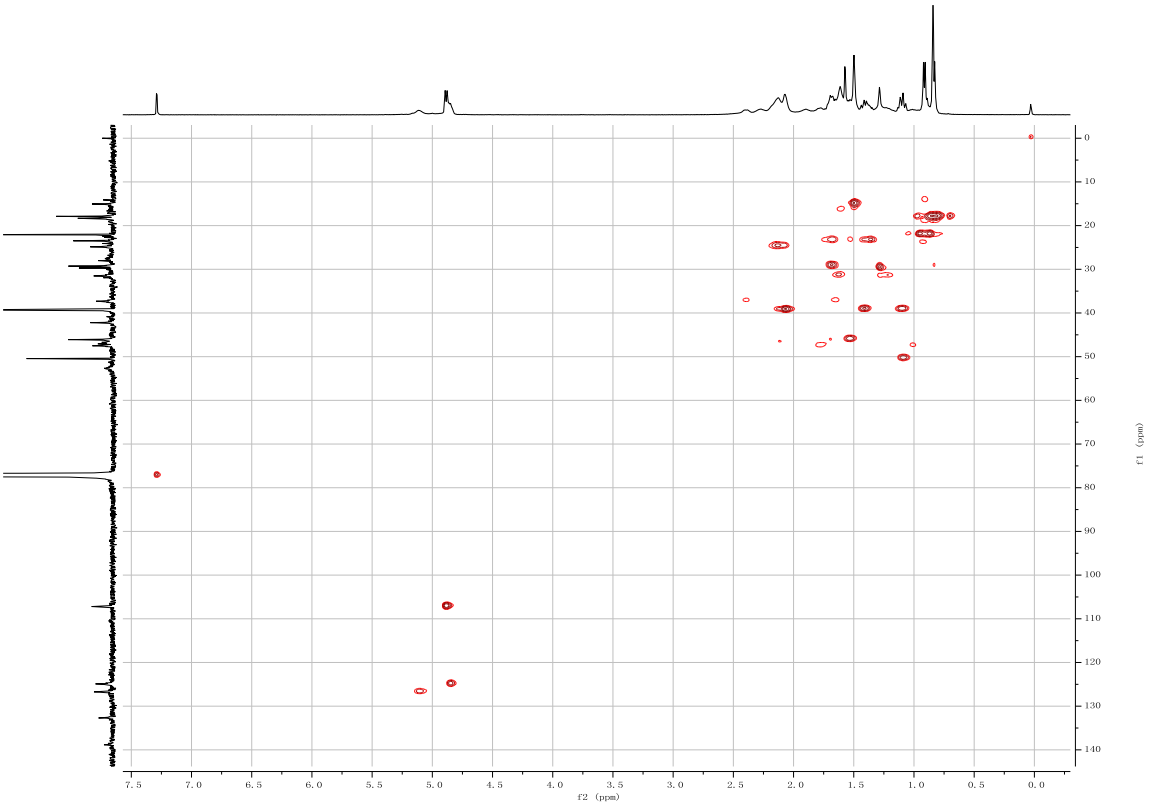



**Figure S5-4. HSQC spectrum of C5 in $CDCl_3$ (500 MHz).**

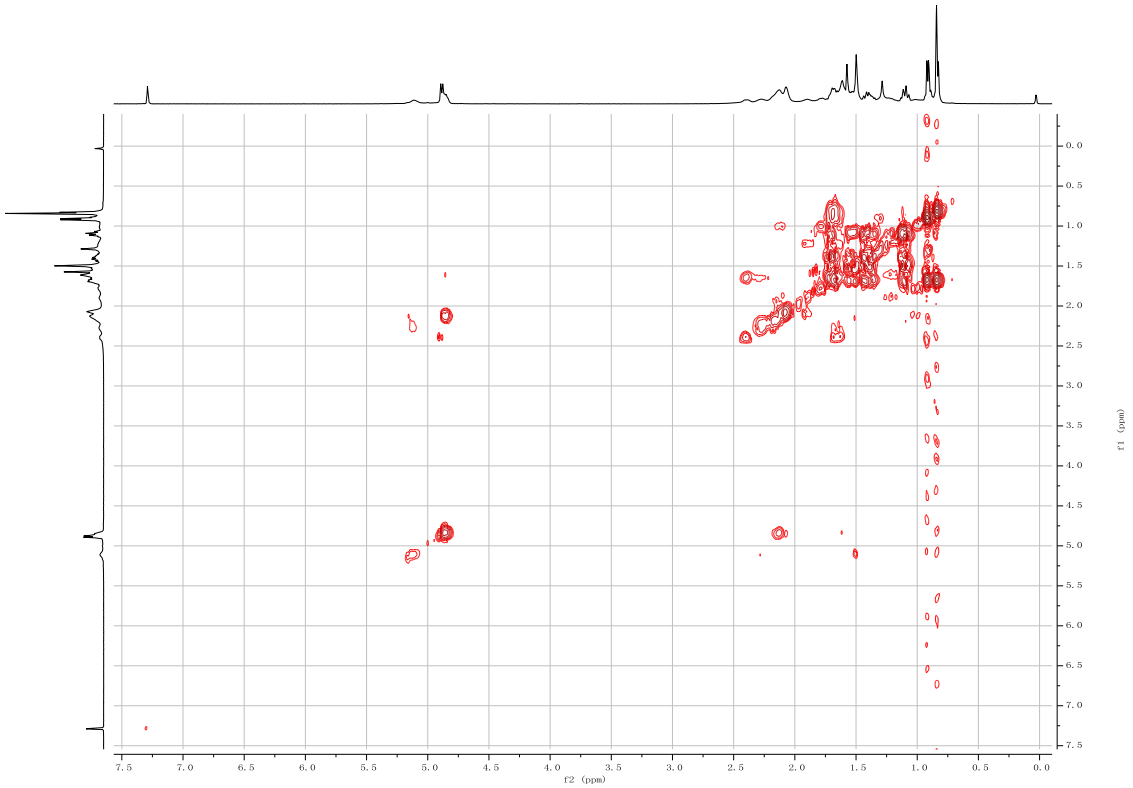



**Figure S5-5. $^{1}H$-$^{1}H$ COSY spectrum of C5 in $CDCl_3$ (500 MHz).**

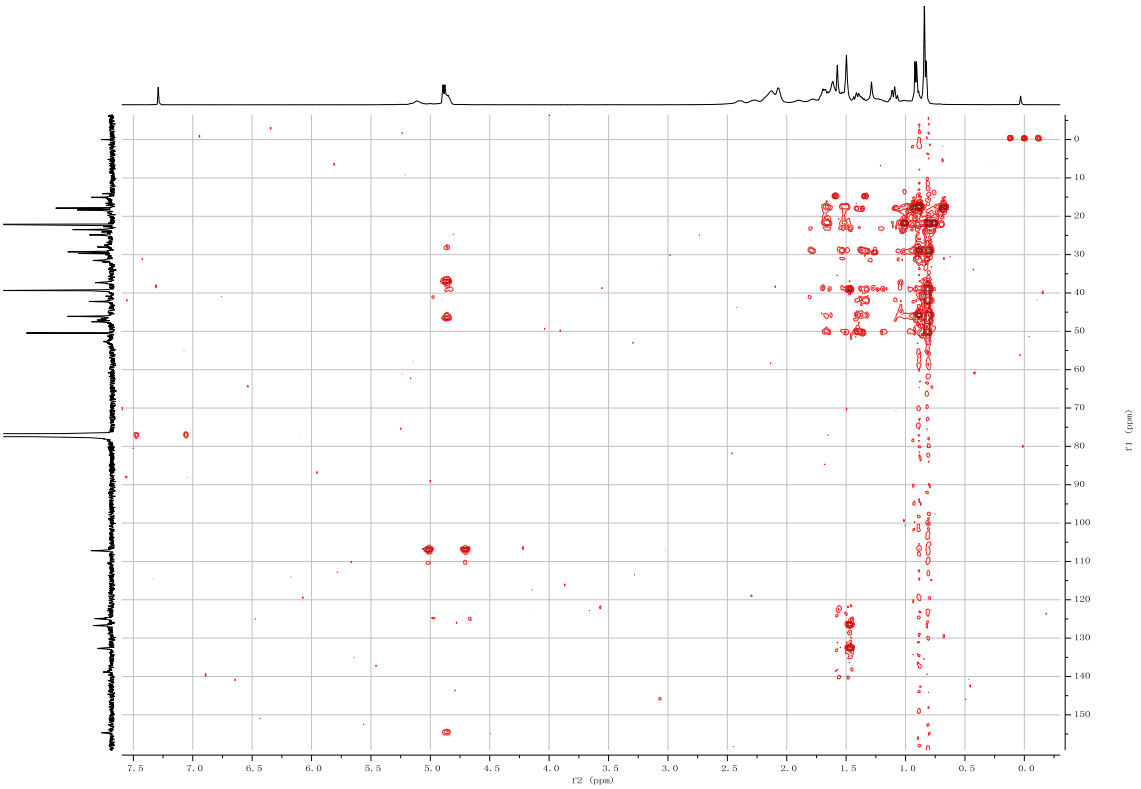



**Figure S5-6. HMBC spectrum of C5 in $CDCl_3$ (500 MHz).**

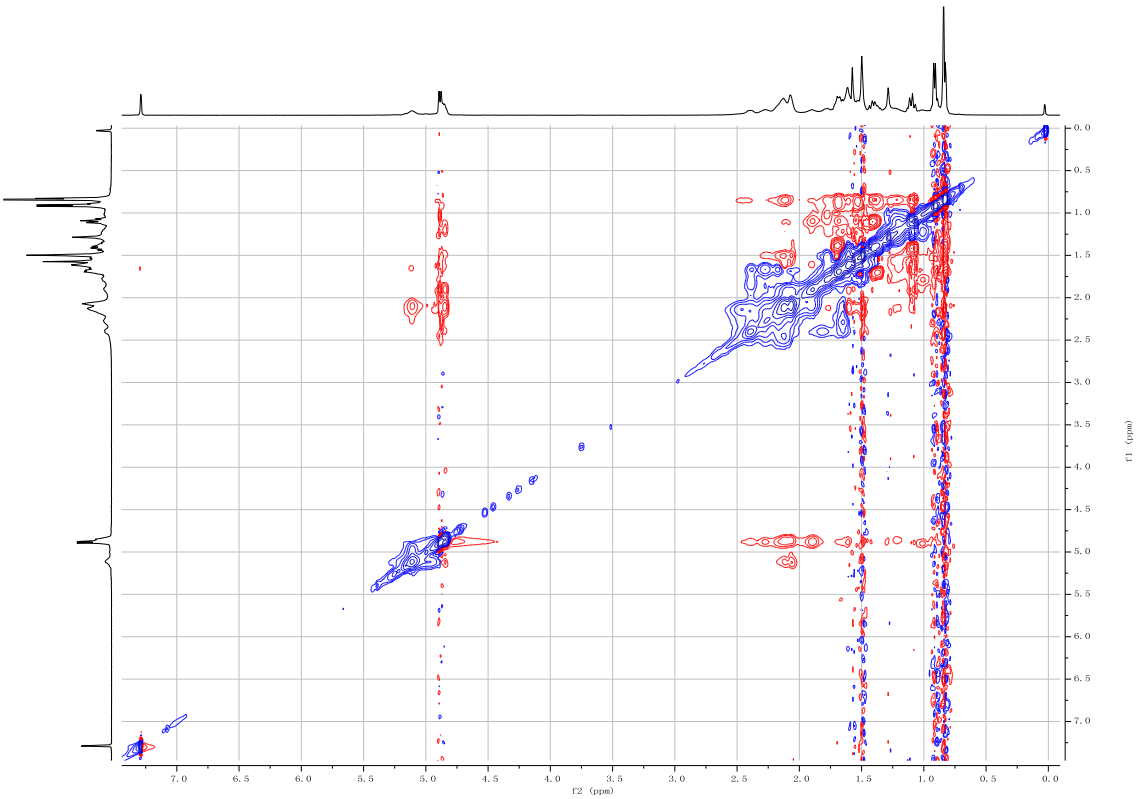



**Figure S5-7. ROESY spectrum of C5 in $CDCl_3$ (500 MHz).**

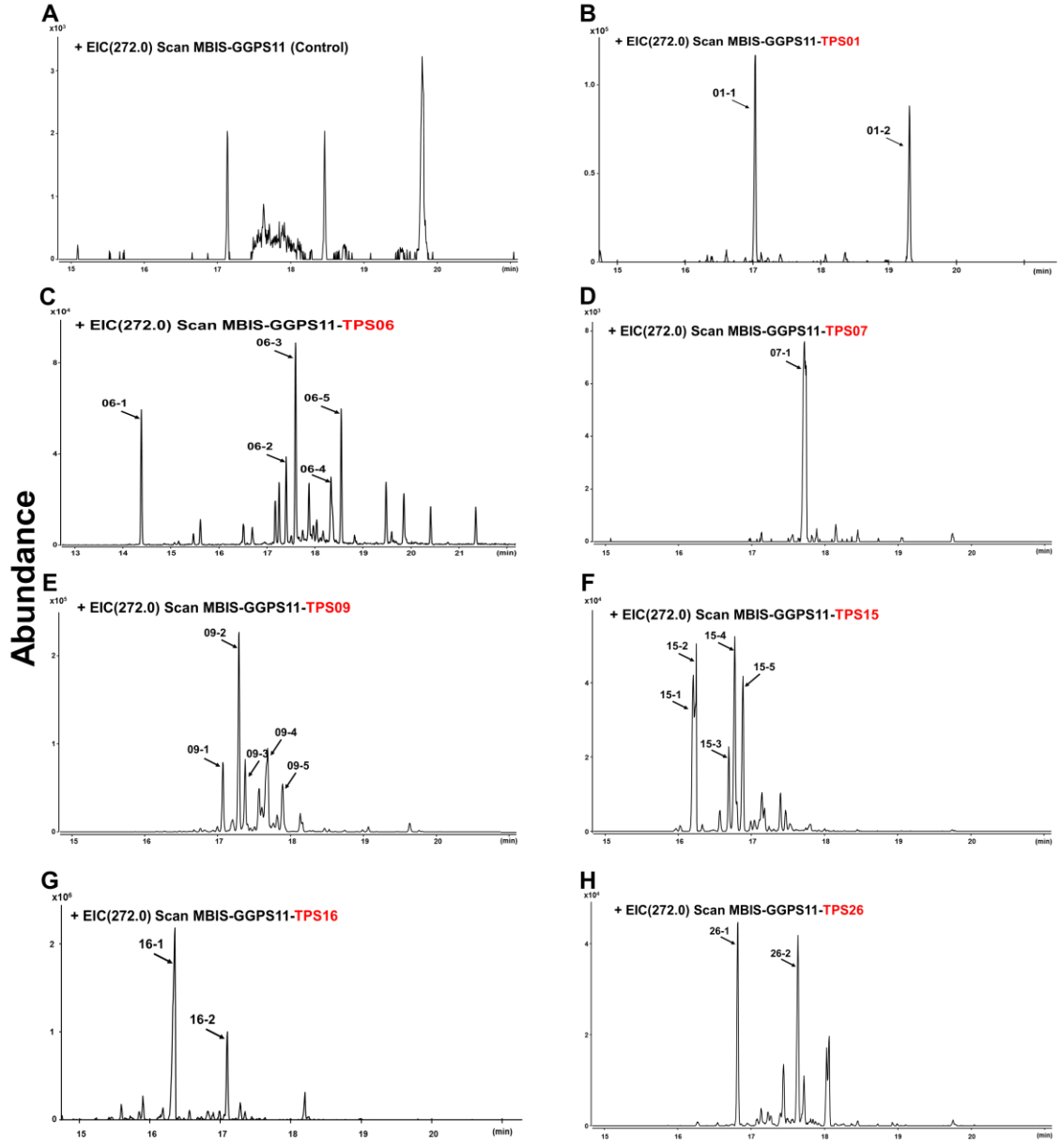



**Figure S6. Di-TPS activity screening of seven Arabidopsis TPS-a members in *E. coli* system. A,** Control, only *AtGGPP11* (*At4g36810*) was expressed in *E. coli*; **B** to **H,** seven TPSs-a members (AtTPS01, AtTPS06, AtTPS07, AtTPS09, AtTPS15, AtTPS16, and AtTPS26) from *Arabidopsis thaliana* (four chemicals, marked with red arrows, were structurally elucidated.

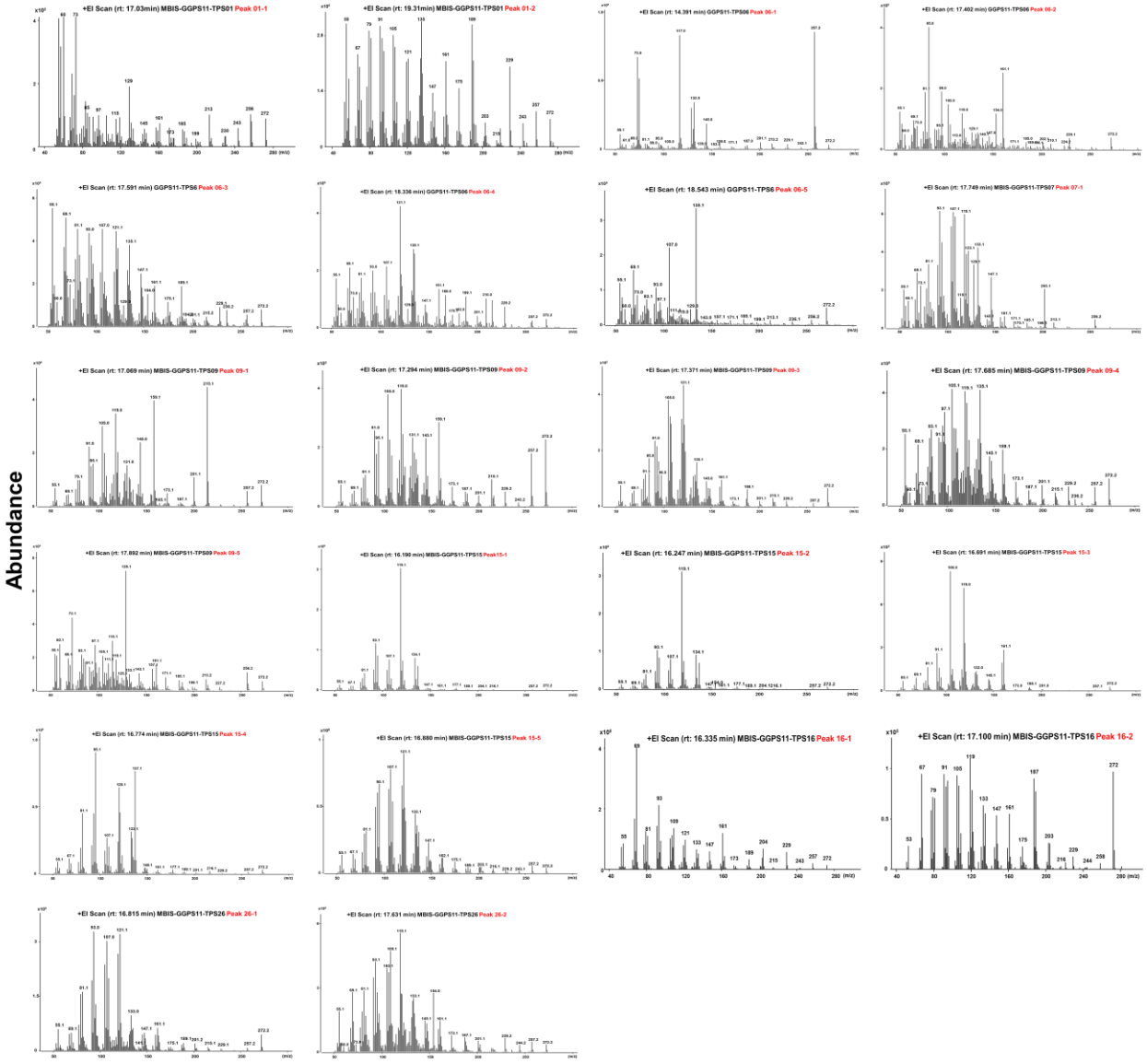



**Figure S7. Mass spectrum of unknown di-terpens generated Arabidopsis TPSs.** AtTPS01 (two main peaks, 01-1 and 01-2), AtTPS06 (five main peaks, 06-1 to 06-5), AtTPS07 (one main peak, 07-1), AtTPS09 (five main peaks, 09-1 to 09-5), AtTPS15 (five main peaks, 15-1 to 15-5), AtTPS16 (two main peaks, 16-1 and 16-2), and AtTPS26 (two main peaks, 26-1 and 26-2).

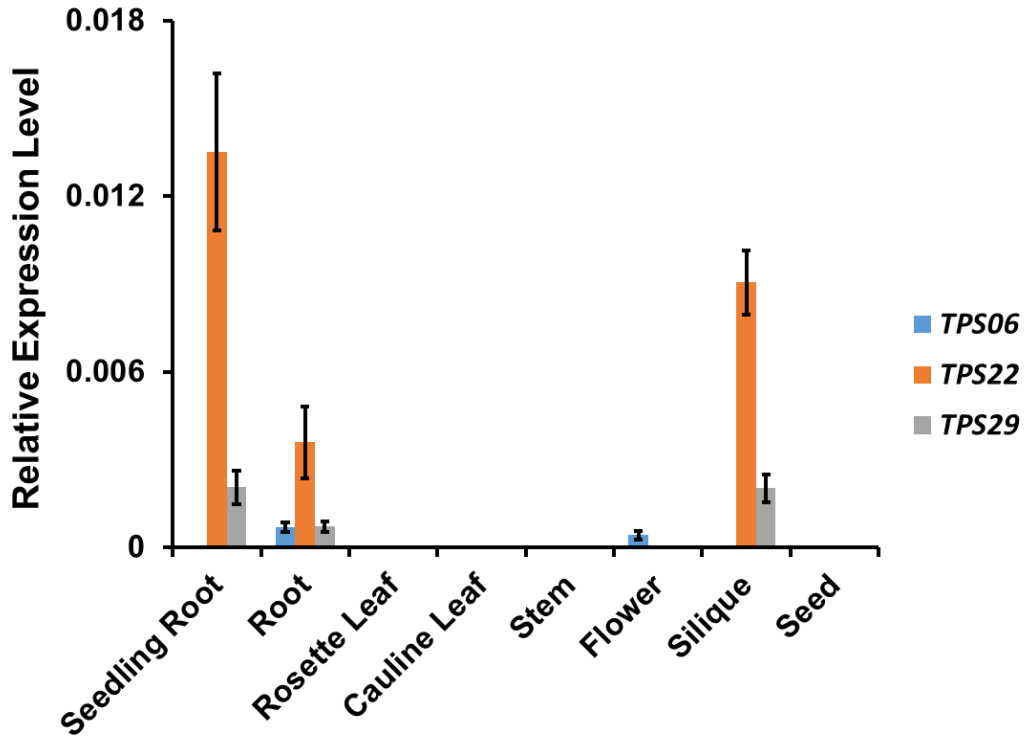



**Figure S8. Tissue-specificity of *TPS06*, *TPS22*, and *TPS29* in *Arabidopsis thaliana*.** Tissue samples were harvested from 8-week-old plants and 2-week-old seedlings. *Actin 2* (*At3g18780*) was used as a reference gene in the analysis. Transcript levels of each gene in the tested tissue are expressed relative to the *Actin 2* gene. Error bars represent means ± SDs ($n$ = 3 biologically independent samples).

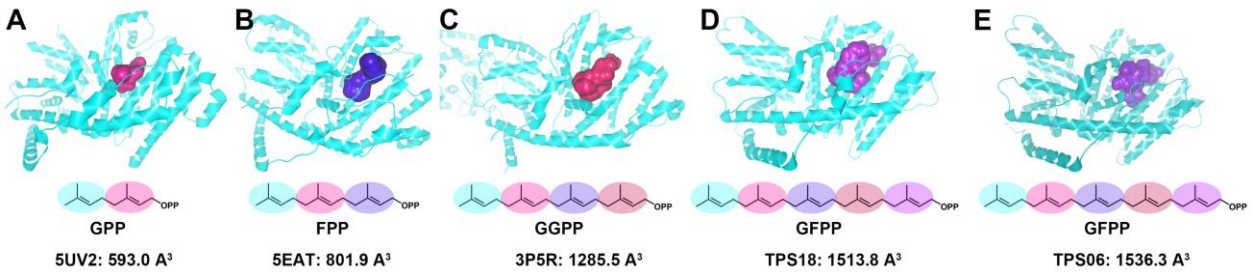



**Figure S9. Catalytic cavity comparison among different types of terpene synthases.** A. mono-TPS ((+)-limonene synthase complexed with 2-fluoroneryl diphosphate, Protein Data Bank entry 5UV2); B, sesqui-TPS (tobacco 5-epi-aristolochene synthase complexed with farnesyl hydroxyphosphonate, Protein Data Bank entry 5EAT); C, di-TPS (taxadiene synthase from *Taxus brevifolia* in complex with $Mg^{2+}$ and 2-fluorogeranylgeranyl diphosphate, Protein Data Bank entry 3P5R); D, AtTPS18; E, AtTPS06.

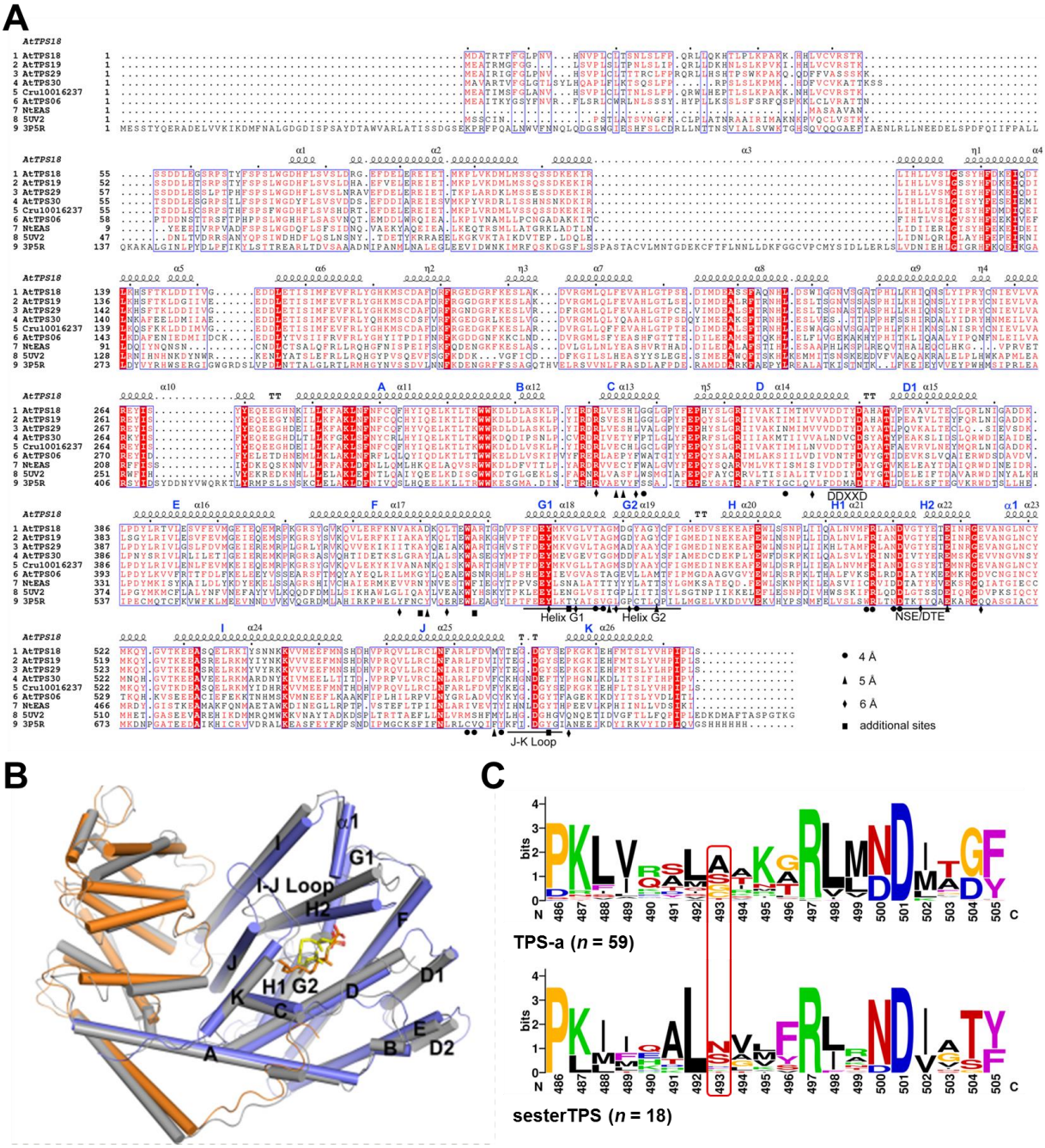



**Figure S10. Amino acid sequence comparison of *Brassicaceae* sesterTPSs and other plant TPSs.**

A. Comparison of amino acid sequences of six sester-TPSs and mono-TPS, sesqui-TPS, and di-TPS. 5UV2 ((+)-limonene synthase complexed with 2-fluoroneryl diphosphate); NtEAS (tobacco 5-epi-aristolochene synthase complexed with farnesyl hydroxyphosphonate, Protein Data Bank entry 5EAT); 3P5R (taxadiene synthase from *Taxus brevifolia* in complex with $Mg^{2+}$ and 2-fluorogeranylgeranyl diphosphate). Helixes A to K (in blue) are labeled on the sequence. The critical amino acid residues are highlighted with different symbols based on the distance between the side chain of amino acid resides and GFPP molecule: black circles, around 4 Å; black

triangles, around 5 Å; black diamonds, around 6 Å. The amino acid residues highlighted with black squares are deduced from published plant TPS structures.

B. Superimposition of the cartoon model of AtTPS18 (orange and blue) and NtEAS (grey, PDB code 5EAT).

C. Helix H1 sequence comparison between active plant sesterTPSs ($n$ = 18) and other TPS-a members ($n$ = 59) from *Brassicaceae* plants, including *Arabidopsis lyrata*, *Arabidopsis thaliana*, *Capsella grandiflora*, *Capsella rubella*, *Boechera stricta*, *Brassica rapa*, *Eutrema salsugineum*, and *Tarenaya hassleriana*. Position 493 was boxed in red. All TPS sequences used in this analysis are extracted from Chen et al (2019).

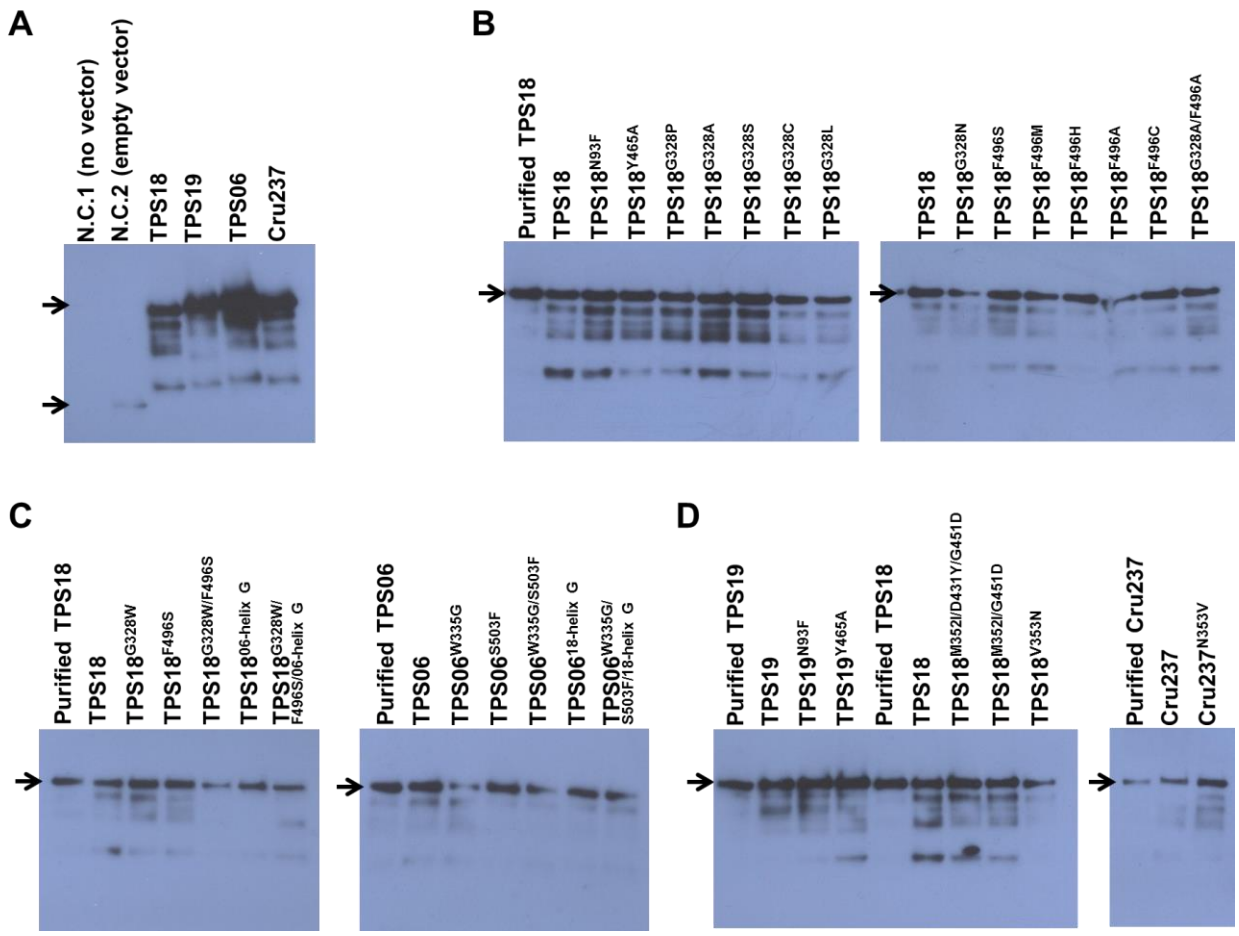



**Figure S11. Measurement of soluble C-terminal MBP-tagged sesterTPSs in *E.coli* with anti-MBP monoclonal antibody.** The calculated molecular weight of each MBP-tagged sesterTPSs is listed below: MBP-tag itself (43.0 KDa), AtTPS06 (107.5 KDa), AtTPS18 (106.6 KDa), AtTPS19 (107.0 KDa), and Cr237 (107.5 KDa). The positions of targeted proteins are marked with arrows.

A. Western blot analysis of soluble recombinant AtTPS06, AtTPS18, AtTPS19, and Cru237. N.C., negative control (one with no pMAL-c2x vector, one with pMAL-c2x empty vector).

B. Western blot analysis of soluble recombinant AtTPS18 and its mutants.

C. Western blot analysis of soluble recombinant AtTPS18, AtTPS06 and their interconverted mutants.

D. Western blot analysis of soluble recombinant AtTPS18, AtTPS19, Cru237 and their interconverted mutants.

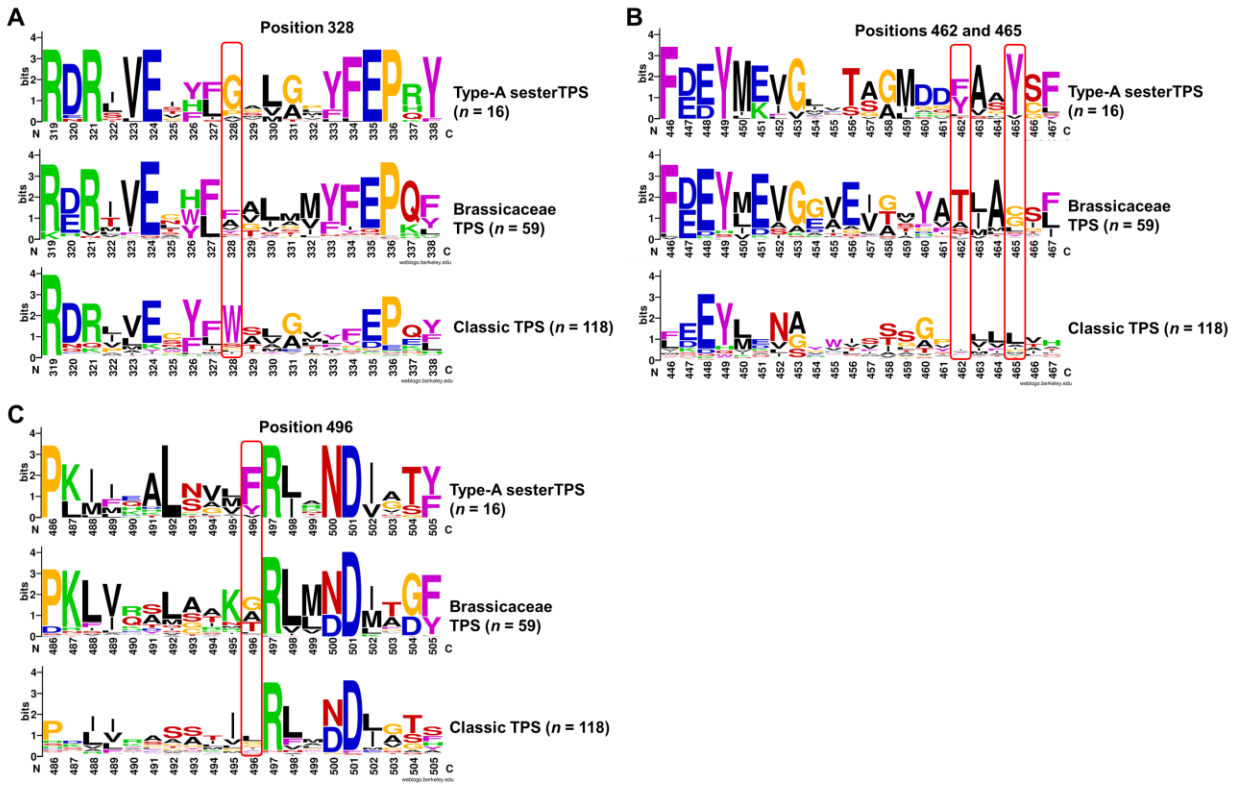



**Figure S12. Sequence comparison between active plant type-A sesterTPSs (*n* = 16) and other TPS-a members (*n* = 59) from *Brassicaceae* plants (see the legend for Figure S11B) and classic TPSs (belong to TPS-a, -b, -g subfamilies; *n* = 118) from other angiosperms plants, including *Aquilegia coerulea* Goldsmith, *Cucumis sativus*, *Glycine max*, *Oryza sativa*, *Solanum lycopersicum*.** All TPS protein sequences from other angiosperms plants are extracted form Phytozome 12.0 (https://phytozome.jgi.doe.gov/pz/portal.html#).

A. the region of position 328 (total 20 residues, from 319-338).

B. the region of positions 462/465 (total 20 residues, from 446-467).

C. the region of position 496 (total 20 residues, from 486-505).

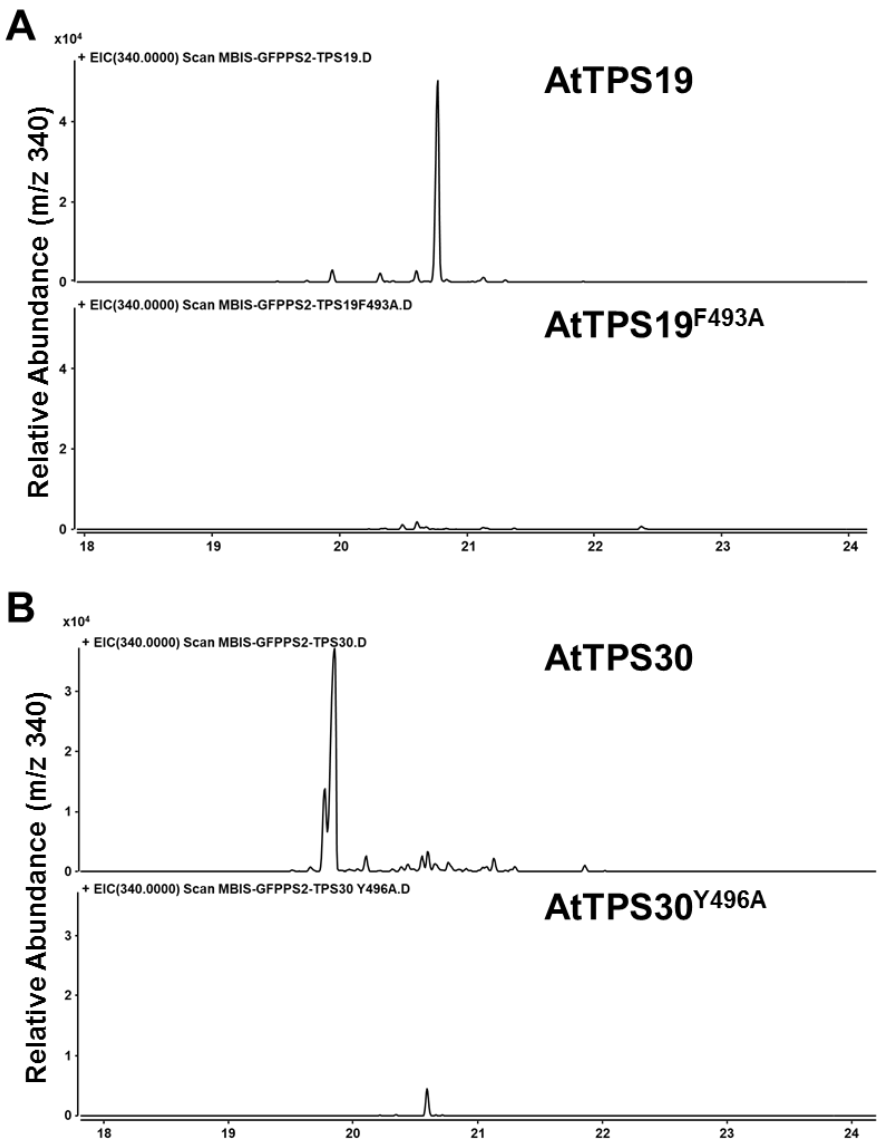



**Figure S13. Sester-TPS activity screening of AtTPS19, AtTPS30 and their mutants in *E. coli* system.**

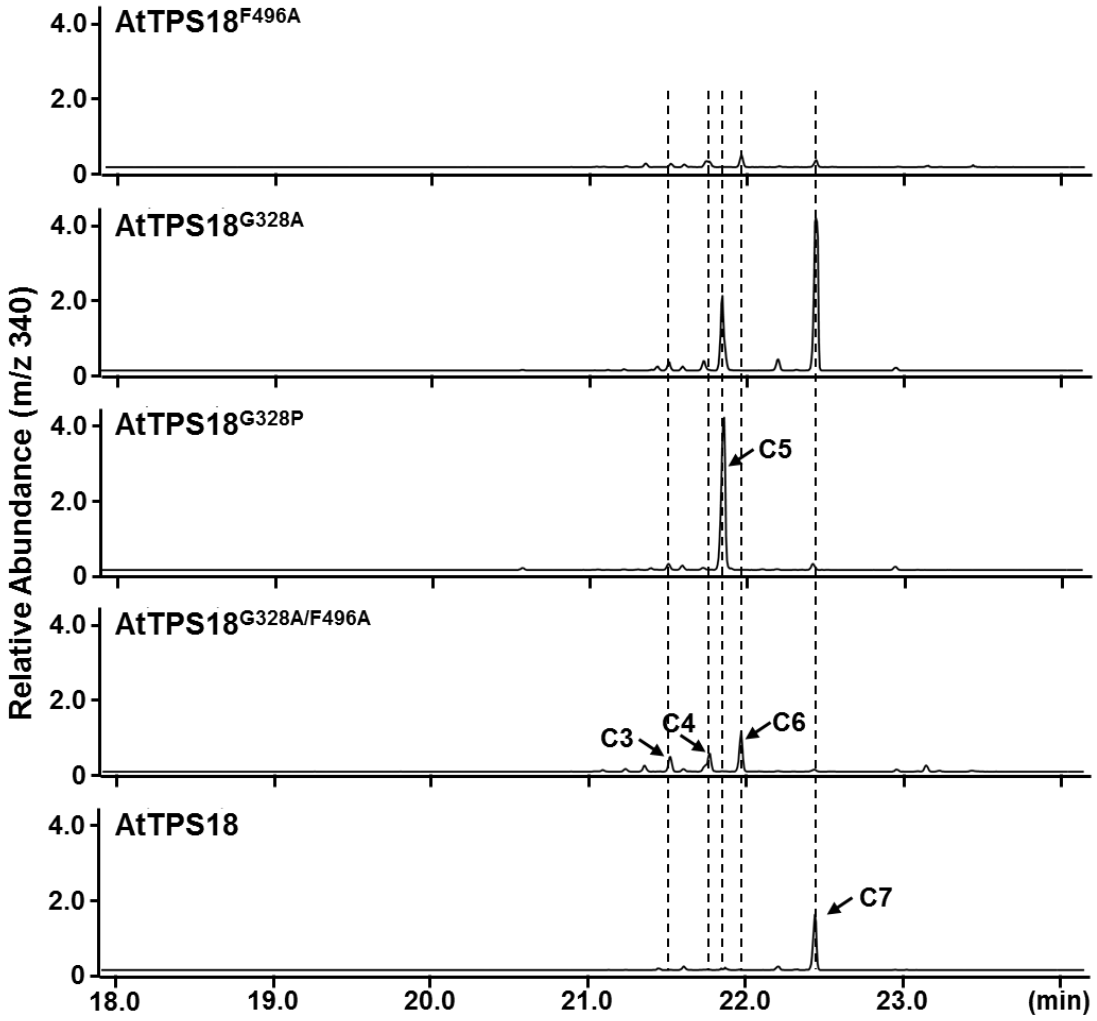



**Figure S14. Sester-TPS activity screening of AtTPS18 and its mutants in *E. coli* system.** The novel sesterterpenes (**C4** and **C5**) were further purified and structurally elucidated in this study.


**Figure 15. Sequence alignment of AtTPS18, AtTPS19, and Cru237 from *Capsella rubella.*** The different amino acids revealed by structural comparison were highlighted with red stars.

## Supplemental Tables

**Table S1.** $^{1}H$ and $^{13}C$ NMR data[a] of Compounds **1** and **2**.

| No. | **Compound 1** | | **Compound 2** | |
|---|---|---|---|---|
| | $\delta_H$ (mult., *J* in Hz) | $\delta_C$ (ppm) | $\delta_H$ (mult., *J* in Hz) | $\delta_C$ (ppm) |
| 1 | 1.67, dd (16.6, 1.9) | 38.7 | 1.72, o | 38.4 |
| | 2.11, dd (8.2, 16.6) | | 2.07, d (16.4) | |
| 2 | 4.89, d (8.2) | 124.6 | 2.57, m | 132.5 |
| 3 | - | 132.5 | - | 34.2 |
| 4 | 1.93, o | 39.2 | 1.17, o | 32.6 |
| | 2.19, d (11.9) | | 1.30, o | |
| 5 | 2.02, o | 25.8 | 1.24, o | 31.2 |
| | 2.26, dddd (12.9, 12.9,12.9, 4.5) | | 1.43, m | |
| 6 | 4.79, dd (12.9, 4.5) | 124.0 | 1.77, d (10.2) | 50.6 |
| 7 | - | 137.8 | 1.66, o | 41.5 |
| 8 | 1.70, dd (12.9, 12.9) | 41.5 | 1.26, o | 31.0 |
| | 1.83, dd (12.9, 8.3) | | 1.89, m | |
| 9 | 0.95, m | 23.8 | 2.21, m | 27.2 |
| | 1.10, m | | 2.28, o | |
| 10 | 1.53 | 39.4 | - | 138.8 |
| 11 | - | 38.8 | - | 128.7 |
| 12 | 1.00, ddd (14.1, 4.3, 2.5) | 34.3 | - | 133.3 |
| | 1.91, o | | | |
| 13 | 1.43, o | 35.9 | 1.96, dd (15.7, 4.5) | 29.3 |
| | 1.51, o | | 2.32, o | |
| 14 | 1.55, o | 49.4 | 1.67, o | 46.4 |
| 15 | - | 43.6 | | 40.0 |
| 16 | 1.21, ddd (10.9, 10.9, 10.9) | 40.1 | 1.16, o | 40.8 |
| | 1.43, o | | 1.57, o | |
| 17 | 1.40, o | 30.4 | 1.53, o | 28.8 |
| | 2.00, o | | 1.80, m | |
| 18 | 2.66, ddd (10.8, 10.8, 5.3) | 47.2 | 1.71, o | 46.6 |
| 19 | - | 148.7 | 1.62, o | 30.9 |
| 20 | 1.49, s | 16.5 | 0.88, d (7.3) | 21.9 |
| 21 | 1.44, s | 16.7 | 0.82, d (7.1) | 20.0 |
| 22 | 0.82, s | 21.0 | 1.64, s | 17.5 |
| 23 | 0.87, s | 17.9 | 0.77, s | 18.6 |
| 24 | 4.60, dd (2.3, 1.3) | 109.9 | 0.95, d (6.3) | 24.2 |
| | 4.68, d (2.2) | | | |
| 25 | 1.72, s | 18.9 | 0.86, d (6.5) | 22.1 |

[a] Bruker Avance III spectrometer (900 MHz for $^{1}H$ and 200 MHz for $^{13}C$ NMR) in $CDCl_3$, chemical shifts (ppm) referred to $CHCl_3$ ($\delta_H$ 7.26) and to $CDCl_3$ ($\delta_C$ 77.0); assignments were deduced from analysis of 1D and 2D spectra.

**Table S2.** $^{1}H$ and $^{13}C$ NMR data[a] of Compounds **4** and **5**.

| No. | **Compound 4** | | **Compound 5** | |
|---|---|---|---|---|
| | $\delta_H$ (mult., *J* in Hz) | $\delta_C$ (ppm) | $\delta_H$ (mult., *J* in Hz) | $\delta_C$ (ppm) |
| 1 | 1.93, m<br>2.03, m | 33.1 | 0.98, m<br>1.75, m | 47.5 |
| 2 | 5.19, t (7.3) | 124.4 | 2.09, m | 46.7 |
| 3 | - | 134.9 | - | 154.7 |
| 4 | 2.12, o<br>2.18, o | 39.1 | 1.62, m<br>2.36, m | 37.3 |
| 5 | 2.18, o<br>2.26, m | 25.0 | 2.14, m<br>2.23, m | 27.5 |
| 6 | 4.98, t (6.0) | 126.1 | 5.07, brs | 126.7 |
| 7 | - | 133.5 | - | 132.6 |
| 8 | 2.06, o | 39.6 | 2.04, m | 39.3 |
| 9 | 2.12, o | 23.9 | 2.10, m | 24.9 |
| 10 | 5.06, t (6.1) | 121.9 | 4.82, brs | 124.9 |
| 11 | - | 134.1 | - | 138.8 |
| 12 | 1.79, o<br>1,95, m | 34.3 | 1.86, m | 52.6 |
| 13 | 1.43, m<br>1.72, o | 28.9 | 1.22, m<br>1.60, m | 31.5 |
| 14 | 2.03, o | 45.0 | 1.06, m | 50.4 |
| 15 | - | 153.6 | - | 42.2 |
| 16 | 1.99, m | 34.5 | 1.07, m<br>1.38, m | 39.2 |
| 17 | 2.10, o | 26.9 | 1.33, m<br>1.65, m | 23.5 |
| 18 | 5.14, ddt (7.0, 5.5, 1.4) | 124.6 | 1.50, m | 46.1 |
| 19 | - | 131.6 | 1.65, m | 29.3 |
| 20 | 1.57, s | 15.6 | 4.85, brs | 107.2 |
| 21 | 1.59, s | 15.4 | 1.47, s | 15.0 |
| 22 | 1.55, s | 18.1 | 1.58, s | 16.6 |
| 23 | 4.74, s<br>4.77, s | 108.3 | 0.82, s | 17.9 |
| 24 | 1.61, s | 17.8 | 0.88, d (6.5) | 22.1 |
| 25 | 1.69 | 25.9 | 0.80, d (6.5) | 18.3 |

[a] Bruker-DRX-500 spectrometer (500 MHz for $^{1}H$ and 125 MHz for $^{13}C$ NMR) in $CDCl_3$, chemical shifts (ppm) referred to $CHCl_3$ ($\delta_H$ 7.26) and to $CDCl_3$ ($\delta_C$ 77.0); assignments were deduced from analysis of 1D and 2D spectra.

**Table S3. Data collection and structural refinement of AtTPS18-PP and AtTPS18-FSPP.**

| Statistic | AtTPS18-PP* | AtTPS18-FSPP |
|---|---|---|
| Wavelength, Å | 0.9798 | 0.9798 |
| Resolution, Å | 39.36 - 2.148 (2.225 - 2.148) | 34.44 - 2.303 (2.385 - 2.303) |
| Space group | P $2_1$ 2 $2_1$ | P $2_1$ 2 $2_1$ |
| Until cell | | |
| *a*, *b*, *c*, Å | 64.562 80.925 125.776 | 64.388 81.537 125.57 |
| α, β, γ, ° | 90 90 90 | 90 90 90 |
| Unique reflections | 34,507 (2,221)† | 28,566 (2,080) |
| Multiplicity | 10.5 | 10.8 |
| I/σ(I) | 26.1 (7.9) | 20.0 (0.4) |
| $R_{merge}$(%) | 0.083 (0.651 ) | 0.115 (0.887) |
| Completeness, % | 99.2 (93.4) | 99.7 (95.1) |
| $R_{work}/R_{free}$‡ | 0.1769/0.2354 | 0.1813/0.2323 |
| Number of atoms | 4,575 | 4,441 |
| Ligand | 11 | 26 |
| Water | 239 | 89 |
| Protein residues | 537 | 537 |
| r.m.s. deviations | | |
| bonds (Å) | 0.009 | 0.008 |
| Angles ( ) | 1.2 | 1.03 |
| Average B factor | 39.8 | 42.28 |
| Protein | 39.75 | 42.14 |
| Ligand | 41.98 | 83.92 |
| Water | 40.7 | 37.04 |
| Ramachandran Favored | | |
| (%) | 94 | 95 |
| Allowed (%) | 3.8 | 3.2 |

* Friedel mates were averaged when calculating reflection statistics.

† Numbers in parentheses represent the highest-resolution shell.

‡ $R = \Sigma hkl||Fo|-|Fc||/\Sigma hkl|Fo|$.