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

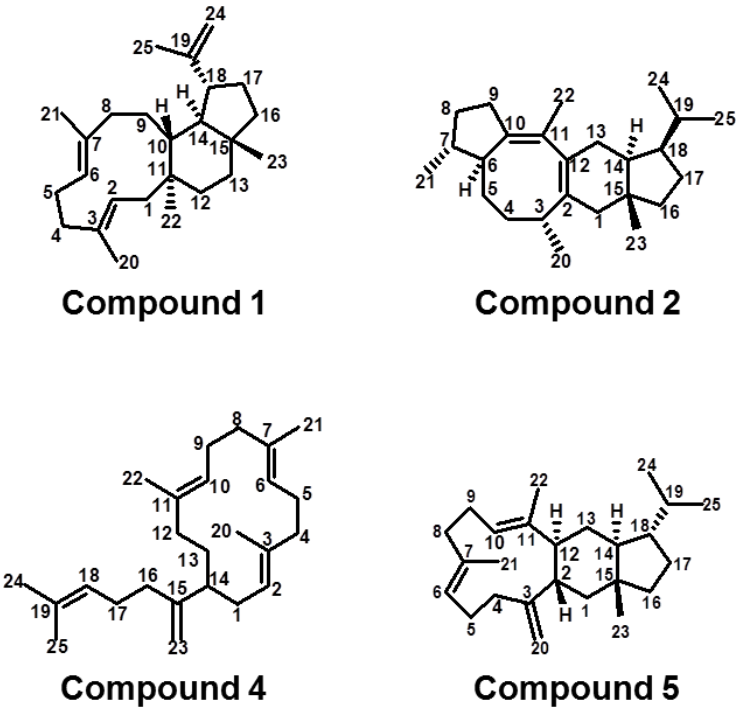


Figure S1. Four novel sesterterpenes (C1, C2, C4, and C5) elucidated in this study (detailed NMR data see Figures S2-S5, Tables S1-S2).

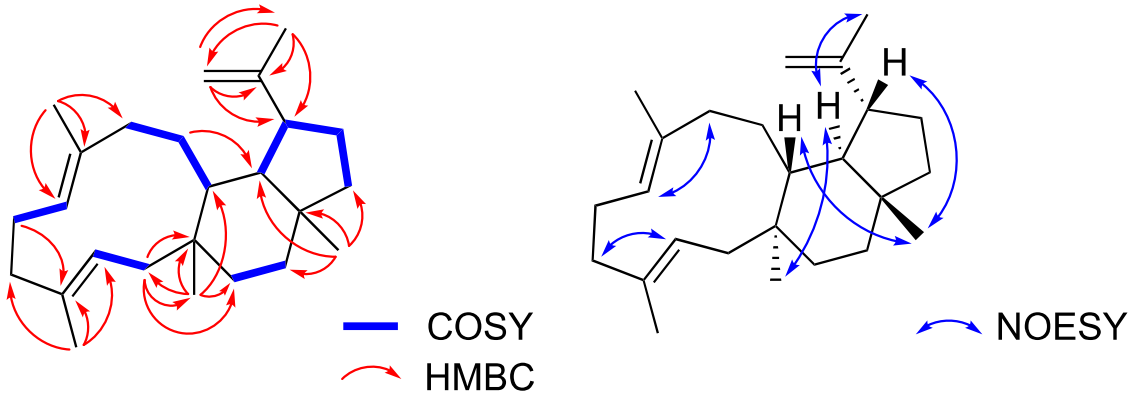


Figure S2. ^{1}H-^{1}H COSY, key HMBC, and Key NOESY correlations in C1.

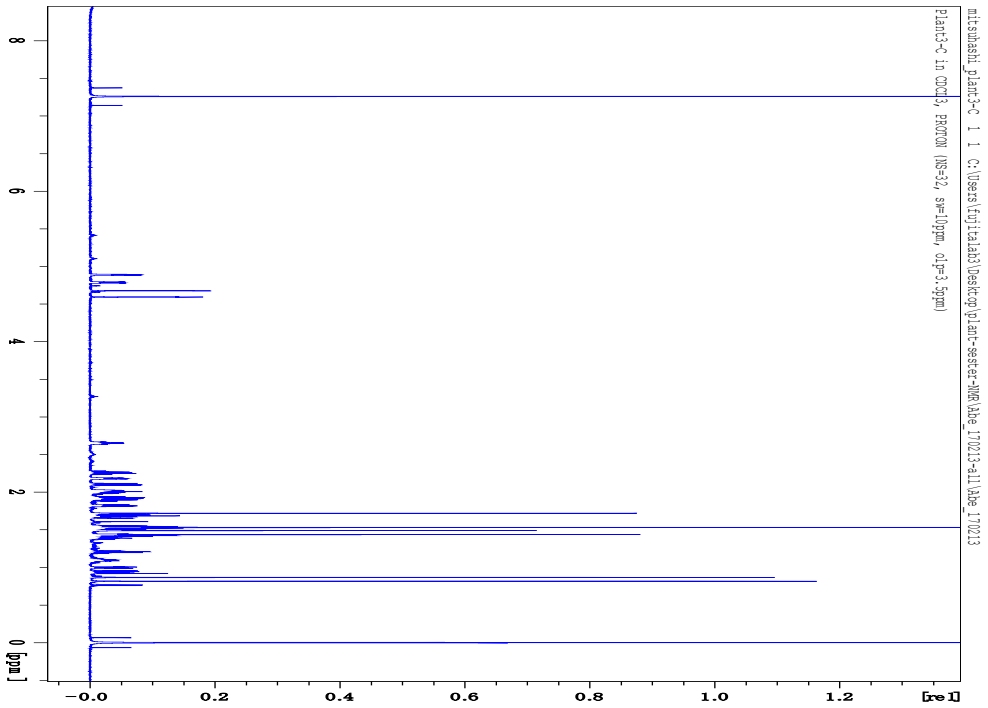


Figure S2-1. ^{1}H NMR spectrum of C1 in $CDCl_3$ (900 MHz).

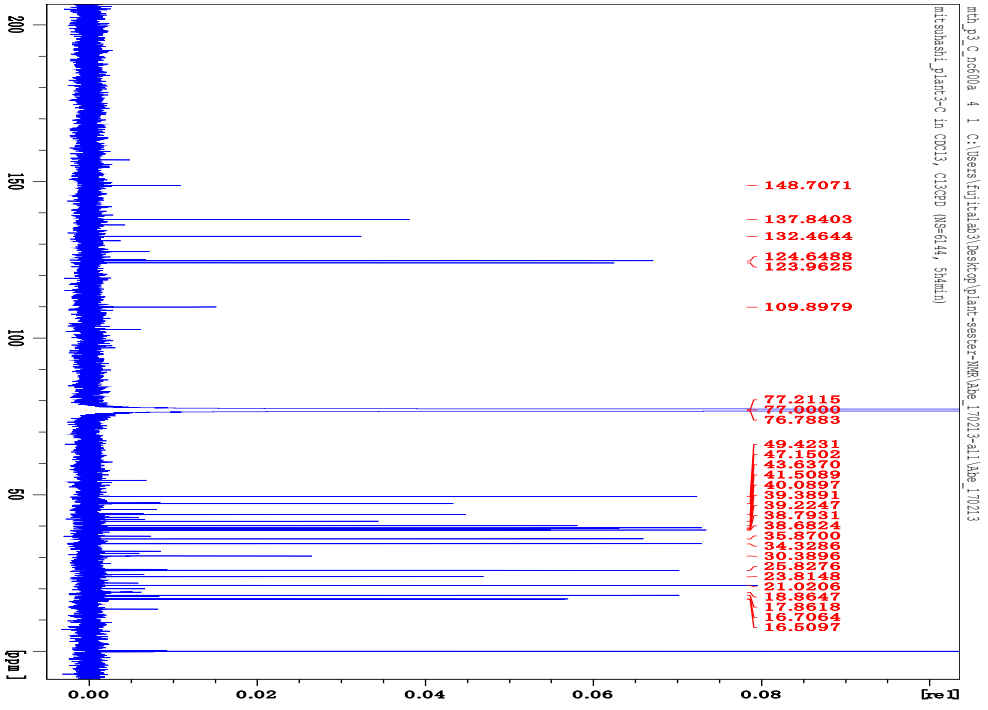


Figure S2-2. ^{13}C NMR spectrum of C1 in $CDCl_3$ (200 MHz).

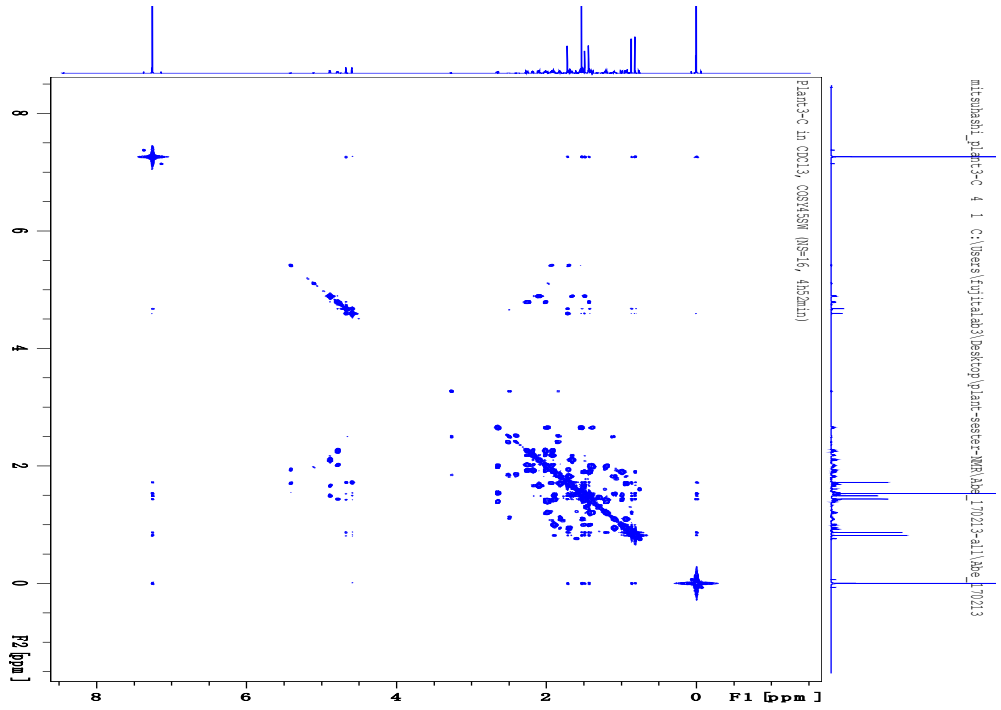


Figure S2-3. ^{1}H-^{1}H COSY spectrum of C1 in $CDCl_3$ (900 MHz).

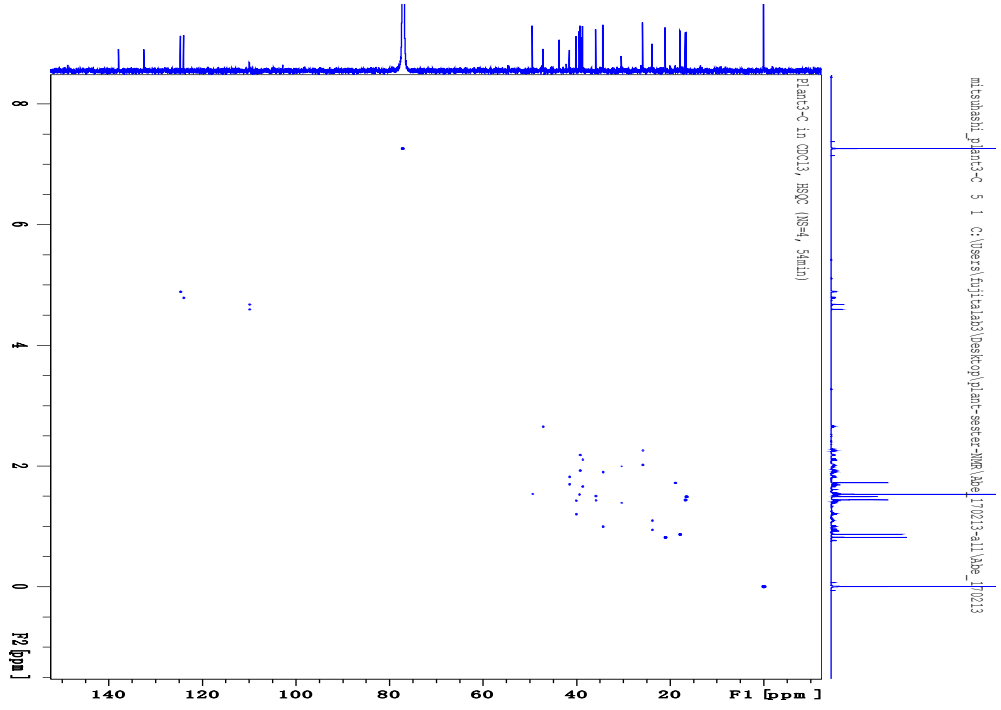


Figure S2-4. HSQC spectrum of C1 in $CDCl_3$ (900 MHz).

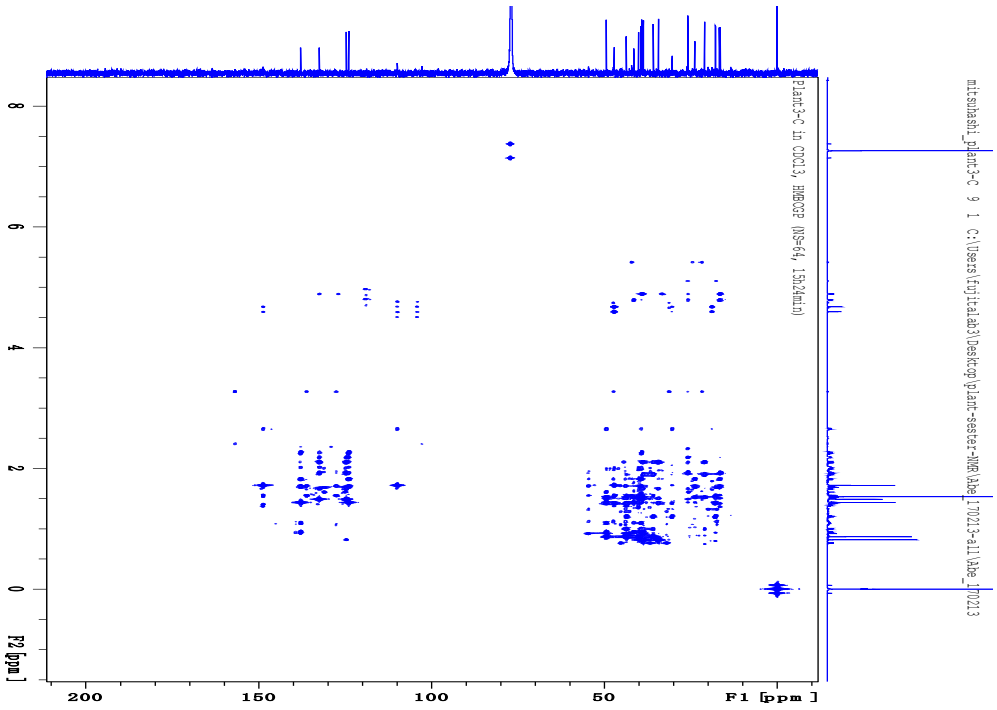


Figure S2-5. HMBC spectrum of C1 in $CDCl_3$ (900 MHz).

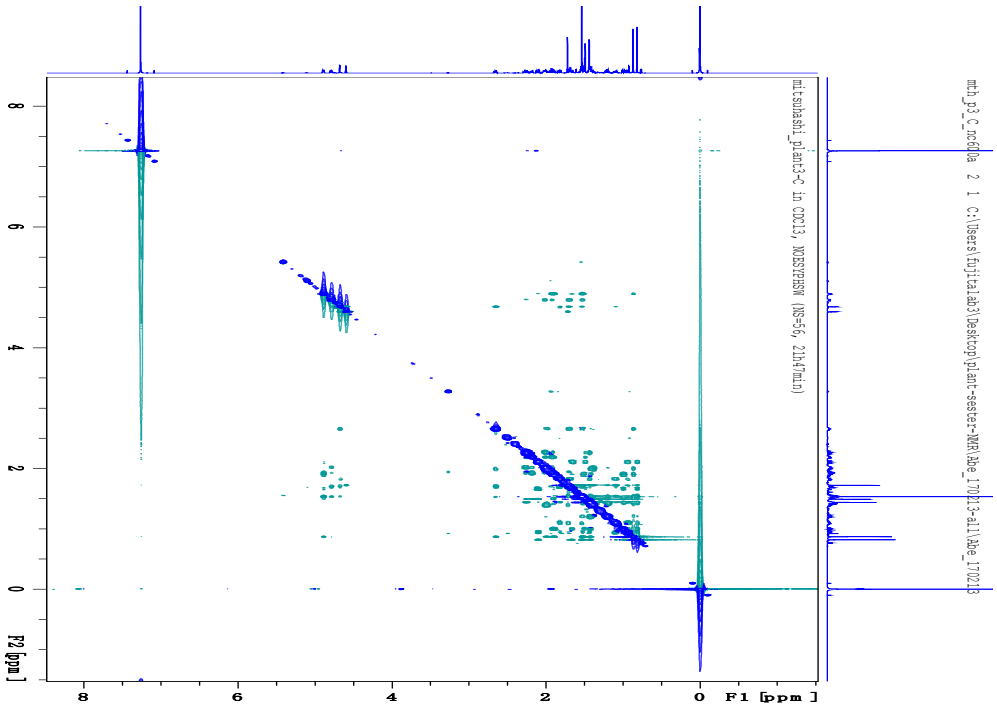


Figure S2-6. NOESY spectrum of C1 in $CDCl_3$ (600 MHz).

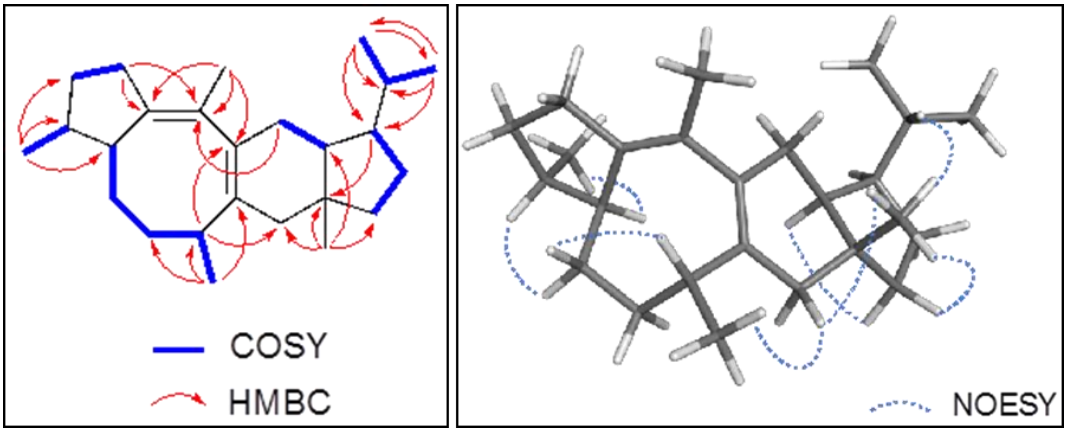


Figure S3. ^{1}H-^{1}H COSY, key HMBC, and Key NOESY correlations in C2.

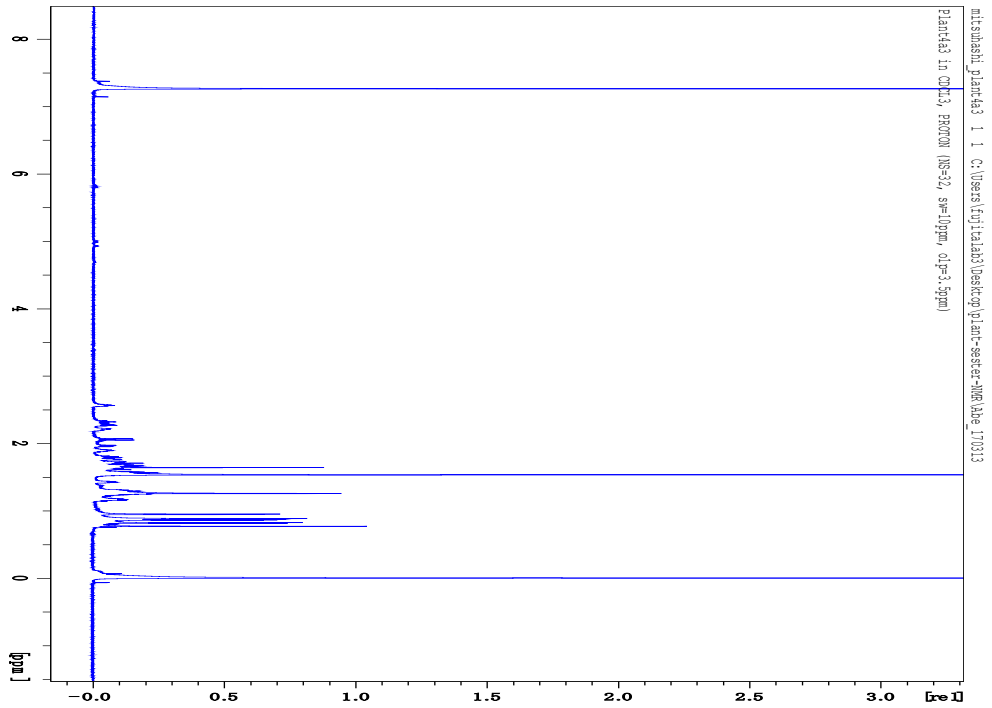


Figure S3-1. ^{1}H NMR spectrum of C2 in $CDCl_3$ (900 MHz).

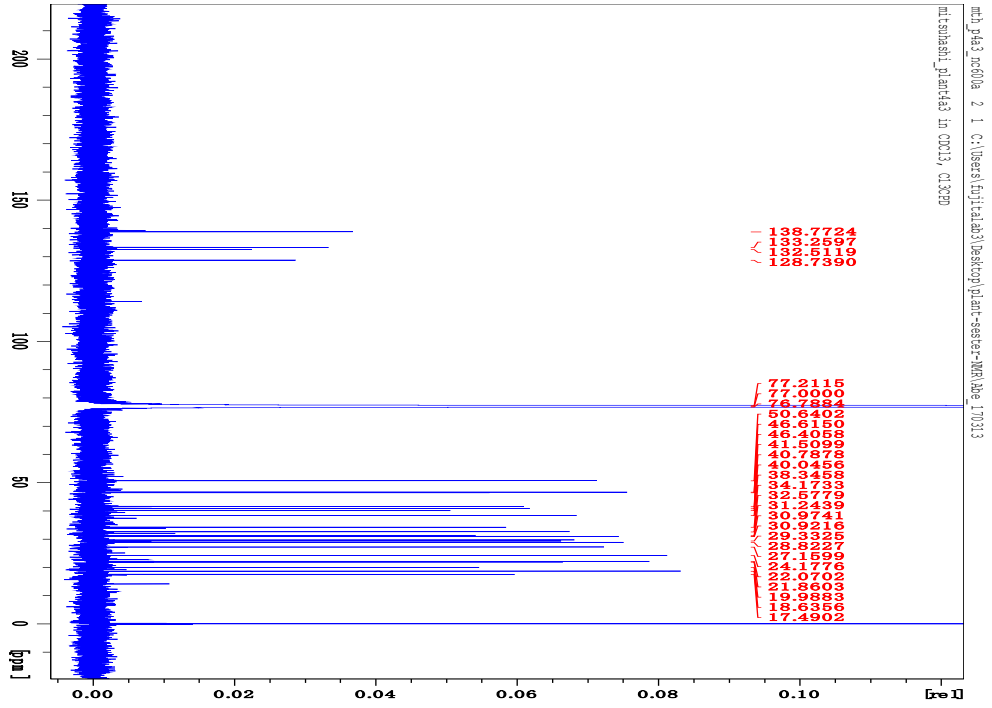


Figure S3-2. ^{13}C NMR spectrum of C2 in $CDCl_3$ (200 MHz).

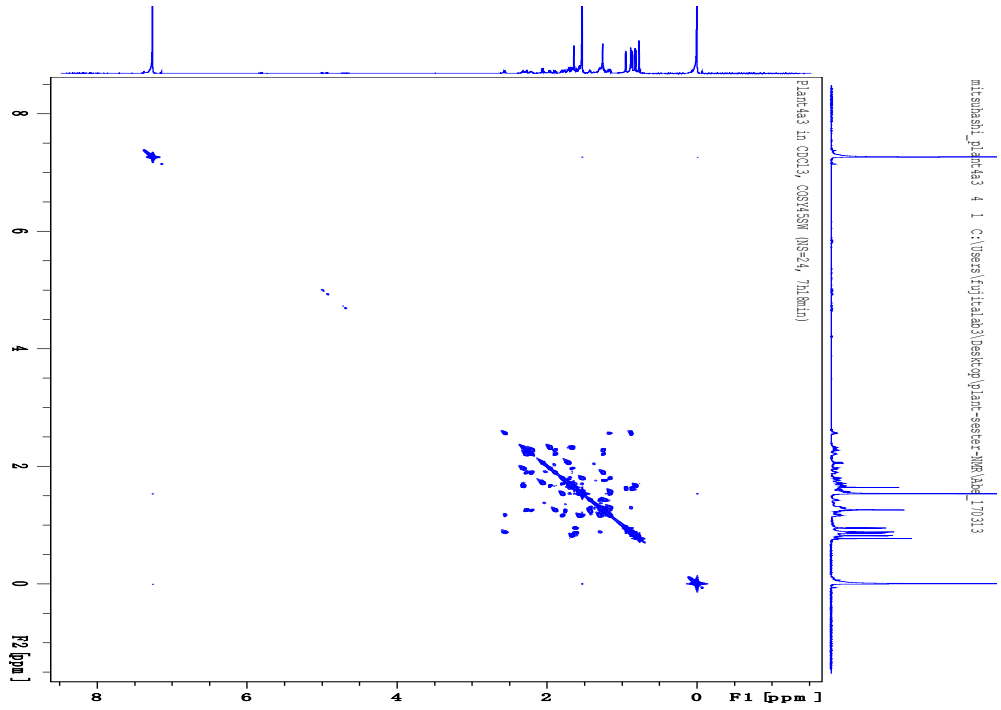


Figure S3-3. ^{1}H-^{1}H COSY spectrum of C2 in $CDCl_3$ (900 MHz).

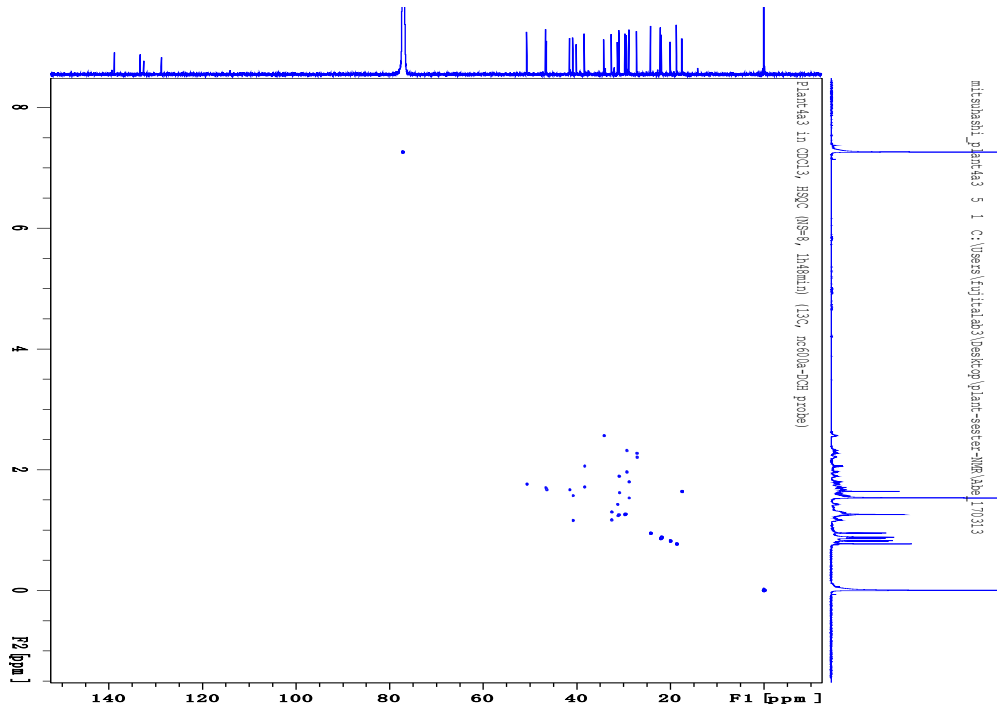


Figure S3-4. HSQC spectrum of C2 in $CDCl_3$ (900 MHz).

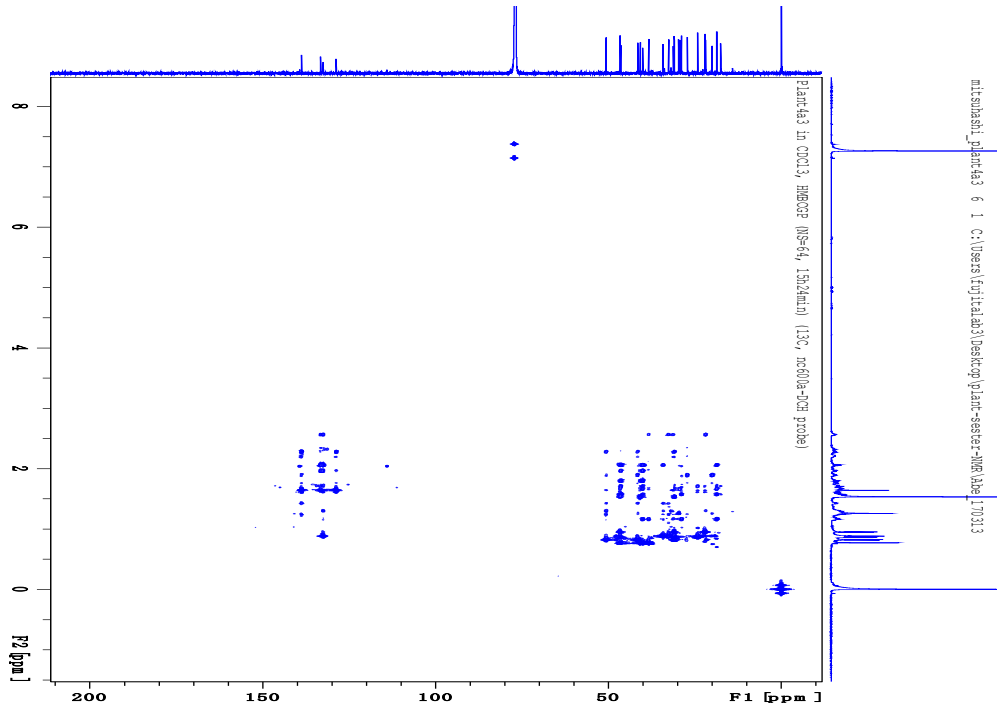


Figure S3-5. HMBC spectrum of C2 in $CDCl_3$ (900 MHz).

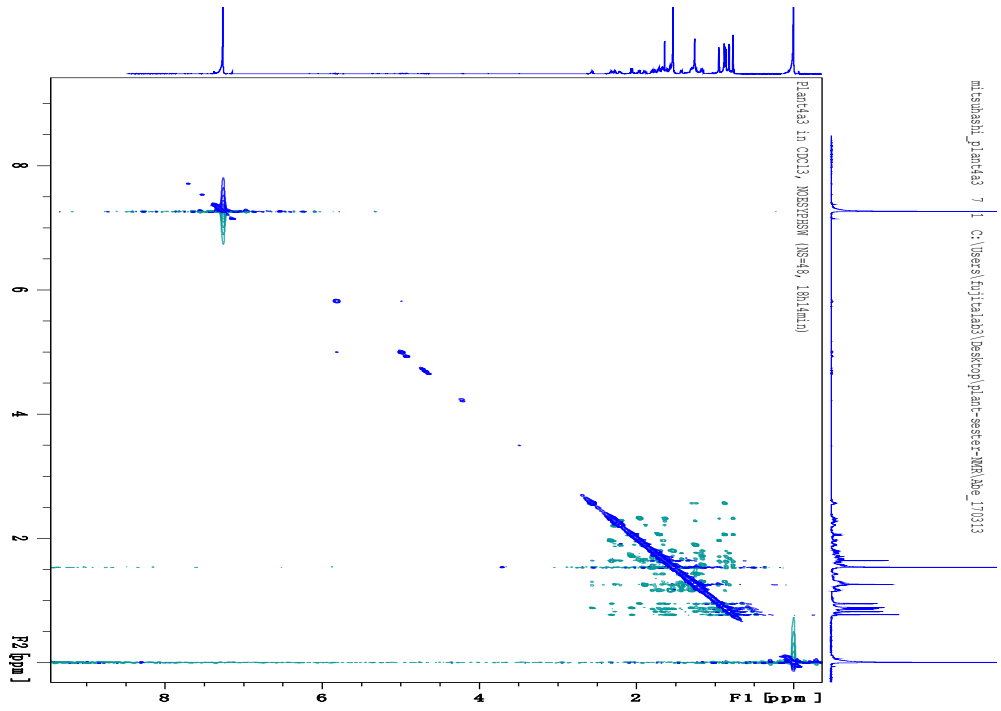


Figure S3-6. NOESY spectrum of C2 in $CDCl_3$ (900 MHz).

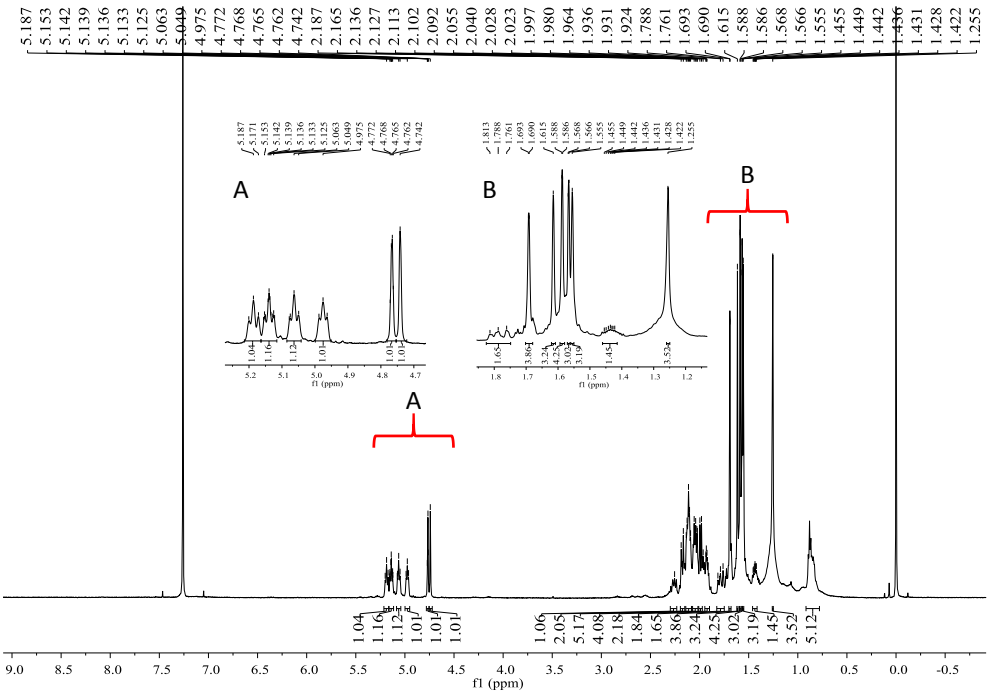


Figure S4-1. ^{1}H NMR spectrum of C4 in $CDCl_3$ (500 MHz).

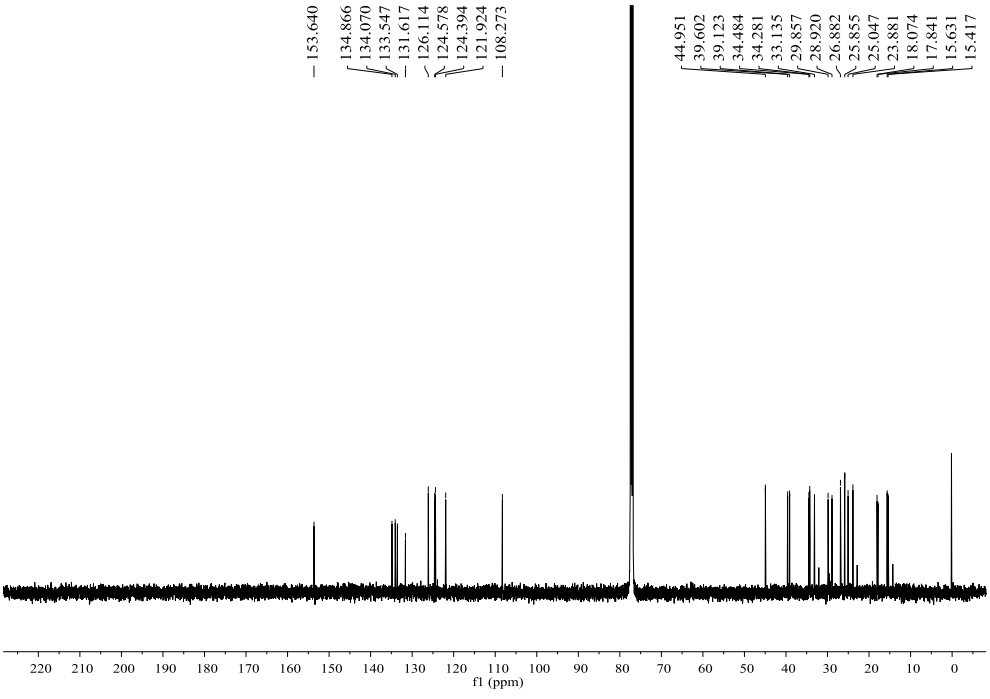


Figure S4-2. ^{13}C NMR spectrum of C4 in $CDCl_3$ (125 MHz).

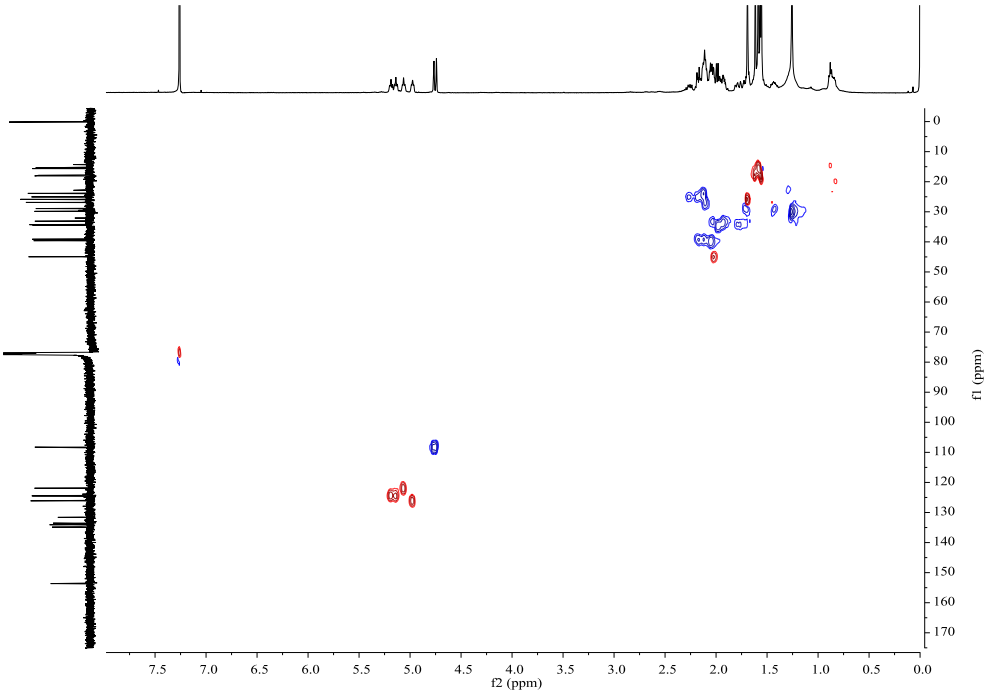


Figure S4-3. HSQC spectrum of C4 in $CDCl_3$ (500 MHz).

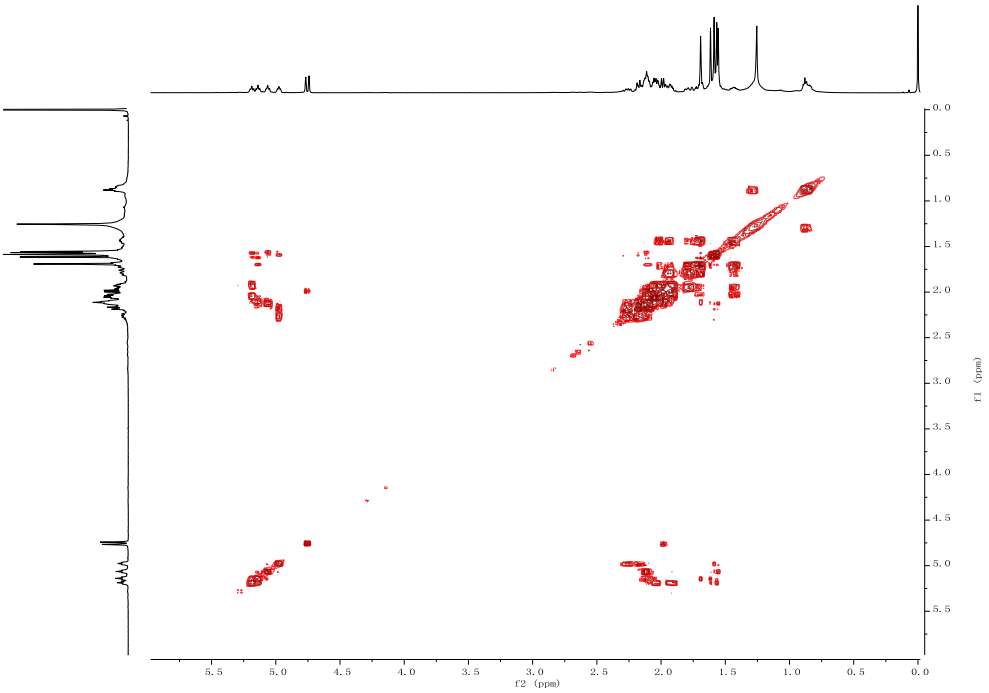


Figure S4-4. ^1H-^1H COSY spectrum of C4 in $CDCl_3$ (500 MHz).

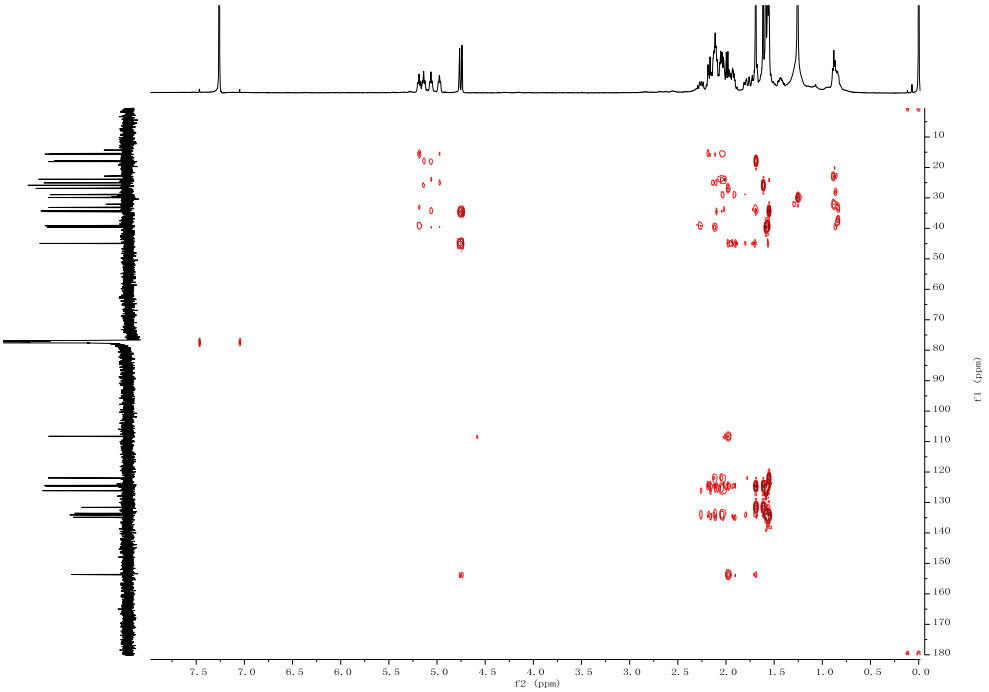
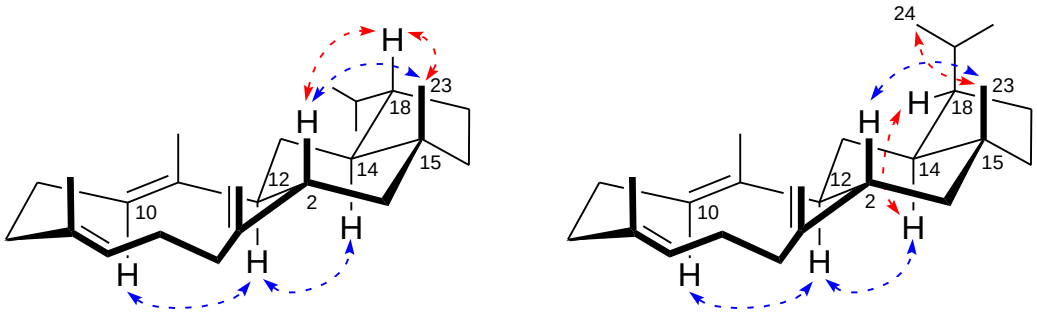


Figure S4-5. HMBC spectrum of C4 in $CDCl_3$ (500 MHz).



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

Figure S5. Key NOESY correlations for C5 and C7.

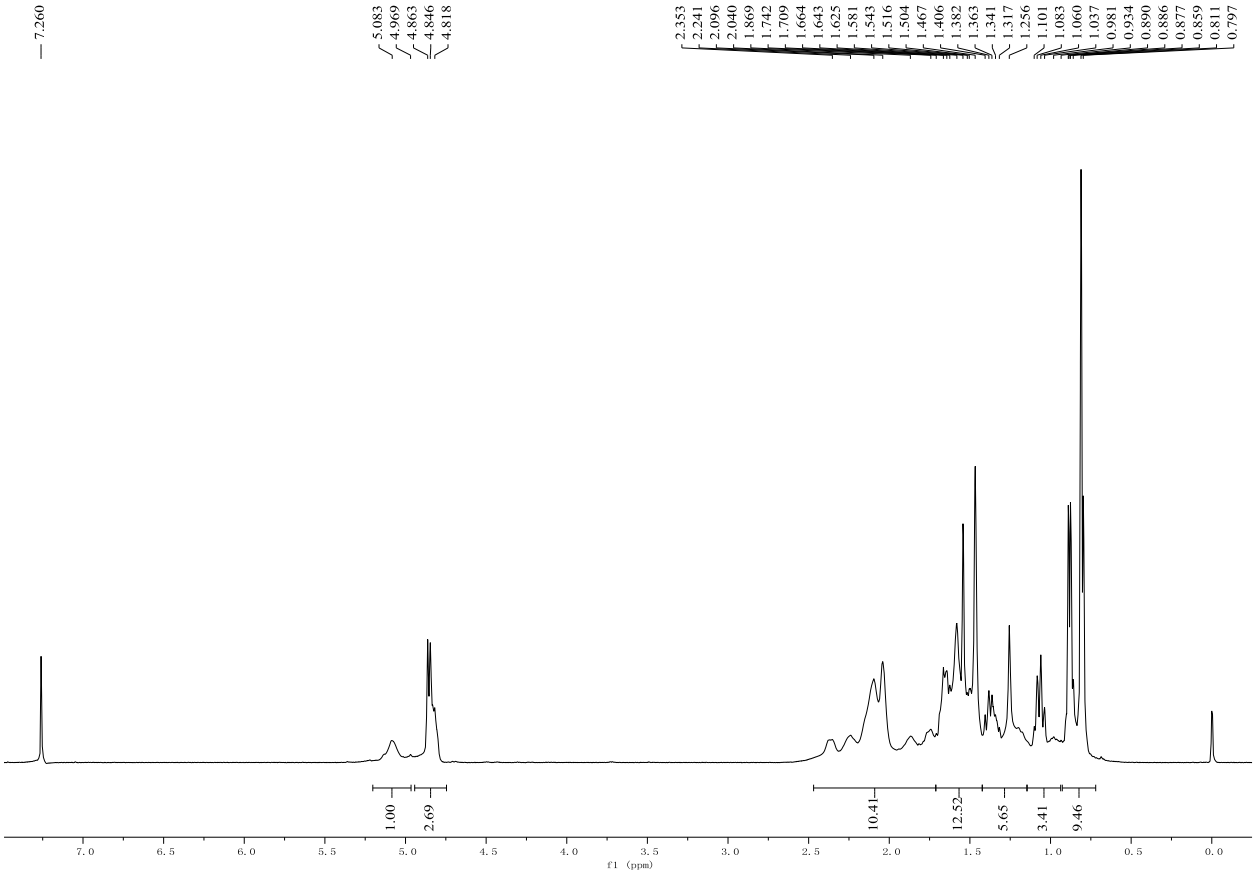


Figure S5-1. ^{1}H NMR spectrum of C5 in $CDCl_3$ (500 Hz).

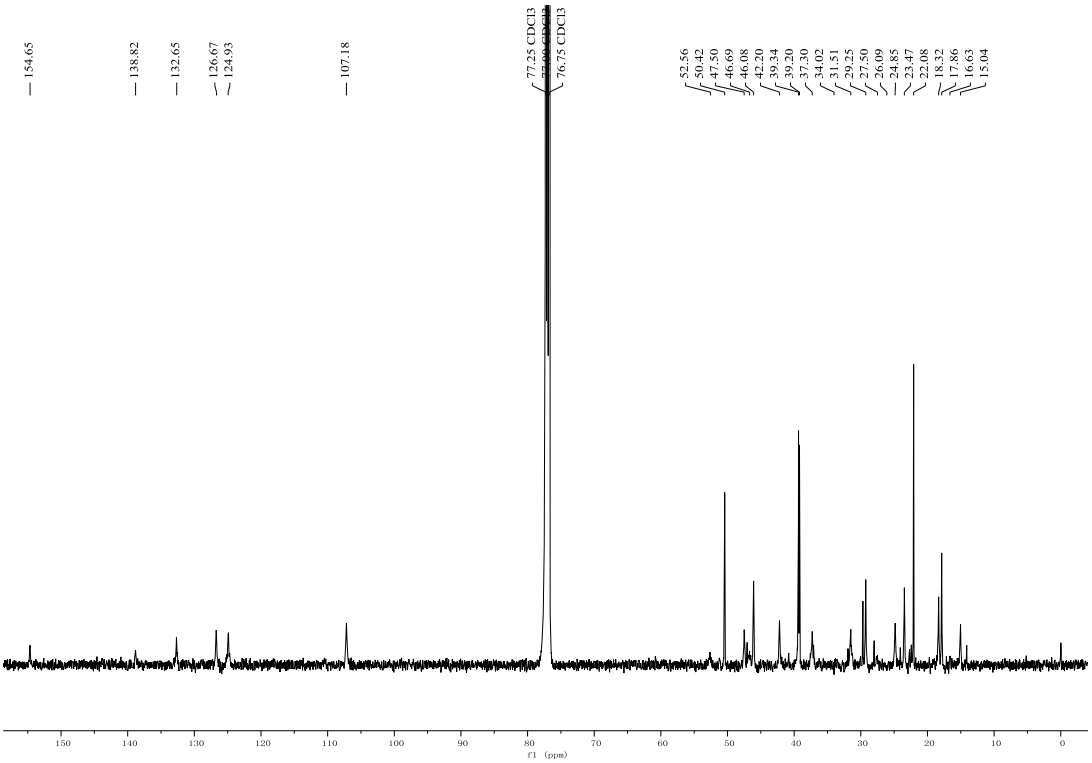


Figure S5-2. ^{13}C NMR spectrum of C5 in $CDCl_3$ (125 MHz).

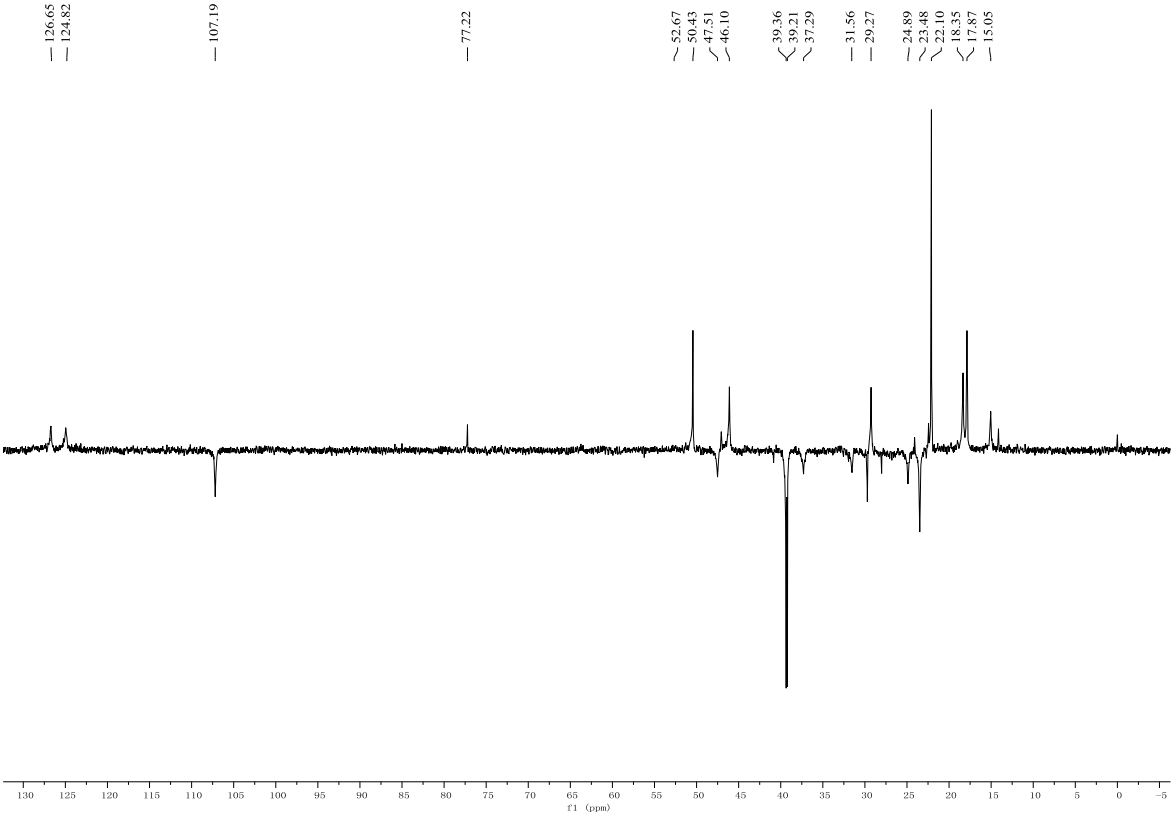


Figure S5-3. DEPT spectrum of C5 in $CDCl_3$ (125 MHz).

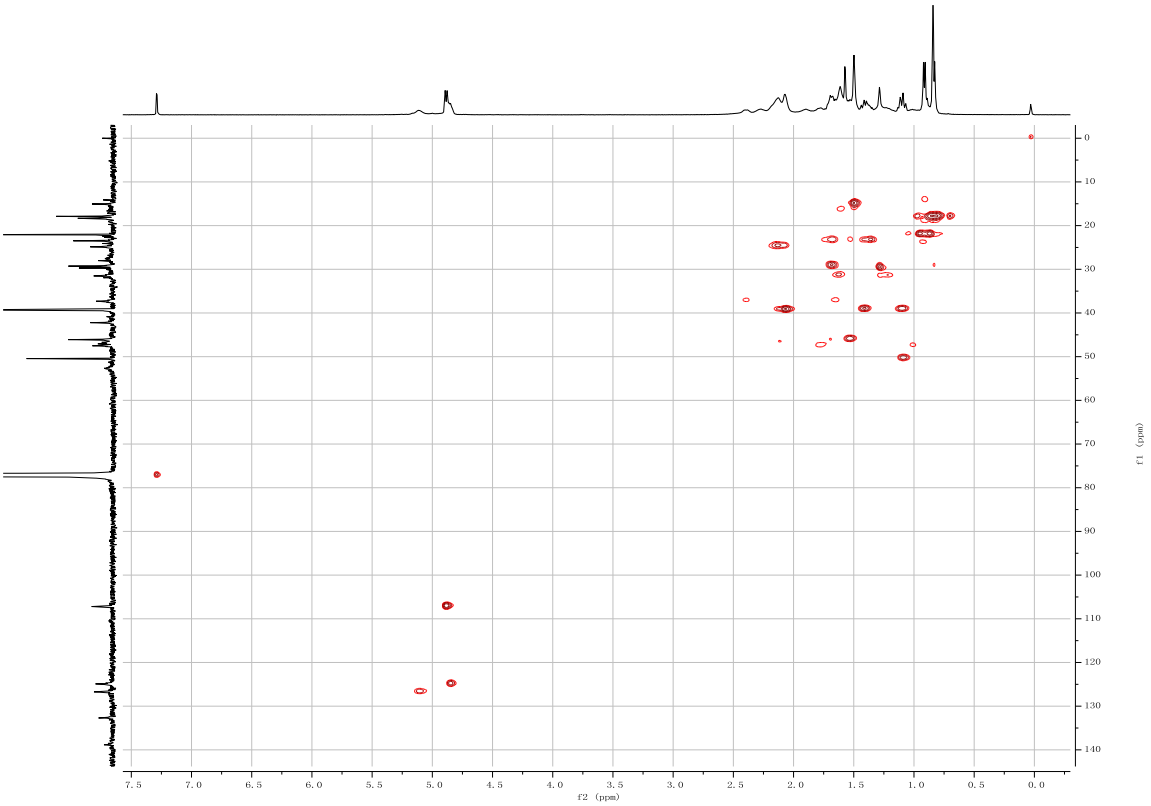


Figure S5-4. HSQC spectrum of C5 in $CDCl_3$ (500 MHz).

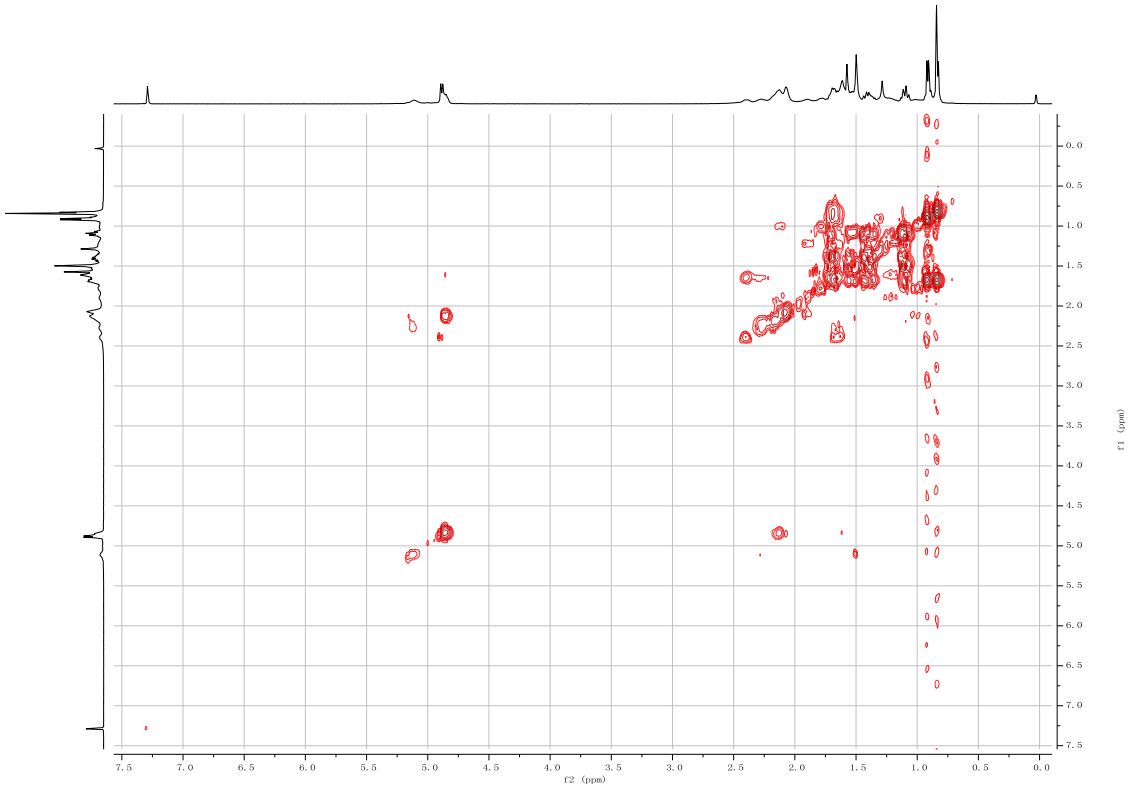


Figure S5-5. ^{1}H-^{1}H COSY spectrum of C5 in $CDCl_3$ (500 MHz).

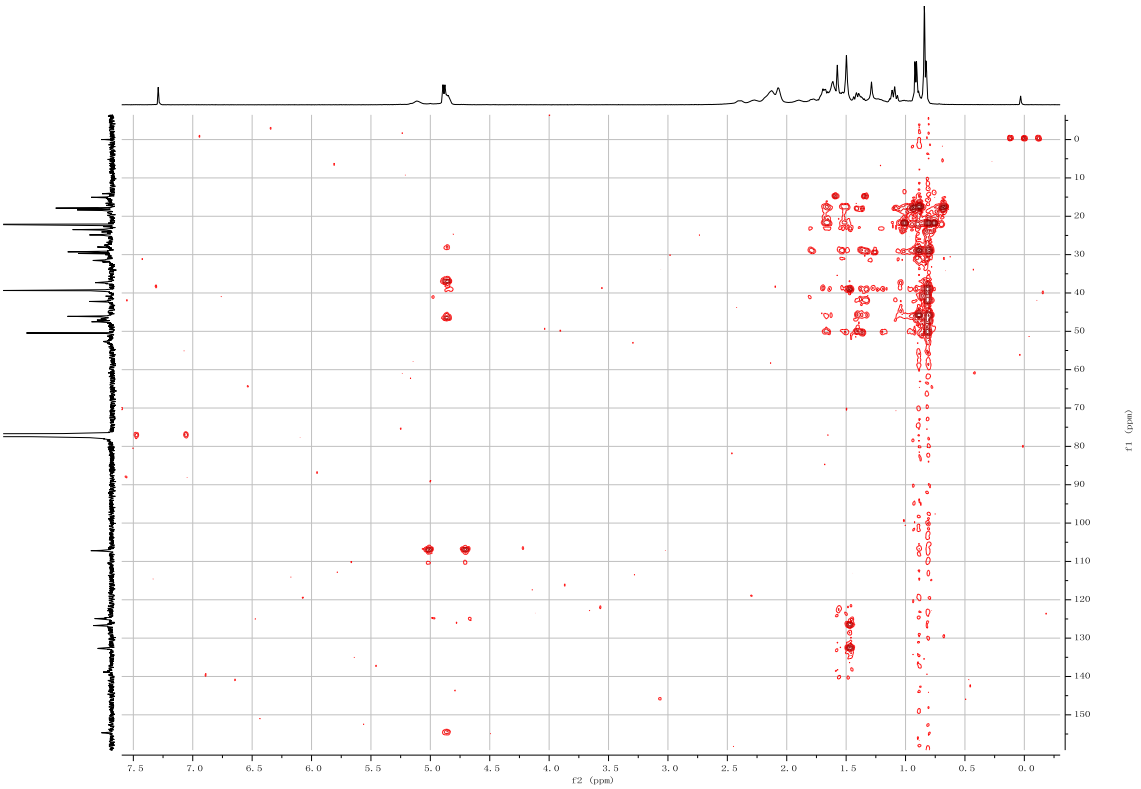


Figure S5-6. HMBC spectrum of C5 in $CDCl_3$ (500 MHz).

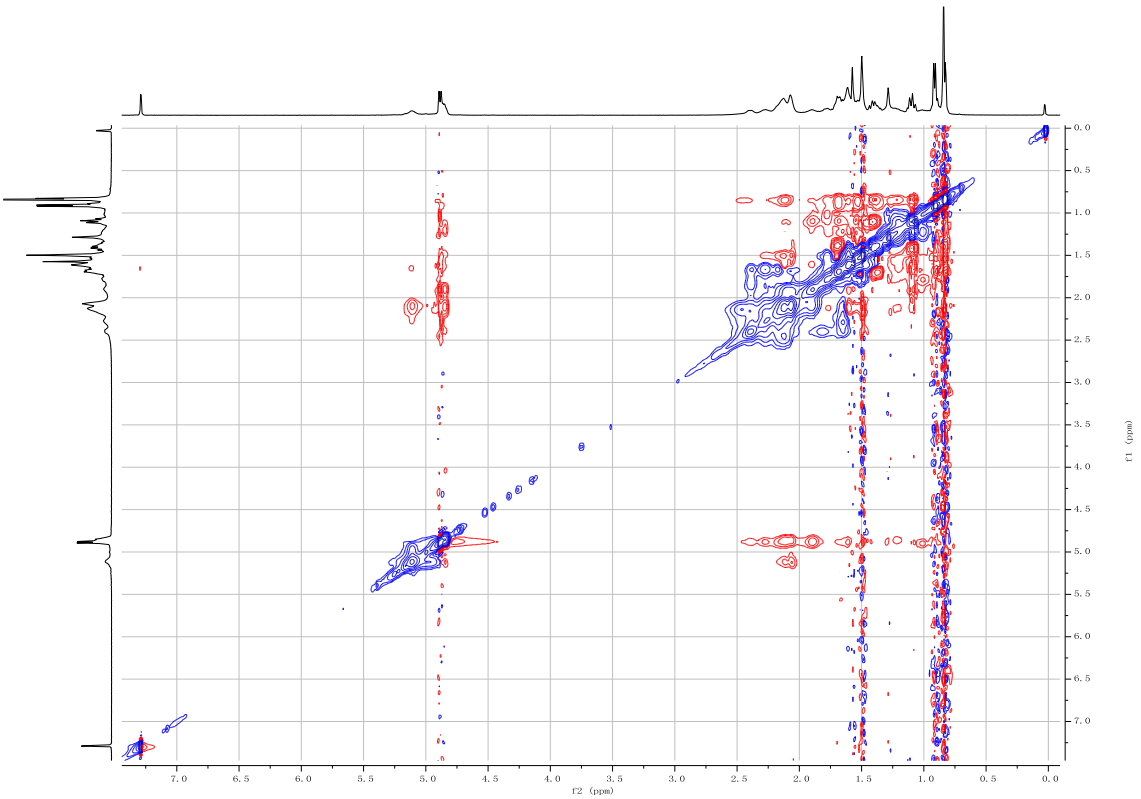


Figure S5-7. ROESY spectrum of C5 in $CDCl_3$ (500 MHz).

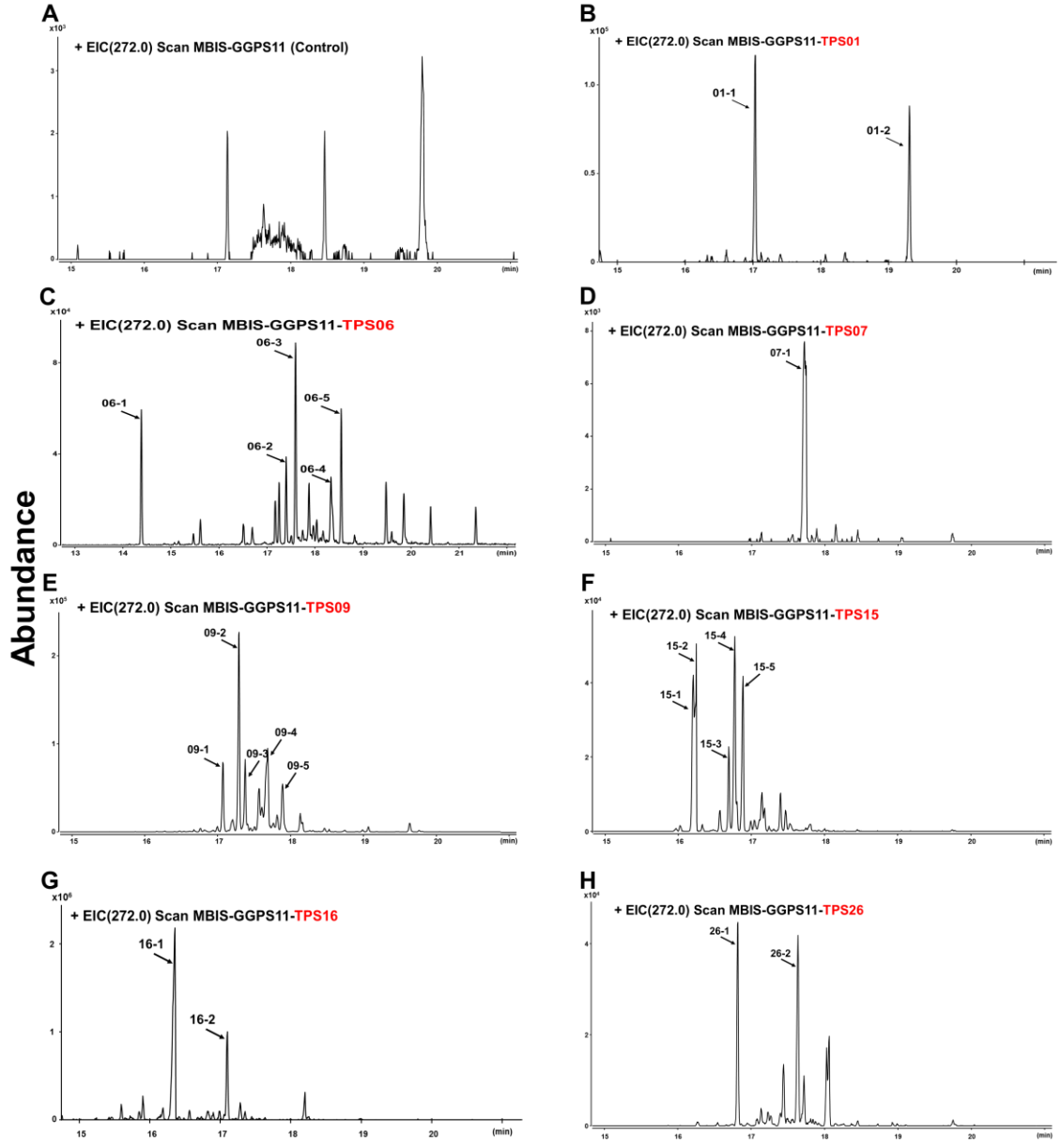


Figure S6. Di-TPS activity screening of seven Arabidopsis TPS-a members in *E. coli* system. A, Control, only *AtGGPP11* (*At4g36810*) was expressed in *E. coli*; **B** to **H,** seven TPSs-a members (AtTPS01, AtTPS06, AtTPS07, AtTPS09, AtTPS15, AtTPS16, and AtTPS26) from *Arabidopsis thaliana* (four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

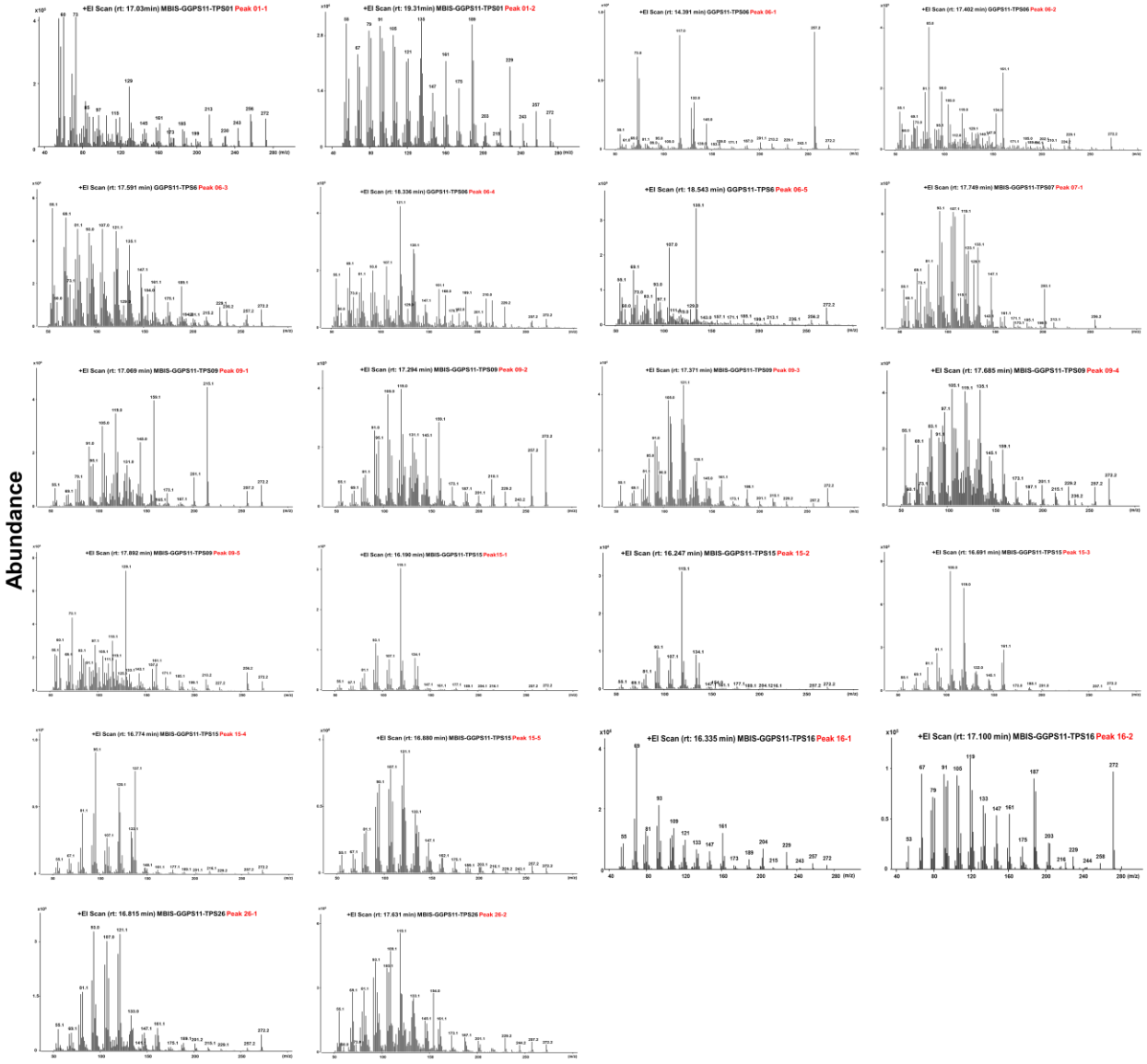


Figure S7. Mass spectrum of unknown di-terpens generated Arabidopsis TPSs. AtTPS01 (two main peaks, 01-1 and 01-2), AtTPS06 (five main peaks, 06-1 to 06-5), AtTPS07 (one main peak, 07-1), AtTPS09 (five main peaks, 09-1 to 09-5), AtTPS15 (five main peaks, 15-1 to 15-5), AtTPS16 (two main peaks, 16-1 and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

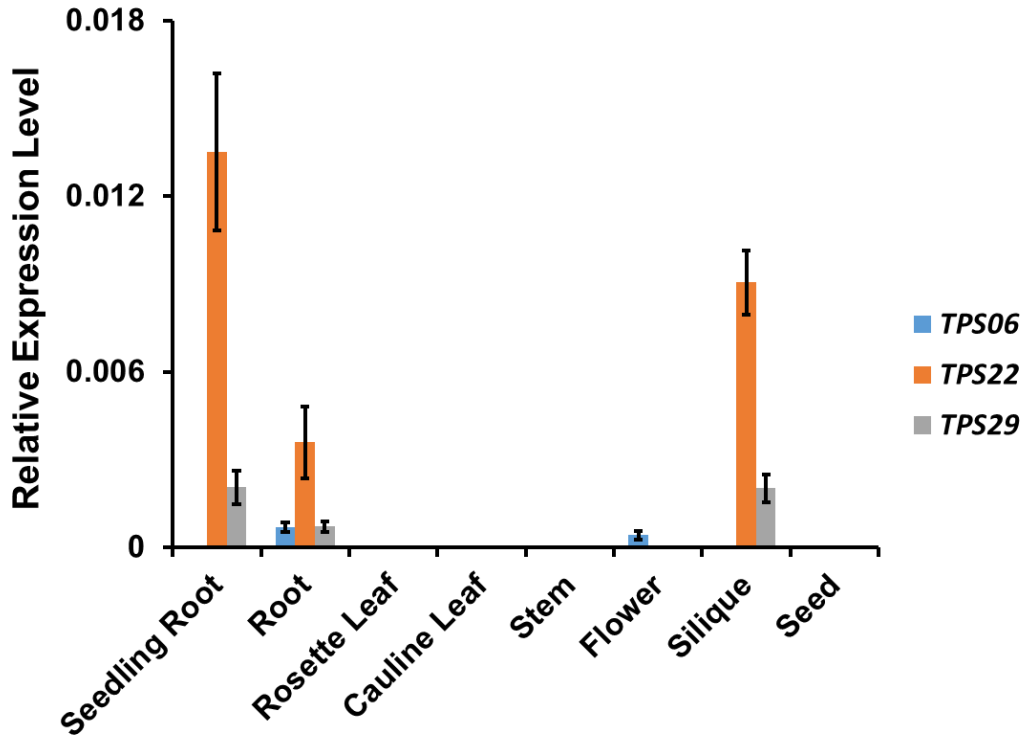


Figure S8. Tissue-specificity of *TPS06*, *TPS22*, and *TPS29* in *Arabidopsis thaliana*. Tissue samples were harvested from 8-week-old plants and 2-week-old seedlings. *Actin 2* (*At3g18780*) was used as a reference gene in the analysis. Transcript levels of each gene in the tested tissue are expressed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

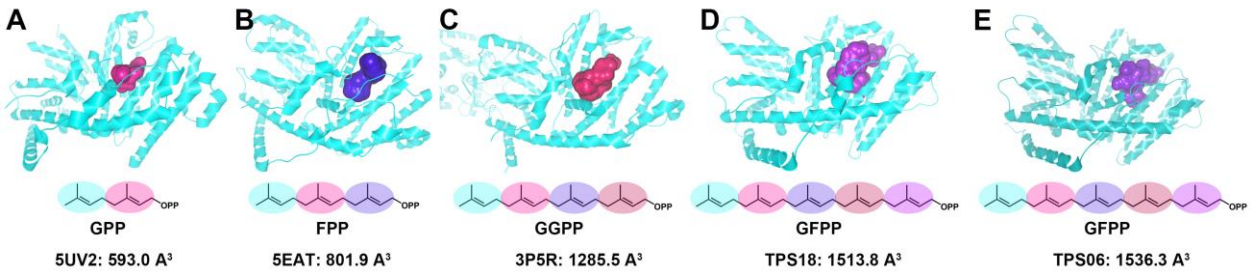


Figure S9. Catalytic cavity comparison among different types of terpene synthases. A. mono-TPS ((+)-limonene synthase complexed with 2-fluoroneryl diphosphate, Protein Data Bank entry 5UV2); B, sesqui-TPS (tobacco 5-epi-aristolochene synthase complexed with farnesyl hydroxyphosphonate, Protein Data Bank 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.

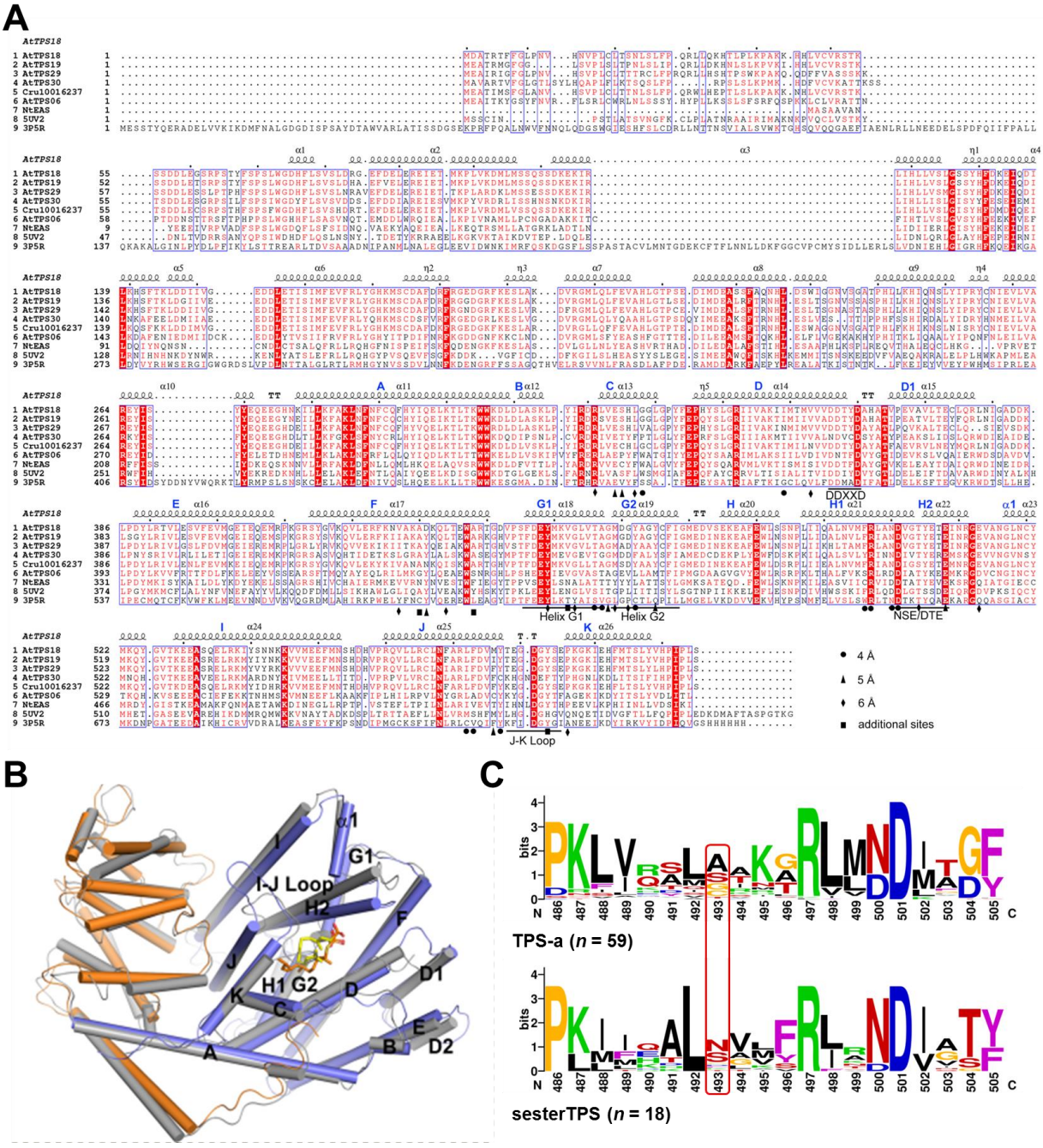


Figure S10. Amino acid sequence comparison of *Brassicaceae* sesterTPSs and other plant TPSs.

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).

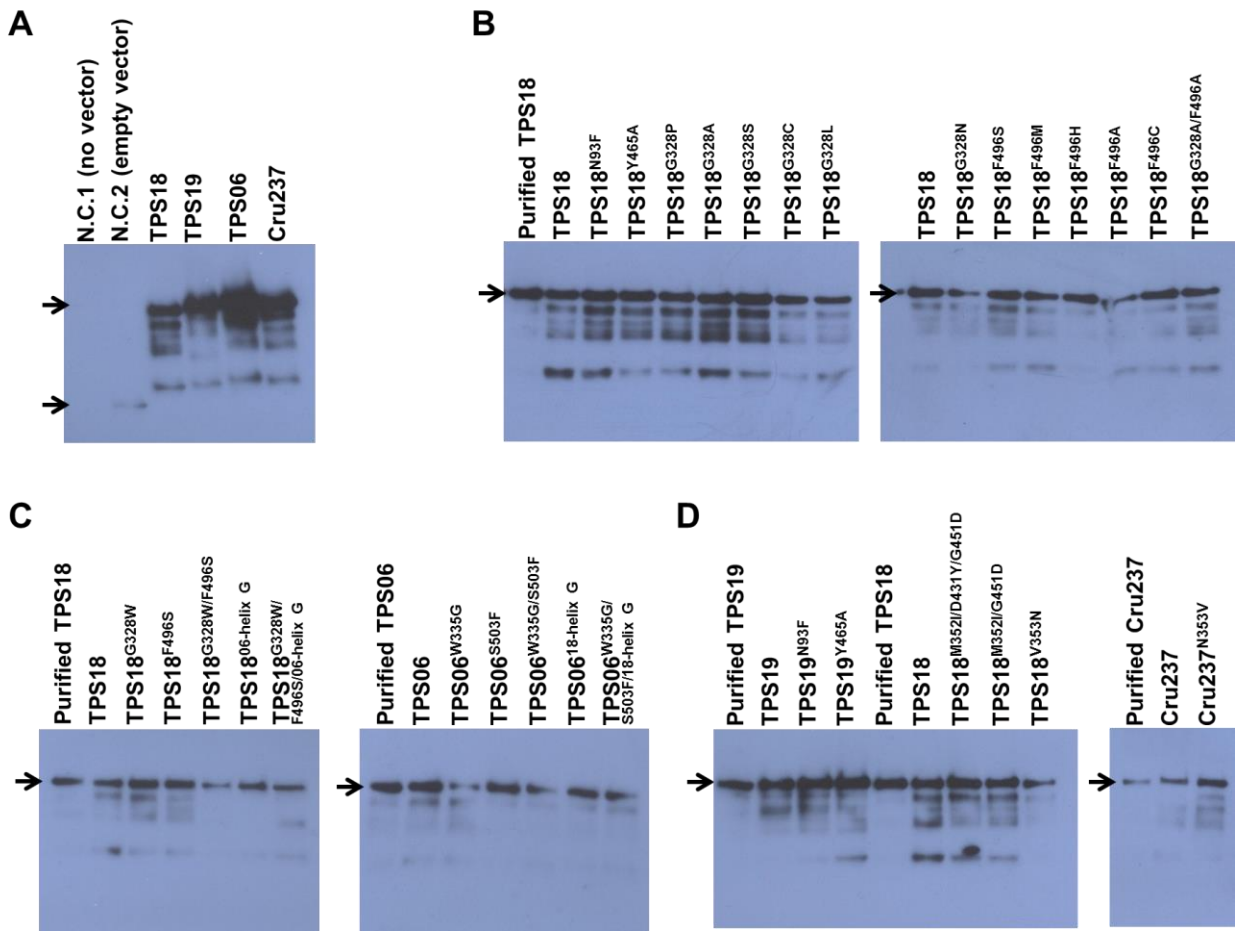


Figure S11. Measurement of soluble C-terminal MBP-tagged sesterTPSs in *E.coli* with anti-MBP monoclonal antibody. The calculated molecular weight of each MBP-tagged sesterTPSs is listed below: MBP-tag itself (43.0 KDa), AtTPS06 (107.5 KDa), AtTPS18 (106.6 KDa), AtTPS19 (107.0 KDa), and Cr237 (107.5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

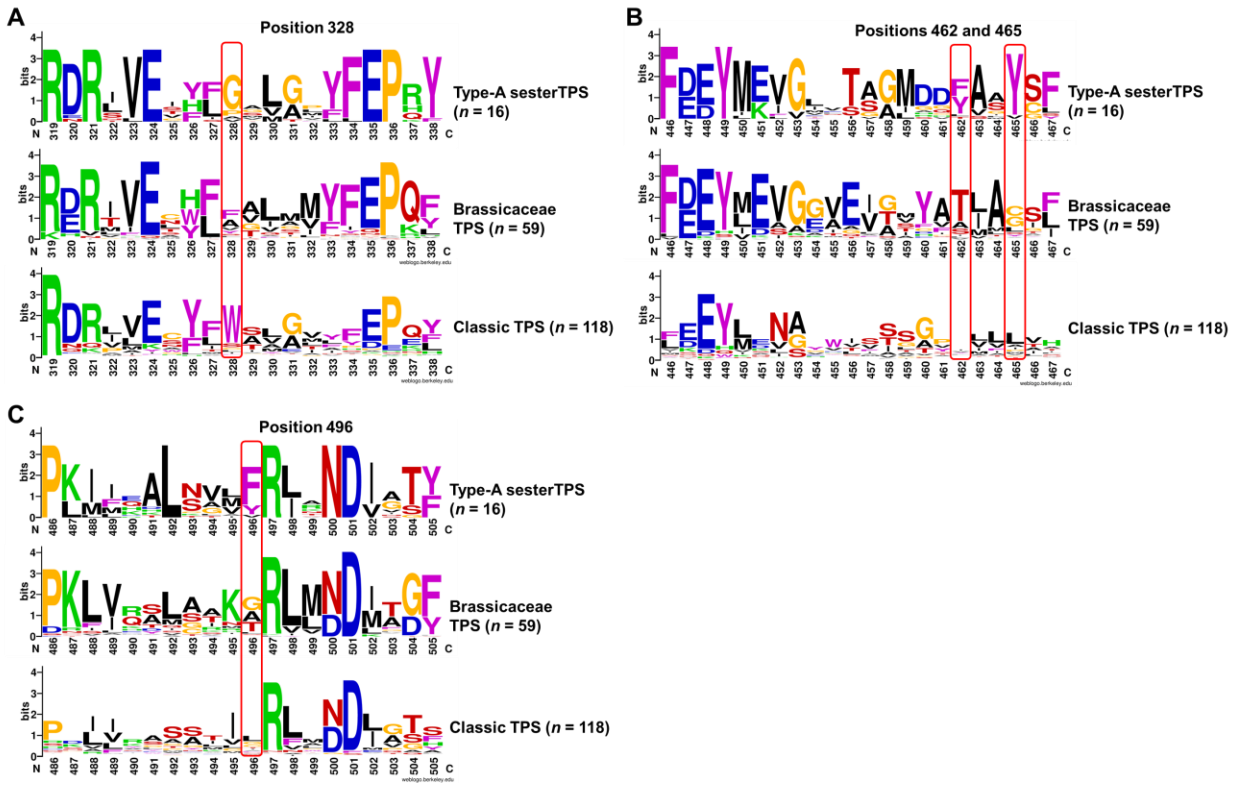


Figure S12. Sequence comparison between active plant type-A sesterTPSs (*n* = 16) and other TPS-a members (*n* = 59) from *Brassicaceae* plants (see the legend for Figure S11B) and classic TPSs (belong to TPS-a, -b, -g subfamilies; *n* = 118) from other angiosperms plants, including *Aquilegia coerulea* 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).

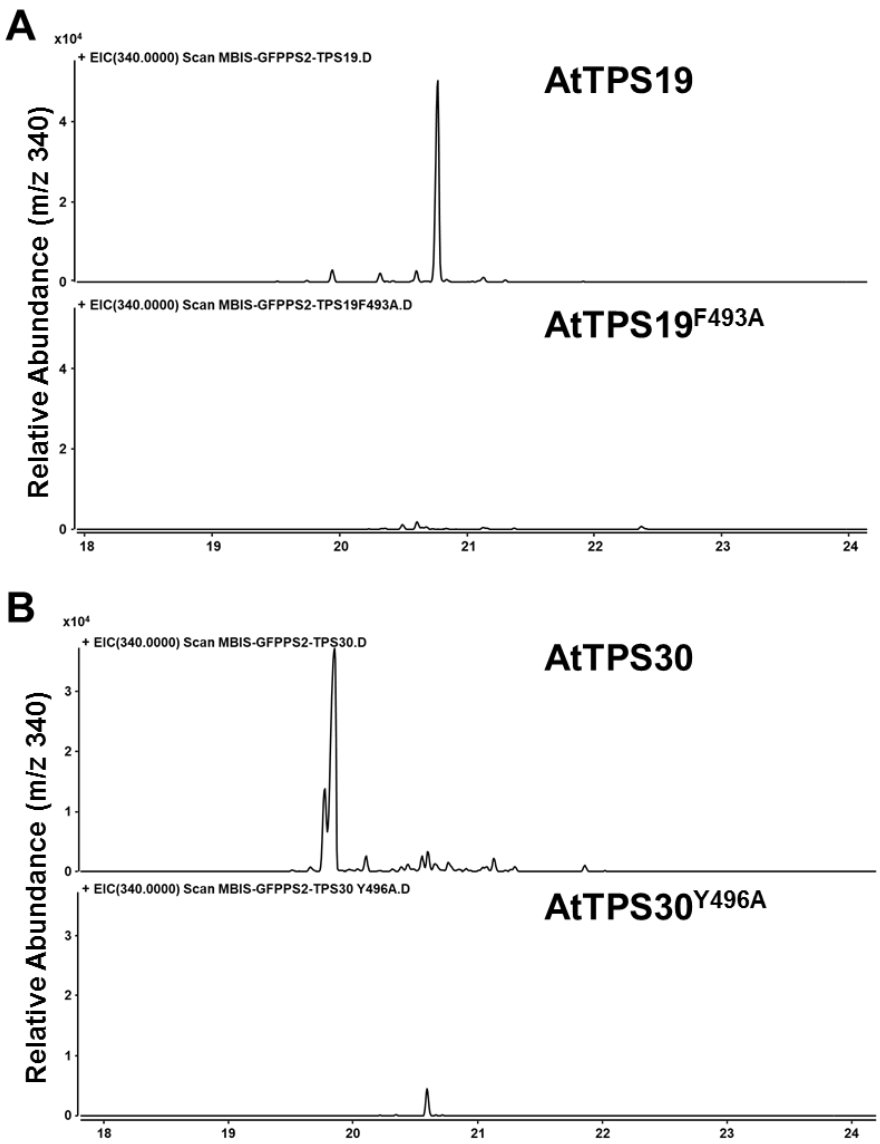


Figure S13. Sester-TPS activity screening of AtTPS19, AtTPS30 and their mutants in *E. coli* system.

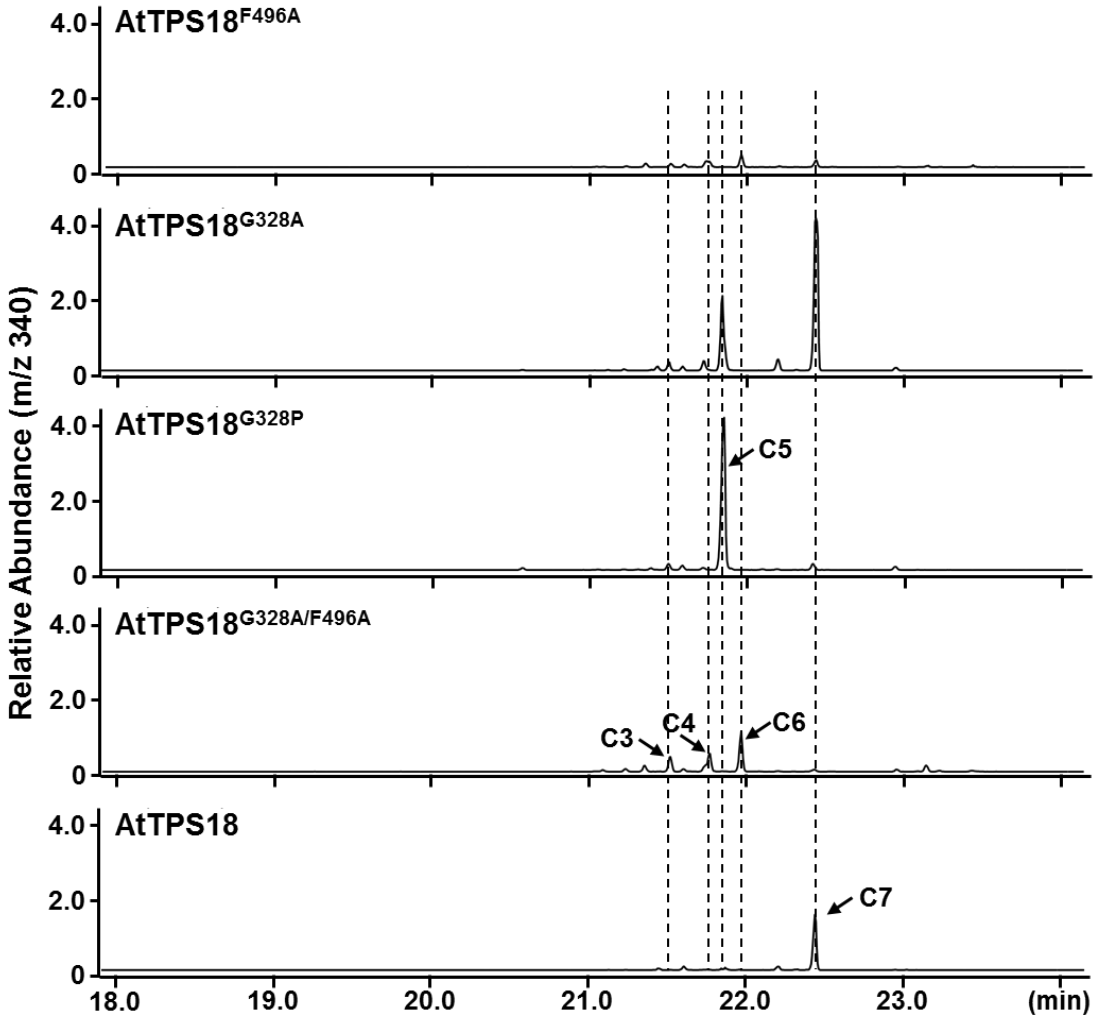


Figure S14. Sester-TPS activity screening of AtTPS18 and its mutants in *E. coli* system. The novel sesterterpenes (**C4** and **C5**) were further purified and structurally elucidated in this study.

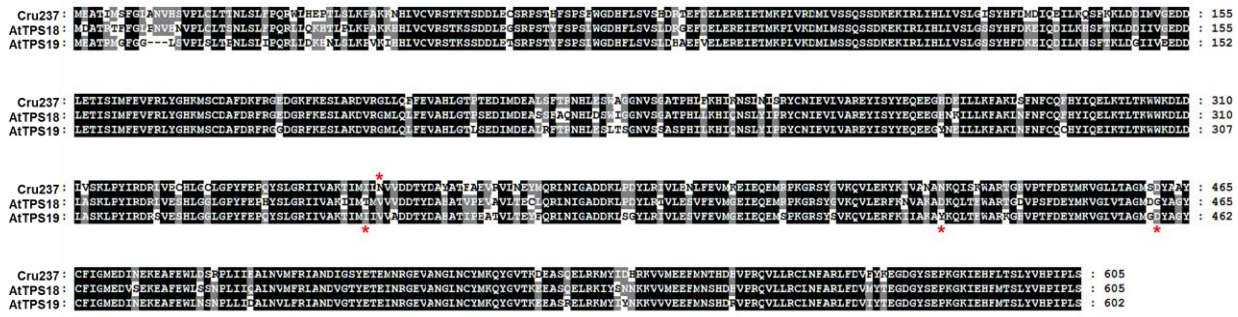


Figure 15. Sequence alignment of AtTPS18, AtTPS19, and Cru237 from *Capsella rubella.* The different amino acids revealed by structural comparison were highlighted with red stars.

Supplemental Tables

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

No.	**Compound 1**		**Compound 2**	
	δ_H (mult., *J* in Hz)	δ_C (ppm)	δ_H (mult., *J* in Hz)	δ_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 $CHCl_3$ (δ_H 7.26) and to $CDCl_3$ (δ_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**	
	δ_H (mult., *J* in Hz)	δ_C (ppm)	δ_H (mult., *J* in Hz)	δ_C (ppm)
1	1.93, m 2.03, m	33.1	0.98, m 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 2.18, o	39.1	1.62, m 2.36, m	37.3
5	2.18, o 2.26, m	25.0	2.14, m 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 1,95, m	34.3	1.86, m	52.6
13	1.43, m 1.72, o	28.9	1.22, m 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 1.38, m	39.2
17	2.10, o	26.9	1.33, m 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 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 $CHCl_3$ (δ_H 7.26) and to $CDCl_3$ (δ_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.

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