

New *Phytologist* Supporting Information

Article title: Promiscuous terpene synthases from *Prunella vulgaris* highlight the importance of substrate and compartment switching in terpene synthase evolution

Authors: Sean R. Johnson, Wajid Waheed Bhat, Radin Sadre, Garret P. Miller, Alekzander Sky Garcia, Björn Hamberger

Article acceptance date: 22 February 2019

The following Supporting Information is available for this article:

Method S1 Nanopore sequencing and assembly

Table S1 Synthetic oligonucleotides used in this study

Fig. S1 Stereo view of homology models for each enzyme characterized in this study

Fig. S2 Multiple sequence alignment of cytosolic and plastidial TPS-a enzymes found in Lamiaceae, Euphorbiaceae, Solanaceae, and Brassicaceae

Fig. S3 GC-MS of hexane extracts TPS activity assays in vitro, in *E. coli*, and in *N. benthamiana*

Fig. S4 Mass spectra of enzyme products and selected reference spectra

Fig. S5 Peak q only appears when PvTPS4 or PvTPS5 is co-expressed with FPPS in *N. benthaminana*

Fig. S6 NMR spectra of PvHVS major product from GGPP, identified as 11-hydroxy vulgarisane.

Fig. S7 UHPLC/MS of 11-Hydroxyvulgarisane in *P. vulgaris*

Table S2 Exact masses and relative abundance for ions of vulgarisin-related compounds detected by UHPLC/MS in *P. vulgaris* root extracts

Fig. S8 UHPLC/MS of vulgarisin A and D isomers in *P. vulgaris*

Methods S1. Nanopore sequencing and assembly.

RNA was extracted from root tissue according to a previously described method (Hamberger *et al.*, 2011), mRNA was selected from 16 µg total RNA using Dynabeads Oligo (dT)25 (ThermoFisher Scientific) yielding 119 ng mRNA. The mRNA was converted to cDNA using the Oxford Nanopore cDNA-PCR Sequencing Kit (SQK-PCS108) and sequenced on the Gridlon. The resulting library consisted of 2,146,282 reads with an average length of 735 bp, and an N50 of 817 bp. The Illumina reads were from PhytoMetSyn were cleaned and normalized using Bbnorm (sourceforge.net/projects/bbmap/), and the raw reads were subjected to error correction using LoRDEC (Salmela & Rivals, 2014) according to settings suggested for transcriptome data (Hoang *et al.*, 2017). A hybrid assembly using the corrected Nanopore reads and the Illumina reads was generated using IDP-denovo (Fu *et al.*, 2018). The resulting assembly seemed to be of much worse quality than the Illumina-only assembly and we did not pursue further analysis.

Table S1. Synthetic oligonucleotides used in this study. Lower case letters indicate vector overhangs to enable In-Fusion (Takara) cloning.

Name	Sequence
PvTPS2 F	GATGAGCTCTCTCACAAATCTCTACAG
PvTPS2 R	TGAAGAATCAAACCTTGATATGGATTGAGC
PvHVS F	CTCAAAATGAGCTCTCTCTCAATTCCC
PvHVS R	GTCTGAATGGGGAGTCAAATATG
PvTPS4 F	ATGGGTGAAGAGAGCAGTGTAAG
PvTPS4 R	CAAGGGAATCAAATGAGAAGAGG
PvTPS5 F	GAAATGGCAGCTGCAGGC
PvTPS5 R	TCAGATGACAAGAGGATTGAGTAGGAG
PvTPS2 pEAQ Infusion F	ttctgccccaaattcgCGAATGAGCTCTCTCACAAATCTCTA
PvTPS2 pEAQ Infusion R	agttaaaggcctcgaTCAAACCTTGATATGGATTGAGCAGC
PvHVS pEAQ Infusion F	ttctgccccaaattcgCGAATGAGCTCTCTCTCAATTCCCTT
PvHVS pEAQ Infusion R	agttaaaggcctcgaTCAAATATGTATAGGATTCATCAATAGAGCAG
PvTPS4 pEAQ-TP Infusion F	ggcgggaagagttaacATGGGTGAAGAGAGCAGTGTA
PvTPS4 pEAQ-TP Infusion R	agttaaaggcctcgaTCAAATGAGAAGAGGATTAAGAAACAAAGC
PvTPS5 pEAQ-TP Infusion F	ggcgggaagagttaacATGGCAGCTGCAGGCATTG
PvTPS5 pEAQ-TP Infusion R	agttaaaggcctcgaTCAGATGACAAGAGGATTGAGTAGG
AtRuBisCO small pEAQ Infusion F	ttctgccccaaattcgATGGCTTCCTCTATGCTCTCTTC
AtRuBisCO small pEAQ Infusion R	agttaaaggcctcgaGTTAACTCTTCCGCCGTTGC
pET28plus PvTPS2 Infusion F	gaaggagatataccatgTCAGCTGTGATGGTGGATCATGC
pET28plus PvTPS2 Infusion R	ggtggtggtggtgctcgaAACTTGATATGGATTGAGCAGC
pET28plus PvHVS Infusion F	gaaggagatataccatgTCAGCTGTGATGGTGGAAAGG
pET28plus PvHVS Infusion R	ggtggtggtggtgctcgaAAATATGTATAGGATTCATCAATAGAGC
pET28plus PvTPS4 Infusion F	gaaggagatataccATGGGTGAAGAGAGCAGTGTAAGT
pET28plus PvTPS4 Infusion R	ggtggtggtggtgctcgaAAATGAGAAGAGGATTAAG
pET28plus PvTPS5 Infusion F	gaaggagatataccATGGCAGCTGCAGGCATTG
pET28plus PvTPS5 Infusion R	ggtggtggtggtgctcgaAGATGACAAGAGGATTGAGTAGG
zFPPS pEAQ-HT F	ttctgccccaaattcgATGAGTTCTTTGGTTCTTCAATGTTG
zFPPS pEAQ-HT R	agttaaaggcctcgaTCAATATGTGTGTCCACCAAAAACG
NNPPS pEAQ-HT Infusion F	ttctgccccaaattcgATGAACTCTTCAATAGTGTCTCAACAC
NNPPS pEAQ-HT Infusion R	agttaaaggcctcgaTCAATATGTGTGTCCACCAAAAAC
NNPPS F	ATGAACTCTTCAATAGTGTCTCAAC
NNPPS R	TCAATATGTGTGTCCACCAAAAACG
NNPPS pACYCDuet Infusion F	aaggagatatacatATGTCTGATCGTGGACTCAGC
NNPPS pACYCDuet Infusion R	cttaccagactcgagTCAATATGTGTGTCCACCAAAAAC
pEAQ_PvTPS2_GFP_For	attctgccccaaattcgATGAGCTCTCTCACAAATCTCT
pEAQ_PvTPS2_GFP_Rev	ctagtcataccggtcgcAACTTGATATGGATTGAGCAG
pEAQ_PvHVS_GFP_For	attctgccccaaattcgATGAGCTCTCTCTCAATTCC

pEAQ_PvHVS_GFP_Rev	ctagtcataccggtcgcAATATGTATAGGATTCATCAATA
pEAQ_PvTPS4_GFP_For	attctgcccaaattcgATGGGTGAAGAGAGCAGT
pEAQ_PvTPS4_GFP_Rev	ctagtcataccggtcgcAATGAGAAGAGGATTAAGAAAC
pEAQ_PvTPS5_GFP_For	attctgcccaaattcgATGGCAGCTGCAGGCATT
pEAQ_PvTPS5_GFP_Rev	ctagtcataccggtcgcGATGACAAGAGGATTGAGT

Fig. S1. Stereo view of homology models for each enzyme characterized in this study. Select positions are highlighted in red. PvHVS: blue; PvTPS2: cyan; PvTPS4: yellow; PvTPS5: tan.

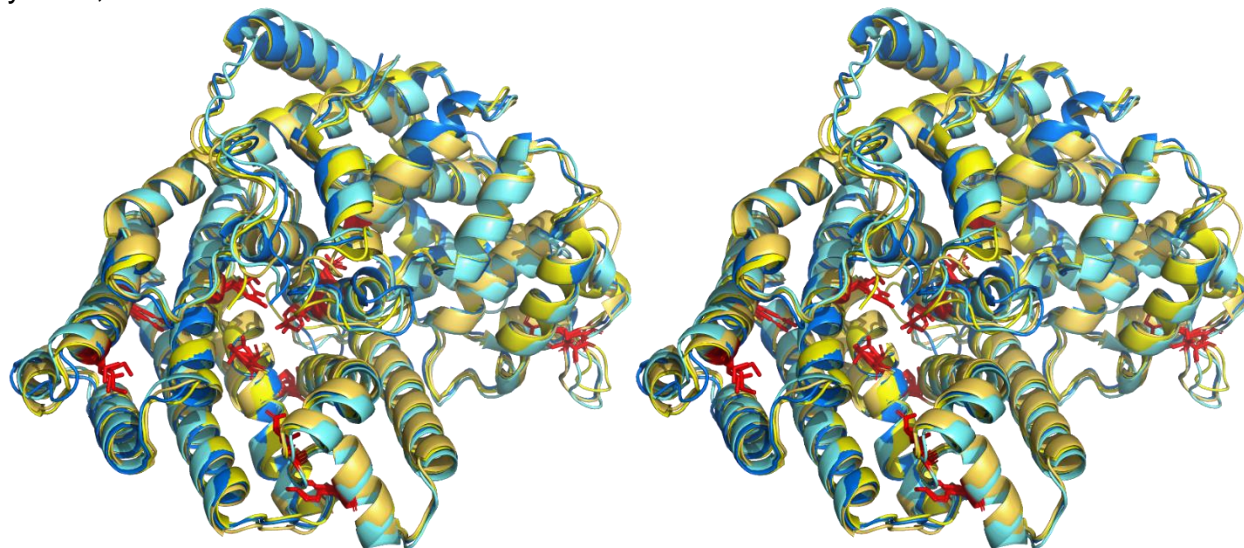


Fig. S2. Multiple sequence alignment of cytosolic and plastidial TPS-a enzymes found in Lamiaceae, Euphorbiaceae, Solanaceae, and Brassicaceae. Cytosolic and plastidial enzymes are above and below the central line, respectively, and enzymes from *P. vulgaris* characterized in this study are within the central box. Column highlights are explained in the main text. Accession numbers and families are indicated in the first line of the alignment. Abbreviations: RcAFS, *Ricinus communis* alpha-farnesene synthase; NtEAS, *Nicotiana tabacum* epi-aristolochene synthase; SlGDS, *Solanum lycopersicum* germacrene D synthase; AtTPS21/20, *Arabidopsis thaliana* terpene synthase 21/20; PvTPS2/4/5, *P. vulgaris* terpene synthase 2/4/5; PvHVS, *P. vulgaris* hydroxy vulgarisane synthase; RcCAS3, *Ricinus communis* casbene synthase; RcCAS2, *Ricinus communis* neocembrene synthase; NsCBTS3, *Nicotiana sylvestris* cembratrienol synthase 3.

AEQ27768.1	RcAFS	-----MKVGPVLCQCQTN-----	14	← Euphorbiaceae
5eau	NtEAS	-----	0	← Solanaceae
XP_019070211.1	SlGDS	-----MV-----VGF	5	← Solanaceae
AAO85539.1	AtTPS21	-----	0	← Brassicaceae
AZB50512.1	PvTPS4	-----MG-	2	← Lamiaceae
AZB50513.1	PvTPS5	-----MAA-AG-----IVS	8	← Lamiaceae Cytosolic
AZB50511.1	PvHVS	MSSLSIPFSSAICTSSIPKISTGHHRR-----ARMPAHDTSRLV-FRPSAV-----MVE	49	← Lamiaceae Plastidial
AZB50510.1	PvTPS2	MSSLTNLYSSIMTISTTKKVAVAHRLRTFNHGHRTRRRTRRLV-FRPSAV-----MVD	54	← Lamiaceae
XM_002513297	RcCAS3	MALPPVAM-----QSNSEKLEK-----YFFYRLPSTRLEYGNRRFH-FLPSSA-----KP-	43	← Euphorbiaceae
XM_002513288	RcCAS2	MALQSLFL-----QANSQNRNF-----CQFLSMPS--IRCCSCRVP-FSSWSA-----KS-	43	← Euphorbiaceae
ADI87448.1	NsCBTS3	-----MS-----QSI-SPLIC-----SHFAKFSQNIWRNCSQLR-VIHSSY-----ASF	38	← Solanaceae
AOA178U9Y5.1	AtTPS20	MEA-----ITK-----NGSLSQLTVLHCGPKLSLSSIPVRCRLRFKSNPF	38	← Brassicaceae
	RcAFS	-----EAFGMMQERRSGNYK-PNIWKYDFLQSLSSKYDE-EKYKTAERLKE	59	
	NtEAS	-----MASAAVANYEEEEIVRPVADFS-PSLWGDQFLSFSIDNQVA-EKYAKEIEALKE	51	
	SlGDS	GGEAE-----ITRRCANHH-PSVWGDHFLTYANLLGAN-EWEEKEHEDQKG	49	
	AtTPS21	-----MGSEVNRPLADFP-ANIWEDPLTSFSSKSDLGT-ETFKEKHSTLKE	43	
	PvTPS4	-----ESSVSCLRNSRPPVTDYV-PSIWADTFSTSSFKEQEQ-QKYEEAIEELKK	51	
	PvTPS5	-----AAVPTNIRQLRPPILKYK-PSMWDGTFATFSLDHQLQ-EKYSKEIETLKK	56	
	PvHVS	G-----SPMTTSSNGKEVQRLLITTFK-PSMWDGTFSTFSDNQVQ-EKYLKEIEELKK	100	
	PvTPS2	H-----AT-----EVRQVKTPYDPVSMWGMNFSTFSLDHQLQ-QKYSEIEELKK	98	
	RcCAS3	---HAKPPAQACLSSTTHQEVRLAHFP-PTVWGNRFASLTFNPSEF-ESYDGRVNVLKE	98	
	RcCAS2	---VTNKSPQACLSSTKQQEFRPLANFP-PTVWGSFASPTFSESEF-GTYDRQANVLQK	98	
	NsCBTS3	GGRRKERVRRMNRAMDLSSSSRHLADFP-STIWGDHFLSYNSEITEITTQEKNEHELKE	97	
	AtTPS20	PKKLVVTRARTSINSDEHAANRPLFQFP-PSLDDRFSLISANQSEI-DSLGRDIEALKA	96	
	RcAFS	DAKHLFIEA-VDLQKLELVDCI I KVLASHFKDEIKKALDTIASSI--KNDKSDAIKNR	116	
	NtEAS	QTRNMLLATGMKLADTLNLDITIERLGISYHFEKEIDDILDQIYNQN--S-----NCNDL	104	
	SlGDS	GVRKMLVLSPSKSLQKLELINTIQLLGVSYHFEHEIEESLSEIYNGY--E-EWIGESHDL	106	
	AtTPS21	AVKEAFMSSKANPIENIKFIDALCRLGVSCHEFKDIVEQLDKSFDCLDFPQMVROEGCDL	103	
	PvTPS4	EARCILLTAAASPRKQMLIDTLERLGLAYHFETEIEHILQQINQ-----QTLQDYDL	104	
	PvTPS5	EVKSMLMAA--TSTQLMVLIDKLERLGLSYHFEAEIEDKQVYDL----EEEVADHDL	109	
	PvHVS	EVRSTLMSA--THRKLFDLIDNLERMGIAYHFETEIEDKQAHASL---EEE-DDYDL	153	
	PvTPS2	NVRSTLMAA--TSTRLLMFIDNLERLGLAYHFETEIEDKQVYDLSL---EKEGEDNDL	152	
	RcCAS3	KVKDLLVSSATDSVETVVLIDLLCRLGVSYHFENDIEELLSKIFNSQ--PDLVDEKECDL	156	
	RcCAS2	KIRELLTSSRSDSVEKIAFIDLCLLGVSYHFENDIEEQLSQIFSCQ--PGLLDEKQYDL	156	
	NsCBTS3	IVRKMLVETLDNSTQKVLIDTIQRLGLAYHFENDEIENSIQNIYFYS--QNSDNDENHL	155	
	AtTPS20	KVSEKLVCM--DVKERIHLIHLVSLGVAYHFQIEEFLKVDENV--E-DMNLGEEDM	151	
	RcAFS	YVTALCFRLLRQHGYEVSQDVFSDFLDENGTFLKA---KSMVKGVLLEFEASYLALESE	173	
	NtEAS	CTSAIQFRLLRQHGFINISPEIFSKFDENGKFKES---LASDVLGLLNLYEASHVRTHAD	161	
	SlGDS	HDVALSFRLLRQGGYVSSDVFRKFTDERGNYKET---LASDVQGLLSLYEAAQLRVDD	163	
	AtTPS21	YTVGIIQVFRQFGFKLSADVFEKFKDENGKFKGH---LVTDAYGMLSLYEAAQWGTGHE	160	
	PvTPS4	FTTTLGFRLLRQHRHVSVCVDFDKFLDQDAKFKESLL--LSSDTESVLSLYDAAHVFRHE	163	
	PvTPS5	FTTALRFRLLRQHGYHVSVDVDFKFKDQKDWLEIISLSRDDIEGVLSLYEAAHVRIRE	169	
	PvHVS	FTTALRFRLLRQHRYHVSVDVDFKFKDQKDKES---LSSDVEGLLSLFEASHLRIHNE	210	
	PvTPS2	FTTALRFRLLRQHRYQVSCDVFQYFEEGKRVKES---HCSVDVEGLLSLYEAAVVGIPNE	209	
	RcCAS3	YTAALVFRVFRQHGFKMSSDVFSKFKDSGKFKES---LRGDARGMLSLFEASHLSVHGE	213	
	RcCAS2	YTVALVFRVFRQHGFKMSSNVFHKFTDSDHGKFKAS---LLSDARGMLSLFEASHLSMHGE	213	
	NsCBTS3	YVAALRFRLLRQGGYQNMSSDVFKQFTNHDGKFKEN---HTNDVQGLLSLYEAAHMRVHDE	212	
	AtTPS20	YSISVIFRVRFRYRHLKSSDVFNRFKEENGDFKCC---LDDVRGMLSFYEASYFGTNT	208	

RcAF5 NILDDAKAFSTTILKDISA--TTESNLYKQVVHALELPPHWRVWFDFVKWHIK-TFQKDK 231
NtEAS DILEDALAFSTIHLESAAP--HLKSPLREQVTHALEQCLHKGVPRVETRFIFISSIYDKEQ 219
SlGDS ILDEAINFTTHLKLPL--NLSDPATQVSNALKFFINNIIVRVATRKYIS-FYQEDK 220
AtTPS21 DIIDEALAFSRSHLEEISS--RSSPHLAIRIKNALKHPYHKGISRIETROYISYEEES 218
PvTPS4 NLLKEAAVSTKQYLRLGIEAE--LLDSSLKEKVNRAKHPHLDVPIFYARFFIS-IYEKDL 221
PvTPS5 KVLDEAAFTIHRKHLVLPQ-LESA-IKEKVELALQHPHKSPLPLNFRSYIN-IYEGEG 226

PvHVS DVLDEAIVFTTHLNRMMPO-LESP-LKEEVKHALRYPLHKCLGILSLRFHID-RYENDK 267
PvTPS2 KILDEAMTITTRQLTLMPLR-LESTILKEKVEHGLKFPVQRTNITFSLRYVE-IYENDE 267
RcCAS3 DILEEAFATKDYLOQSSAVE--LFPNLKRHTNALEQPFHSGVPRLEARKFIDLYEADIE 271
RcCAS2 DILEEAFATKDYLESSAVDQYLCPNLQKHITNALEQPFHKGIPRLEARKYIDLYEGD-E 272
NsCBTS3 EILEEALIFTTHLESVIP--NLSNSLKVQVTEALSHPIRKTIIPRVGARKYIY-IYENIG 269
AtTPS20 EILDEAMGFTRKHLELFVGG-SNEEHLSGHIKNVLYLSQQENAIVVMSREYIQ-FYEQET 266

RcAF5 SINKTLLDLAKVNFVQATLQNDLKEISRWRNGLIENLK-FSRDRLVESFLCTVGLV 290
NtEAS SKNNVLRFAKLDNFNLQMLHKQELAQVSRWKKDLDFVTTLP-YARDRVVECYFWALGVY 278
SlGDS SHNEMLLHFAKLDNFNLQRLHKRELCDITRWKKDSEIAQALP-FARDRVVELYFWSLGVY 279
AtTPS21 C-DPTLLEFAKIDFNLLQILHREELACVTRWHHEMFKSKVT-YTRHRIETAYLWSLGT 276
PvTPS4 SRNELLAKLNFNFQNLKYKEELFQLTGWNKFDVKSCLT-YARDRLVEAYLWGVAYH 280
PvTPS5 LTNESVVKLSKFNFNFLQNIYRKELBELTSWNNKYELSKLP-YARDRLVECYIWGAALR 285
PvHVS SRDEVVLRGQVNFNYMNIYMNELYEITTWNNKLMQTSKVP-YFRDRLVECYMWGLAYH 326
PvTPS2 SRDALVLRKALNFNFQNIYRSELSEATAWNNYENLASEVP-YIRDRMVECYVSGVTYR 326
RcCAS3 CRNETLLEFAKLDYNRVQLLHQELCQFSKWKDLNLASDIP-YARDRMAEIFFWAVAMY 330
RcCAS2 CRNETVLEFAKLDYNRVQLLHQELSQFSQFSTWKKDLNLASEIP-YARDRMAEIFFWAVAMY 331
NsCBTS3 THNDLLKFAKLDNFNLQKHLKELNELTSWKKDLDCANKEP-YAKDRLVEAYFWTVGIY 328
AtTPS20 HHDETLLKFAKINFKFMQLHYVQELQITVKKWKELDLESKIPNYRVRVAVCELYWAMAVY 326

RcAF5 FEPQYSSFRWLTKVIMILVIDDVYDIYGSLEELQHTNAINRWDTALEQ--LP-EYM 347
NtEAS FEPQYSQARVMLVKTISMISIVDDTFDAYGTVKELEAYTDAIQRWDINEIDR--LP-DYM 335
SlGDS FEPHYSVARKILTKVLCFCSIMDDTYDTYGTLDLTLTAAIERWDIDASEQ--LP-SYM 336
AtTPS21 FEPQYSQARVITMALILFTALDDMYDAYGTMELELELFTDAMDEWLPVVPDEIPIP-DSM 335
PvTPS4 YEPQYVNRIGLVKGIQITGIMDDTYDNYATLNEAQIFQTLDRWNANEVDG--LP-DYM 337
PvTPS5 YEPQFAYLRAIVAKTMQVLSIMDDTYDNYGTLEEDDLTDLLEKWNLEIDV--LP-DFM 342
PvHVS FEPEYAPVRVILITKYMTATTVDVDDTYDNYATLEIELELFTQAI DRWSEDEIDQ--LPDEYL 384
PvTPS2 FEPQYSQIRMMVTKYVILGSLVDDTYDYATPOEIDLMTQAMESWNTQETEK--LP-EYL 383
RcCAS3 FEPDYAHTRIIAKVVLISLIDDTIDAYATMEETHILAEAVARWDMSCLEK--LP-DYM 387
RcCAS2 FEPKYAQARMIIAKVVLISLVDVDFDAYATIEETHLLEAFERWDKSCLDQ--LP-DYM 388
NsCBTS3 FEPQYSRSRSMITKVVKMNSIIDDTYDAYATFDELVLFTDAIQRWDEGAMDS--LP-TYL 385
AtTPS20 MEPQYSVARIILSKSLVLTWITIDDLVYDAYCTLPEAIAFTENMERWETDAKDM---P-DHM 382

RcAF5 KICFKTLHTITGETAHEMQREKRWQEQETHLKKVWADFQALFVEAKWFNKGYTPSVQ 407
NtEAS KISYKAILDLKDYEKELSSAGR---SHIVCHAIERMKEVVRNRYNVSTWFIEGYTPPV 392
SlGDS KLSYRALVQVNETEKELENLGNKM-TYRVKYSINEMKLLRAYFQEAkWYHGKDVPTME 395
AtTPS21 KFIYNVTVVEFYDKLDEELEKEGR---SGCGFHLKKSLOKTANGYMQEAkWLLKDYIATFD 392
PvTPS4 KIVYDFILSIYEDYKCSASKHEK---GFSVPYFKEAVQELSSAYNQLEWVTERNMPSFI 394
PvTPS5 KIVYRFIMSVYEDFEREAEOQK---SFAAPYSESLEKLSRAVYKQKQWIMERKMPFFE 399
PvHVS KIVYKGLMNFTEEFRRDAEERGG---GYVVPYFIEETKRATQGYANEQRWIMKREMPSE 441
PvTPS2 KLVLKAFMRFTTEFACEAEKQK---SYVVPYVIEQMKVEVGRRYAEAQEWTRAQESPTFO 440
RcCAS3 KVIYKLLNTESEFEKELTAEKG---SYSVKYGREAFQELVIRGYYLEAVWRDEGKIPSF 444
RcCAS2 KVIYKLLNTESEFENDLAKEG---SYSVRYGREAFQELVIRGYYLEAMWRDEGKIPSF 445
NsCBTS3 RSIYQGLLDVFNEMEEVLAKEG---ADRIYAKKEMKLVAAVYFKEAQWLNANYIPKCE 442
AtTPS20 KLMRSFIDLHEDFKREVILEGR---LYSVEYIGDECKRLFPREDLKLKSWARTGYIPNDY 439

RcAF5 EYLKTACISSGSLLSVHSFFLIMNEGTRMLHFLEKNOE-MFYNISLIIRLCNDLGTSV 466
NtEAS EYLSNALATTTYYLATTSYLGMKS-ATEQDFEWSKNPK-ILEASVICRVIDDTATYE 450
SlGDS QYIKNGIPSSYLLLATTSWLGMDVATKDAFDWISNET-ILVASSIARLLNDLVTHE 454
AtTPS21 EYKENAILSSCYGLIAMTFVRMTDVAKLDAFEWLNHPK-IRVASEIISRTDDISSYE 451
PvTPS4 EYARNSEITSSLYMMFAVIPGLKS-LTPPETIDWVKTEPL-ITVSTAMIGRYMDDIGSQH 452
PvTPS5 EYMTNSVFTSCYIPMFIAFVPGMKS-VTEKEVQWLLSEPK-IVISTAKMGRFTLEDLGSHE 457
PvHVS EYMNRSRVTSIMYVTVYVAVVAVIES-ATKETVDWALSDD-IFVYTNDIGRLIDDLATHR 499
PvTPS2 EYLANGRVTVLHYVLFPIIATFTKA-ATEETIHLLDDSSKLVTVSADINRLDLDLATHQ 499
RcCAS3 DYLYNGSMTTGLPLVSTASFMGVQEITGLNEFQWLETNPK-LSYASGAFIRLVNDLTSV 503
RcCAS2 EYIRNGSLSSGLPLVSTASFMGVQEITGIREFQWLRTPK-LNHFSGAVGRIMNDIMSHV 504
NsCBTS3 EYMKNGVVTSTGTMYGIISLVVMEIITKEAFEWLANEPL-ILRAASTICRLMDDMADHE 501
AtTPS20 EYMEVGIVTAGIDMTVAFAFIGMGE-AGKEAFDWIRSRPK-FIQTLDIKGRLRDDVATYK 497

RcAF5 AEQERGAASSIVCHMREMEVLEEEARSYLKGIIGNYKVKVNEKCFQSP--EMQLFINI 524
NtEAS VEKSRGQIATGIECCMRDYDISTKEAMAKFQNMABETAWKDINEGLLRPT--VSTFELTP 508
SlGDS IEVERGDVASGIECYMNEYGATKEEAYMEIRKIIENNWKDLNRCGLKPTT--VPRVLLMP 512
AtTPS21 FEHKREHVATGIDCYMQQFGVSKERAVEVMGNIVSDAWKDLNQEMLRPHV--FPFPLLMR 509
PvTPS4 RESKGGVLTAVDCYMKQYVNSKEETLSKFGELVEDAWEDLNKEWVETTS-FLPTEIAVQ 511
PvTPS5 ENRDGEMETVDDCYMKDKGVSKQRTIYEFQVLENGWKVDYTAEWAKENS--VCKEMVEH 515
PvHVS RERKDGIMLTSMDYMKYEGGTMEEGEAAFRKLMEEKWLLNAAWVDTINGKESKEIVVQ 559
PvTPS2 RESKNGVLTTPMDYMKYEGVSRQEAIEVTKLQETWKEINAWEVWVCKT--VPKEICND 557
RcCAS3 TEQQRGHVASCIDCYMNQHGVSKEAVKILQKMATDCKWKEINEECMRRSQ--VSVGHLMR 561
RcCAS2 SEQNRGHVASCIDCYMKQYVESKEEAIKEMQKMASDAWKDINEGYMRPAQ--VSVSELMR 562
NsCBTS3 VEQQREHVASFVECYMEEYGVSKQEAIVEMRKKITNACKDINKELLRFTA--VPMFLER 559
AtTPS20 DEMARGEIATGINDCYMKQYKVTBEEAFLEFHRRRIKHTSKLVNEEYFKTTV---PLKLVRI 554

RcAF5	NVNMARVVHNLV-QN-RDGFVQDQHNKKQILSLLVHFFKLD-----	564
NtEAS	ILNLRARIVEVTYIHN-LDGYTHPEKVLKPHIINLLVDSIKI-----	548
SlGDS	VLNLRVAEF-FYKD-EDAYTFSKNNLKDVISMVLIDFIKA-----	551
AtTPS21	VLNLSRVLDV-FYRY-QDAYTNPV-LLKEHIVSLLIETIPI-----	547
PvTPS4	FLNYARMCDACYNNSNDGDGYTDPVIFK-SNVVALFLNPLLI-----	551
PvTPS5	LLNYGRVIAEITYNNK-EDGYTEPEKYLGLAAALLLNPLVI-----	555
PvHVS	VLDLARICTLYGDE-EDGFTYPEKNFAPLVAALLMNPVHI-----	599
PvTPS2	LVEISRLSEIYYKSG-EDCYTNPEKNLAPIMAAVLLNPFYQV-----	597
RcCAS3	IVNLRARLTDV-SYKY-GDGYTDSQ-QLKQFVKGLFVDFISI-----	599
RcCAS2	VVNLRARLTDV-SYKY-GDGYTDPQ-HLKQFVKGLFIDFVPLPNQIRKGETKTKHV	614
NsCBTS3	TLNFIKRLVGT-FLKD-DDGYTNPKSQVKDSIALLFVESIDI-----	598
AtTPS20	AFNVGRAIDTNY-KH-GDGLTYGG-IVEGQITSLFLDLITI-----	592

Fig. S3. GC-MS of hexane extracts TPS activity assays *in vitro*, in *E. coli*, and in *N. benthamiana*. *In vitro*, purified enzymes were fed with pure precursors. For the *N. benthamiana* assays, DXS is overexpressed in all conditions. Chromatograms labeled *P. vulgaris* root are from methyl tert-butyl ether extracts of fresh root. Key shows the feature annotations. Annotations in black are based on comparisons to the Adams or NIST17 libraries. Annotations in blue are based on comparisons to authentic standards, or NMR of the purified compound.

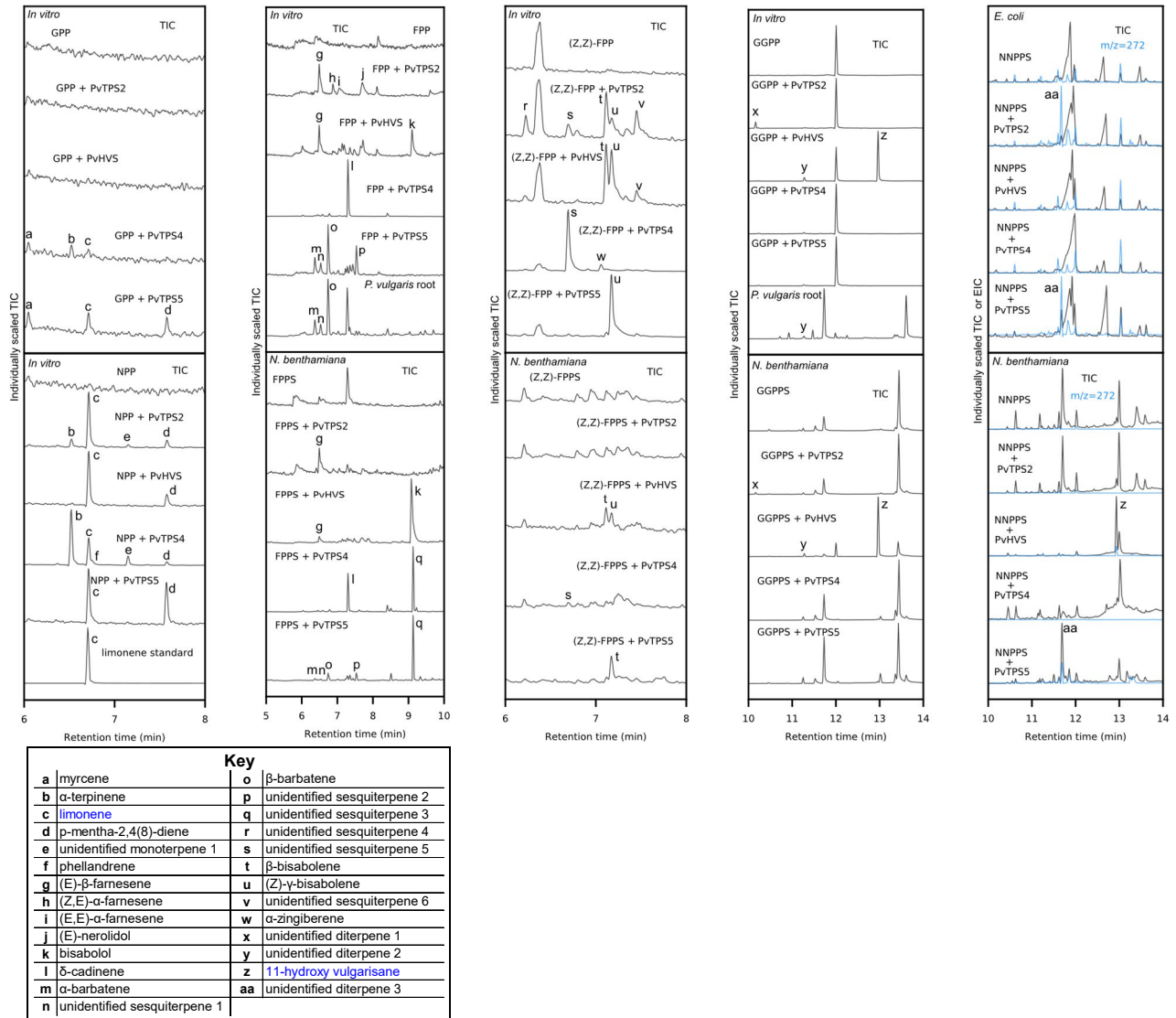


Fig. S4. Mass spectra of enzyme products and selected reference spectra.

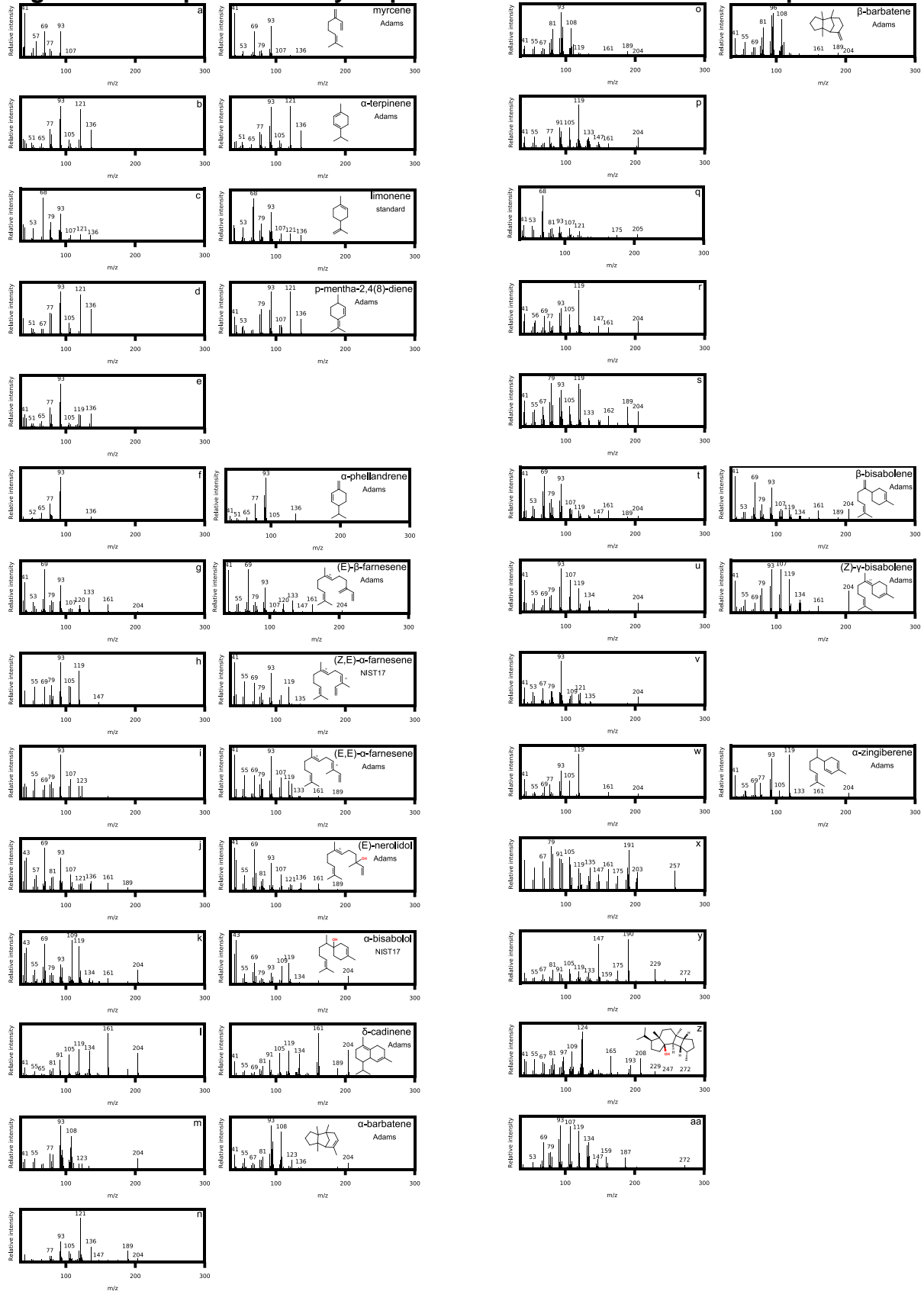


Fig. S5. Peak q only appears when PvTPS4 or PvTPS5 is co-expressed with FPPS in *N. benthamiana*.

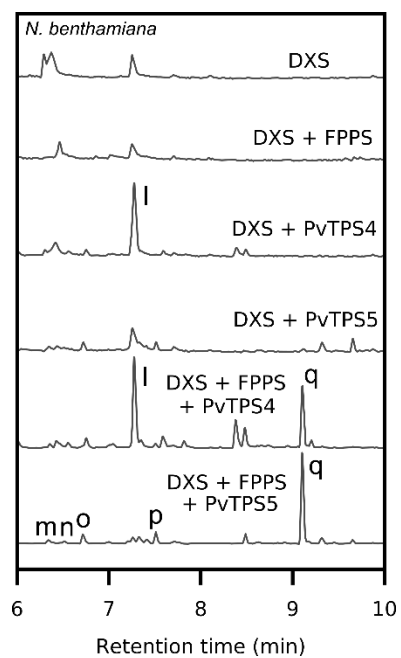
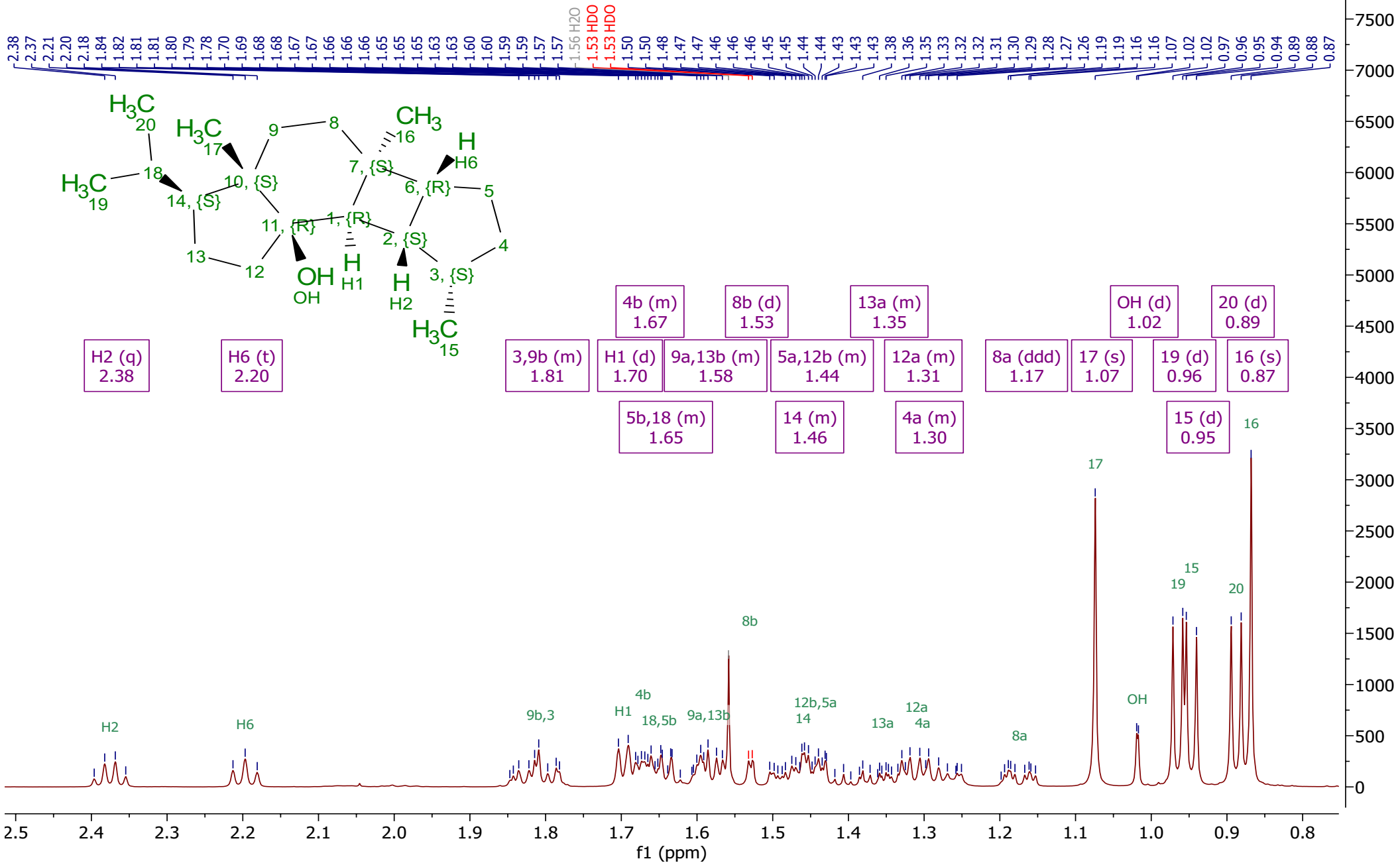


Fig. S6. NMR spectra of PvHVS major product from GGPP, identified as 11-hydroxy vulgarisane.

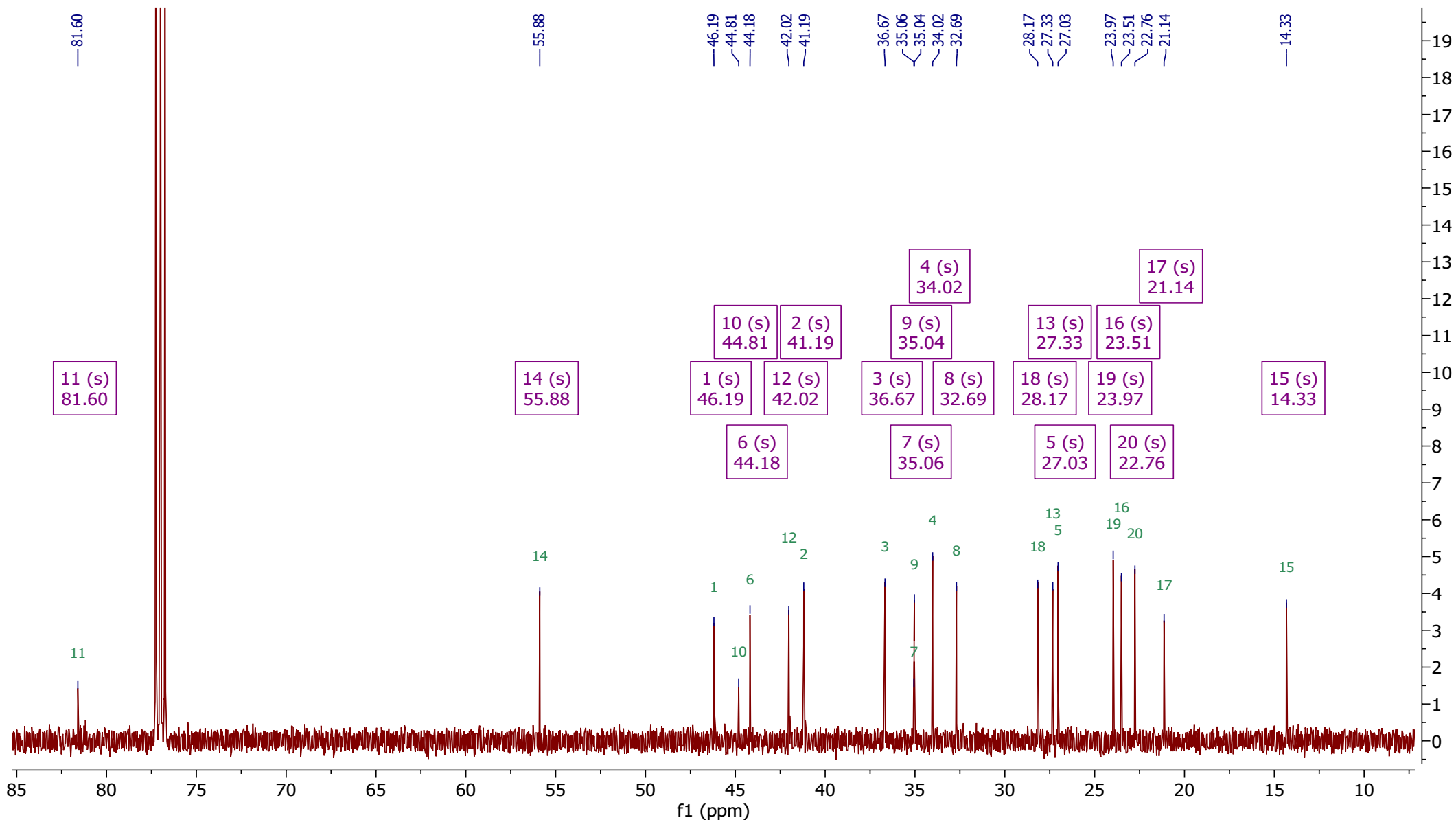
^1H NMR (500 MHz, cdcl_3) δ 0.87 (s, 3H, H-16), 0.89 (d, $J = 6.6$ Hz, 3H, H-20), 0.95 (d, $J = 6.7$ Hz, 3H, H-15), 0.96 (d, $J = 6.5$ Hz, 3H, H-19), 1.02 (d, $J = 1.2$ Hz, 1H, OH-11), 1.07 (s, 3H, H-17), 1.17 (ddd, $J = 13.6, 4.1, 2.8$ Hz, 1H, H-8a), 1.30 (m, 1H, H-4a), 1.31 (m, 1H, H-12a), 1.35 (m, 1H, H-13a), 1.44 (m, 2H, H-5a, H-12b), 1.46 (m, H-14), 1.53 (d, $J = 27.4$ Hz, 1H, H-8b), 1.58 (m, 2H, 9a, H-13b), 1.65 (m, 2H, H-5b, H-18), 1.67 (m, 1H, H-4b), 1.70 (d, $J = 6.4$ Hz, 1H, H-1), 1.81 (m, 2H, H-3, H-9b), 2.20 (t, $J = 8.0$ Hz, 1H, H-6), 2.38 (q, $J = 6.9$ Hz, 1H, H-2).

^{13}C NMR (126 MHz, cdcl_3) δ 14.33 (C-15), 21.14 (C-17), 22.76 (C-20), 23.51 (C-16), 23.97 (C-19), 27.03 (C-5), 27.33 (C-13), 28.17 (C-18), 32.69 (C-8), 34.02 (C-4), 35.04 (C-9), 35.06 (C-7), 36.67 (C-3), 41.19 (C-2), 42.02 (C-12), 44.18 (C-6), 44.81 (C-10), 46.19 (C-1), 55.88 (C-14), 81.60 (C-11).

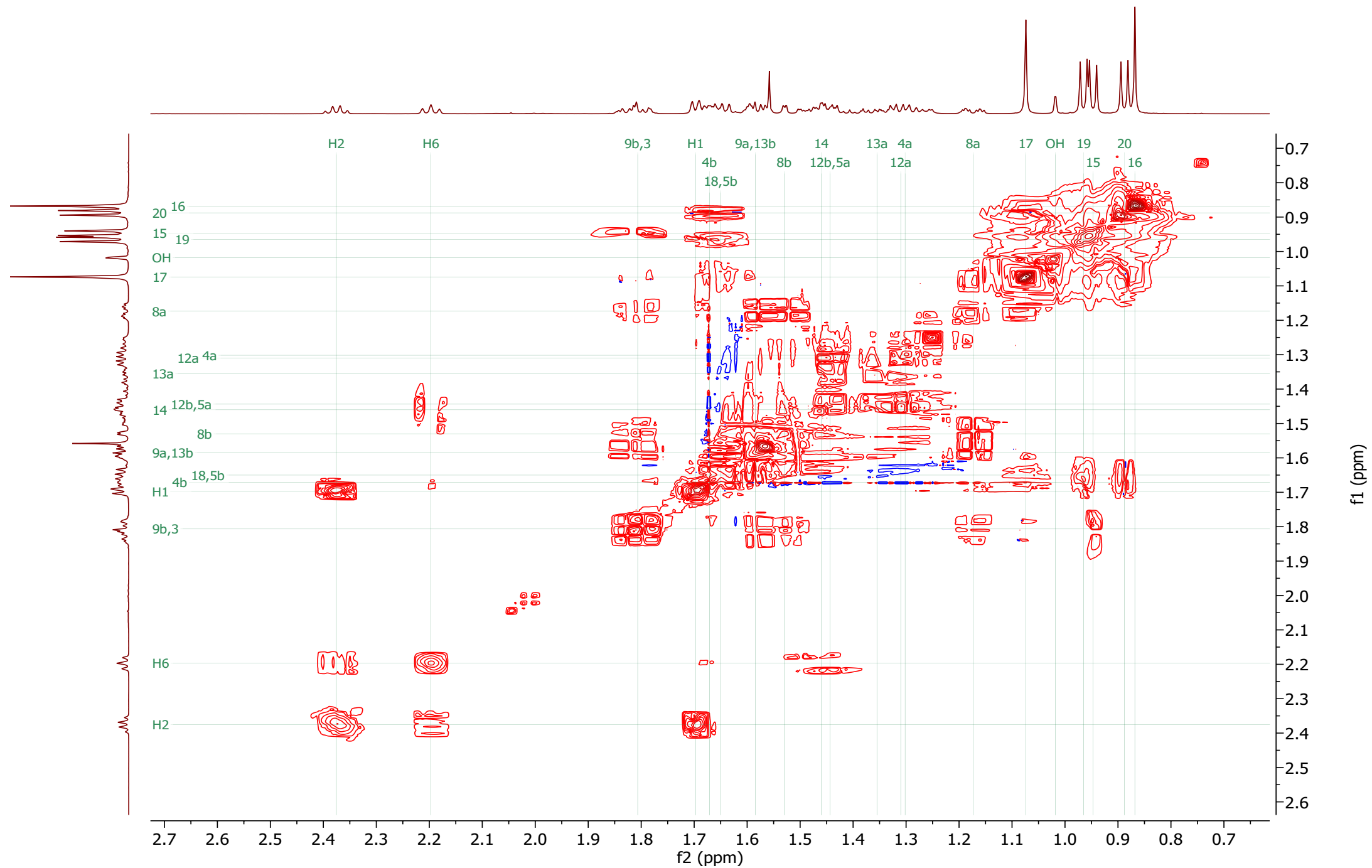
¹H-NMR



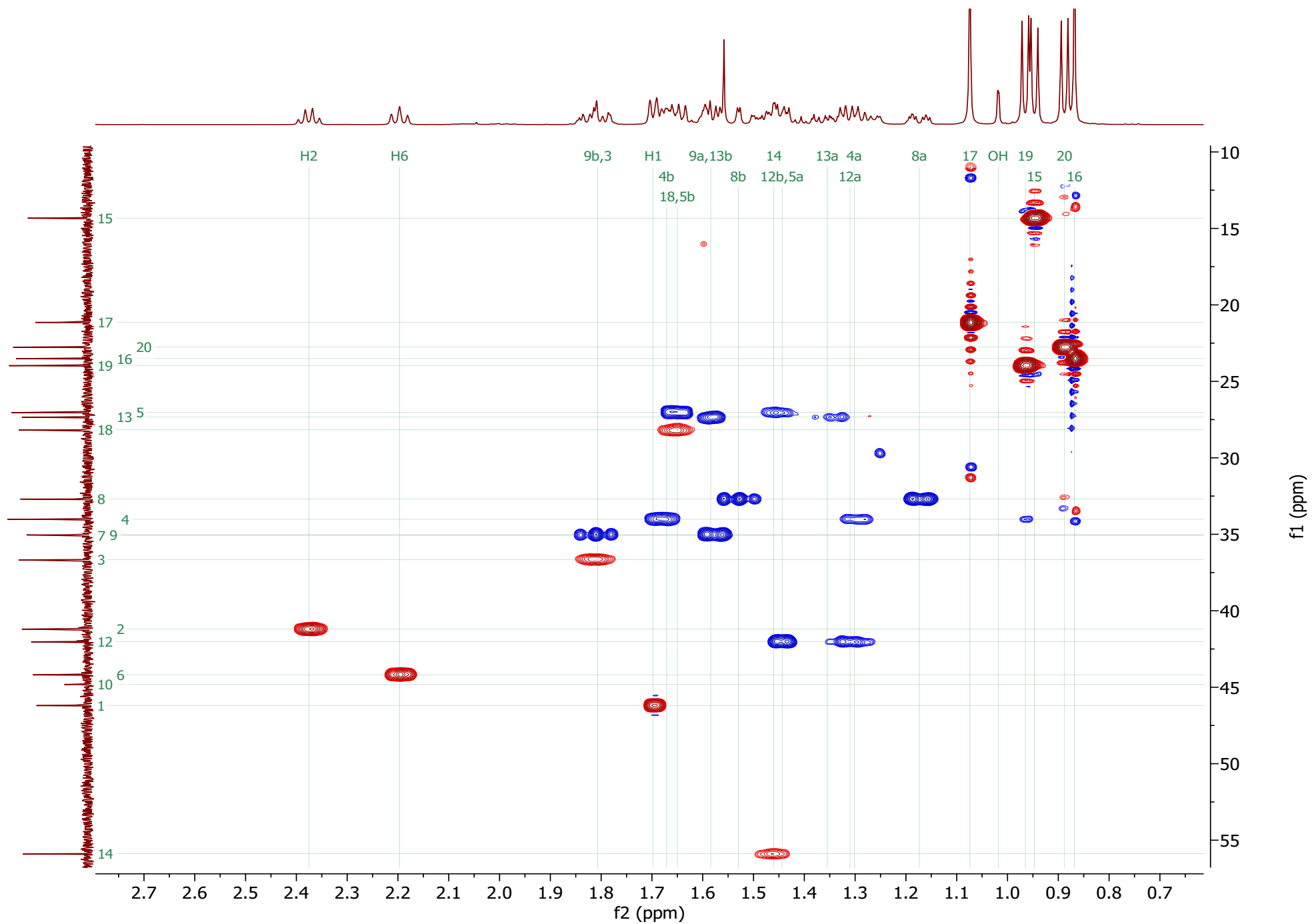
^{13}C -NMR



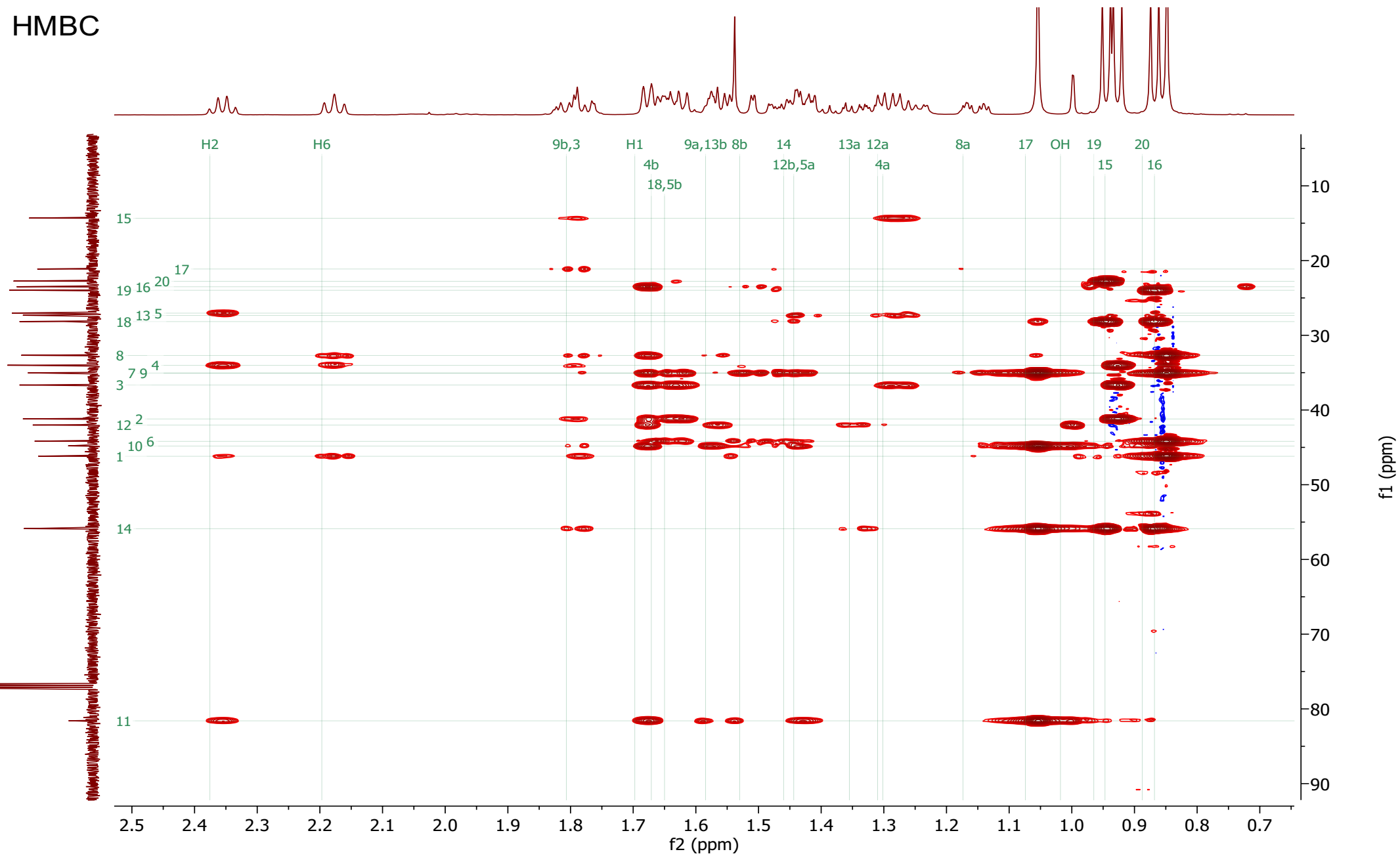
COSY



HSQC



HMBC



NOESY

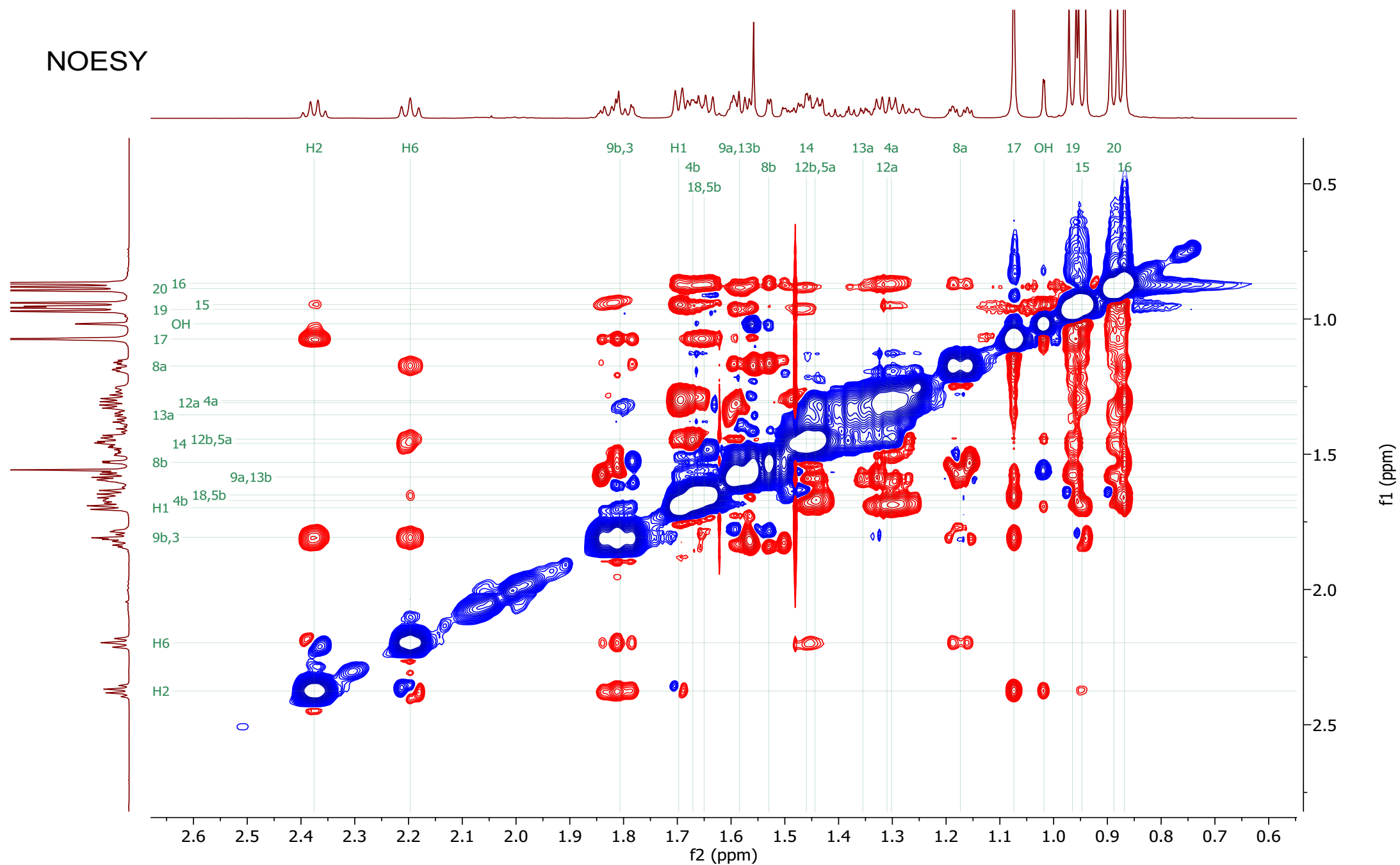


Fig. S7. 11-Hydroxyvulgarisane (z) in *P. vulgaris*. A 70% methanol extract of root tissue was subjected to non-targeted UHPLC/MS^E analysis using a 31-min chromatographic separation. (a) Extracted ion chromatogram for m/z 273.3 ($[M+H-H_2O]^+$) for the crude *P. vulgaris* extract and (b) extracted ion chromatogram for m/z 273.3 for the purified 11-hydroxy vulgarisane, (c) MS/MS spectrum of the purified 11-hydroxyvulgarisane. The identity of purified 11-hydroxyvulgarisane was confirmed by NMR.

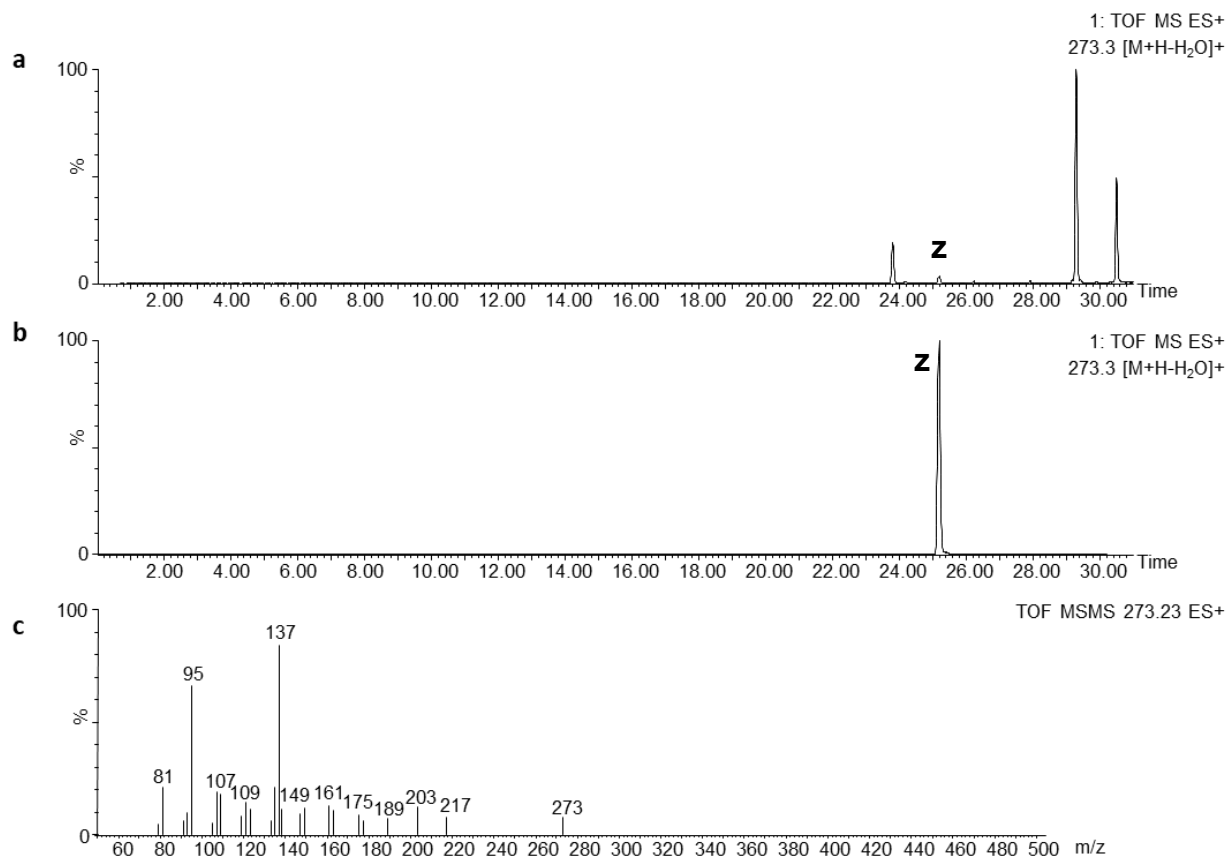
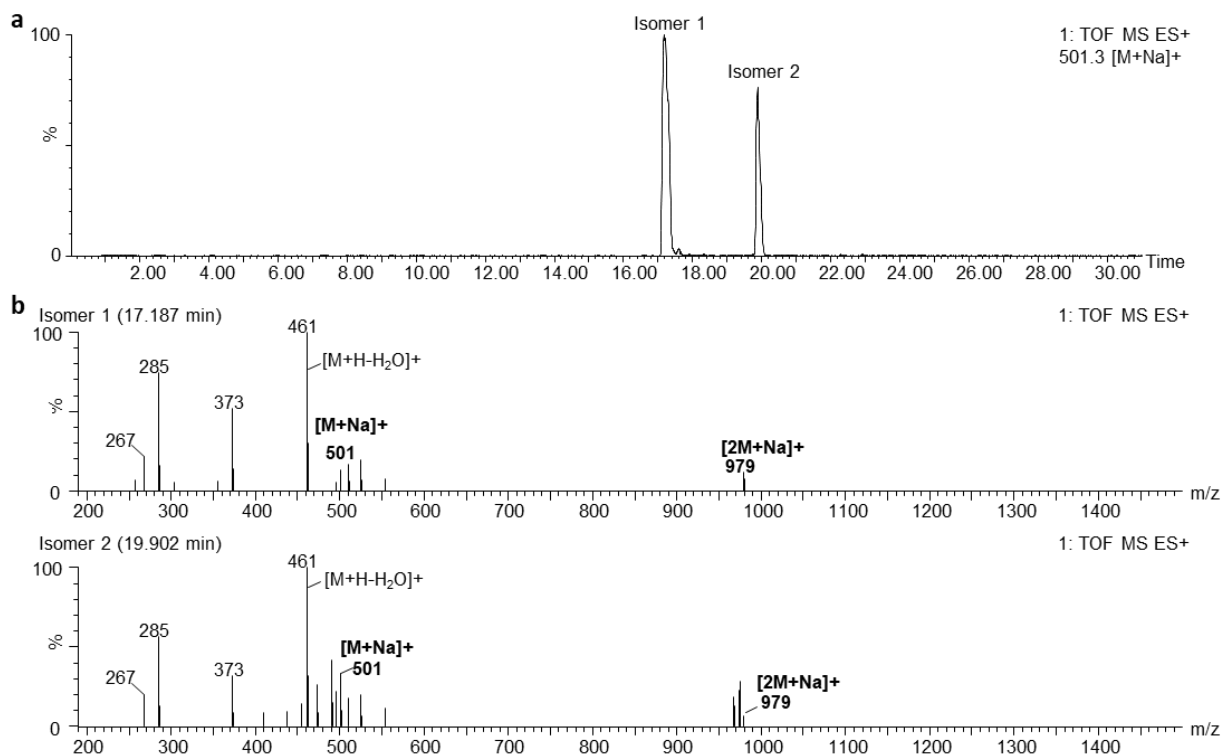


Table S2. Exact masses and relative abundance for ions of vulgarisin-related compounds detected by UHPLC/MS^E in *P. vulgaris* root extracts.

Annotated Metabolite	Formula	Retention time [min]	Calculated m/z	Experimental m/z	Precursor and fragment ion(s) m/z and relative ion abundances (%)
11-Hydroxy vulgarisane (Fig. S7 c)	C ₂₀ H ₃₄ O	25.19	273.2582 [M+H-H ₂ O] ⁺	273.2585 [M+H-H ₂ O] ⁺	273.26 (10), 217.20 (10), 203.18 (15), 189.16 (9), 137.13 (100), 107.08 (24), 95.08 (85)
Vulgarisin A isomer 1 (Fig. S8 b, top)	C ₂₈ H ₄₆ O ₆	17.17	501.3192 [M+Na] ⁺	501.3191 [M+Na] ⁺	501.32 (15), 461.33 (100), 373.27 (51), 285.22 (74), 267.21 (21)
Vulgarisin A isomer 2 (Fig. S8 b, bottom)	C ₂₈ H ₄₆ O ₆	19.88	501.3192 [M+Na] ⁺	501.3184 [M+Na] ⁺	501.32 (34), 461.33 (100), 373.27 (33), 285.22 (58), 267.21 (20)

Fig. S8. Vulgarisin A and D isomers in *P. vulgaris*. A 70% methanol extract of root tissue was subjected to non-targeted UHPLC/MS^E analysis using a 31-min chromatographic separation and electrospray ionization in positive-ion mode. (a) Extracted ion chromatogram for m/z 501.3 ($[M+Na]^+$) of the *P. vulgaris* extract is shown in function 1 (b). The mass spectra of the two peaks are shown, displaying prominent $[M+H-H_2O]^+$ ions in each case at m/z 461 and subsequent loss of isobutyric acid (88 Da) at m/z 373. Note that the mass spectra of the two isomers are consistent with the previously reported spectra for vulgarisin A (Lou *et al.*, 2014) and vulgarisin D (Lou *et al.*, 2017).



Additional references:

Fu S, Ma Y, Yao H, Xu Z, Chen S, Song J, Au KF, Berger B. 2018. IDP-denovo: de novo transcriptome assembly and isoform annotation by hybrid sequencing. *Bioinformatics* **34**: 2168–2176.

Hamberger B, Ohnishi T, Hamberger B, Séguin A, Bohlmann J. 2011. Evolution of diterpene metabolism: Sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects. *Plant Physiology* **157**: 1677–1695.

Hoang NV, Furtado A, Mason PJ, Marquardt A, Kasirajan L, Thirugnanasambandam PP, Botha FC, Henry RJ. 2017. A survey of the complex transcriptome from the highly polyploid sugarcane genome using full-length isoform sequencing and de novo assembly from short read sequencing. *BMC Genomics* **18**.

Lou H-Y, Jin L, Huang T, Wang D-P, Liang G-Y, Pan W-D. 2017. Vulgarisins B–D, three novel diterpenoids with a rare skeleton isolated from *Prunella vulgaris* Linn. *Tetrahedron Letters* **58**: 401–404.

Lou H, Zheng S, Li T, Zhang J, Fei Y, Hao X, Liang G, Pan W. 2014. Vulgarisin A, a new diterpenoid with a rare 5/6/4/5 ring skeleton from the Chinese medicinal plant *Prunella vulgaris*. *Organic Letters* **16**: 2696–2699.

Salmela L, Rivals E. 2014. LoRDEC: accurate and efficient long read error correction. *Bioinformatics* **30**: 3506–3514.