

ELUCIDATION AND REPURPOSING OF PLANT DITERPENOID BIOSYNTHETIC
PATHWAYS

By

Garret P. Miller

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biochemistry and Molecular Biology—Doctor of Philosophy

2022

ABSTRACT

ELUCIDATION AND REPURPOSING OF PLANT DITERPENOID BIOSYNTHETIC PATHWAYS

By

Garret P. Miller

Terpenoids are the largest class of specialized metabolites in plants, with widespread uses ranging from fragrances and cosmetics to biofuels, antifeedants, and pharmaceuticals. Terpenoids are derived from a small set of prenyl diphosphate substrates which are cyclized into different terpene scaffolds by terpene synthases. These scaffolds are then modified by various tailoring enzymes—typically starting with cytochrome P450s—into functionalized terpenoids. Given the structural complexity of many of these metabolites, total chemical synthesis is often challenging to achieve at a relevant scale and cost, and as such, biosynthetic methods are increasingly being employed as an alternative for their production. The work presented in this dissertation describes the elucidation of two terpenoid biosynthetic pathways and the repurposing of known pathways to convert synthetic substrates not found in nature. First, three steps constituting the full biosynthetic pathway to leubethanol, an antimicrobial diterpenoid active against multidrug-resistant TB, was elucidated in the Texas Sage (*Leucophyllum frutescens*). Second, seven steps in the biosynthetic pathway towards structurally complex diterpenoid alkaloids were elucidated in the Siberian Larkspur (*Delphinium grandiflorum*). Third, twenty-four terpene synthases were screened for activity against twenty synthetic substrate analogs not found in nature, resulting in fifty-six new products and demonstrating the ability to derivatize terpene scaffolds through the derivatization of a starting substrate. In all, this work expands access to different classes of terpenoids through the elucidation of biosynthetic pathways and semi-biosynthesis of terpene scaffolds not found in nature, allowing for more feasible and sustainable production of these structurally complex compounds.

TABLE OF CONTENTS

LIST OF TABLES	v
LIST OF FIGURES	vi
LIST OF SCHEMES.....	x
KEY TO ABBREVIATIONS.....	xi
CHAPTER 1 Plant Terpenoid Biosynthesis and Elucidation of Biosynthetic Pathways	1
Specialized Metabolism	2
Terpenoids	3
Evolutionary Considerations in Specialized Metabolism	5
Signatures of Biosynthetic Pathways and Strategies to Find Them.....	7
Biochemical Considerations in Elucidating Biosynthetic Pathways.....	10
Work Presented in this Dissertation	12
REFERENCES	13
CHAPTER 2 Elucidating the Biosynthetic Pathway to the Antimicrobial Diterpenoid Leubethanol in <i>Leucophyllum frutescens</i>	22
Abstract	23
Significance Statement	24
Introduction	24
Results	27
Accumulation of leubethanol guided tissue-specific RNA sequencing.	27
Identification of TPS candidates from <i>L. frutescens</i>	28
<i>Lj</i> TPS1 exclusively cyclizes neryl neryl diphosphate into the serrulatane backbone.	30
<i>Lj</i> CPT1, a short chain <i>cis</i> -prenyl transferase, supplies NNPP in serrulatane biosynthesis. ..	32
A cytochrome P450 converts the serrulatane backbone to leubethanol.	33
Discussion	37
Materials and Methods	41
Plant material, RNA Isolation and cDNA synthesis, and metabolite analysis	41
<i>L. frutescens</i> and <i>E. serrulata</i> de novo transcriptome assembly and analysis.....	41
Cloning and sources of genes used.....	42
In vitro assays	43
Transient expression in <i>N. benthamiana</i>	44
<i>E. coli</i> in vivo assays	44
Dihydroserrulatene production scale-up and NMR.....	45
GC-MS.....	45
Homology Modeling.....	46
Data Availability	46
Acknowledgments	46
APPENDIX.....	48
REFERENCES	67

CHAPTER 3 Identifying Entry Steps in the Biosynthetic Pathway to Diterpenoid Alkaloids in <i>Delphinium grandiflorum</i>	75
Abstract	76
Introduction	77
Results	79
Proposal of an Initial Biosynthetic Pathway.....	79
RNA Sequencing and Transcriptome Assembly	81
A Pair of TPSs Cyclizes GGPP to <i>ent</i> -atiserene.....	81
Two Pairs of Cytochrome P450s With Overlapping Functions Oxidize <i>ent</i> -atiserene	84
Continuation of the Previously Proposed Biosynthetic Pathway	89
Coexpression Analysis Reveals that a Predicted Reductase is Active in the Pathway	92
Discussion	94
Materials and Methods	98
Plant material, RNA isolation, and cDNA synthesis.....	98
<i>D. grandiflorum</i> and <i>Aconitum</i> genera de novo transcriptome assembly and analysis.....	99
Coexpression analysis.....	100
Cloning	100
Transient expression in <i>N. benthamiana</i> , product scale-up, and NMR analysis	100
GC-MS analysis.....	101
LC-MS analysis	102
APPENDIX.....	103
REFERENCES	120
CHAPTER 4 Repurposing Terpene Synthases for the Conversion of Synthetic Geranylgeranyl Diphosphate Derivatives	126
Abstract	127
Introduction	128
Results and Discussion.....	130
TPSs Utilized in this Study and Screening Process.....	130
Screening of Class II diTPSs	130
Screening of Class II/Class I Combinations	135
Screening of Single-Step class I TPSs.....	137
General Trends	139
Future Perspectives.....	141
Materials and Methods	144
Cloning and Sources of Genes Used	144
Phylogenetic Tree	144
Enzyme Expression and Purification.....	144
In vitro assays	145
GC-FID/MS analysis	146
Scaleup and NMR.....	146
APPENDIX.....	148
REFERENCES	208

LIST OF TABLES

Table 2.S1: ^{13}C and ^1H chemical shifts for NMR spectra of dihydroserulatene	53
Table 3.S1: ^1H and ^{13}C chemical shifts for <i>ent</i> -atiserene	104
Table 4.1: List of enzymes used in this study	131
Table 4.S1: ^1H and ^{13}C chemical shifts for (+)-14-methylcopal-8,13-ol (24)	149
Table 4.S2: ^1H and ^{13}C chemical shifts for (+)-14-fluorocopal-8,13-ol (31)	153
Table 4.S3: ^1H and ^{13}C chemical shifts for (+)-11-oxo-copal-13-ol (34)	157
Table 4.S4: ^1H and ^{13}C chemical shifts for (+)-11-oxo-copal-8,13-ol (46)	161
Table 4.S5: ^1H and ^{13}C chemical shifts for <i>ent</i> -11-oxo-copal-8,13-ol (53)	165
Table 4.S6: ^1H and ^{13}C chemical shifts for 13-oxo-casbene (61)	169

LIST OF FIGURES

Figure 1.1: Biosynthetic origin of terpenoids and diversity in diterpenoid biosynthesis	4
Figure 2.1: Distribution of serrulatane diterpenoids in members of the Scrophulariaceae family	26
Figure 2.2: Maximum likelihood phylogenetic tree of TPS candidates	29
Figure 2.3: Dihydroserrulatene production by <i>Lf</i> TPS1	31
Figure 2.4: NNPP production by <i>Lf</i> CPT1	34
Figure 2.5: Leubethanol production by CYP71D616	36
Figure 2.S1: GC-MS analysis of MTBE extracts of <i>L. frutescens</i> flower, leaf, and root tissue ...	49
Figure 2.S2: First 120 positions of a sequence alignment of each reference and candidate TPS-a from Figure 2.2 in the main text	50
Figure 2.S3: CG-MS chromatograms of initial screening of <i>Lf</i> TPS1 and <i>Lf</i> TPS2	51
Figure 2.S4: GC-MS chromatograms for initial screening of <i>Lf</i> TPS1 and <i>Lf</i> TPS2 activity against NNPP in independent systems	52
Figure 2.S5: ¹ H, ¹³ C, HSQC, H2BC, HMBC, COSY, and NOESY NMR spectra for dihydroserrulatene	54
Figure 2.S6: GC-MS chromatograms for initial screening of <i>Lf</i> CPT1-3 co-expressed with <i>Lf</i> TPS1 in <i>N. benthamiana</i>	59
Figure 2.S7: Maximum likelihood phylogenetic tree of candidate cytochrome P450s	60
Figure 2.S8: GC-MS chromatograms for initial screening of <i>L. frutescens</i> P450 candidates co-expressed with <i>Lf</i> CPT1 and <i>Lf</i> TPS1 in <i>N. benthamiana</i>	61
Figure 2.S9: GC-MS chromatograms of transient expression of the <i>Eremophila serrulata</i> orthologue to <i>Lf</i> TPS1 (<i>Es</i> TPS1)	62
Figure 2.S10: Stereo-view of two different homology models of <i>Lf</i> CPT1 aligned with model templates	63
Figure 2.S11: Sequence alignment of reference <i>cis</i> -PTs from <i>S. lycopersicum</i> , <i>Li</i> LPPS, each <i>L. frutescens cis</i> -PT tested in this study, and the <i>Lf</i> CPT1 orthologue from <i>E. serrulata</i> (<i>Es</i> CPT1)	64

Figure 2.S12: Maximum likelihood phylogenetic tree of TPSs shown in Figure 2.2 in the main text, with recently identified TPSs added from three other <i>Eremophila</i> species.....	65
Figure 2.S13: Maximum likelihood phylogenetic tree of <i>cis</i> -PTs shown in Figure 2.4 in the main text, with recently identified <i>cis</i> -PTs added from three other <i>Eremophila</i> species	66
Figure 3.1: Common structural features of diterpenoid alkaloids and proposed biosynthetic pathway	80
Figure 3.2: Maximum likelihood phylogenetic tree of predicted <i>D. grandiflorum</i> TPS sequences	82
Figure 3.3: <i>Dgr</i> TPS1 is an <i>ent</i> -CPP synthase	83
Figure 3.4: <i>Dgr</i> TPS7a and <i>Dgr</i> TPS7b convert <i>ent</i> -CPP to <i>ent</i> -atiserene	84
Figure 3.5: Process of filtering Cytochrome P450 transcripts for candidate selection.....	85
Figure 3.6: CYP701A127 and CYP71FH1 convert <i>ent</i> -atiserene to oxidized products.....	86
Figure 3.7: Coexpression of CYP701A127 and CYP71FH1 lead to an accumulation of the same products.....	88
Figure 3.8: CYP729G1 and CYP71FK1 have redundant functions	89
Figure 3.9: Nitrogen incorporation into diterpenoid alkaloids likely involves iminium cation resolution through reduction <i>and</i> substitution	91
Figure 3.10: Coexpression analysis on <i>Aconitum vilmorinianum</i> and BLAST search back against the four <i>Delphinium/Aconitum</i> transcriptome assemblies	92
Figure 3.11: Coexpression with SangRed produces an isomer of what is produced upon supplementation with ethylamine	94
Figure 3.S1: ¹ H, ¹³ C, HSQC, H2BC, HMBC, COSY, and NOESY NMR spectra for <i>ent</i> -atiserene	105
Figure 3.S2: Maximum likelihood phylogenetic tree of candidate P450s from the CYP71 clan.....	108
Figure 3.S3: Maximum likelihood phylogenetic tree of candidate P450s from the CYP72 clan.....	109
Figure 3.S4: Maximum likelihood phylogenetic tree of candidate P450s from the CYP85 clan.....	110

Figure 3.S5: Maximum likelihood phylogenetic tree of candidate P450s from the CYP97 clan	111
Figure 3.S6: Mass spectra for all compounds shown in Figure 3.6 in the main text	112
Figure 3.S7: ^1H , ^{13}C , HSQC, H2BC, HMBC, COSY, and NOESY NMR spectra of <i>ent</i> -atiserene-20-al	113
Figure 3.S8: Select HMBC correlations for <i>ent</i> -atiserene-20-al.....	116
Figure 3.S9: Mass spectra for all compounds shown in Figure 3.7 in the main text	117
Figure 3.S10: Mass spectra for all compounds shown in Figure 3.8 in the main text	118
Figure 3.S11: CYP729G1 and CYP71FK1 still have similar activity when coexpressed with SangRed	119
Figure 4.1: Summary of active substrate and TPS combinations	132
Figure 4.2: Example of class II TPS conversion of modified substrates to derivatives of native products.....	134
Figure 4.3: Structures of select products.....	137
Figure 4.4: <i>Pv</i> TPS4 converts six modified substrates to cadinene	140
Figure 4.S1: NMR spectra (^1H , ^{13}C , HSQC, H2BC, HMBC, NOESY, COSY) for (+)-14-methylcopal-8,13-ol (24)	150
Figure 4.S2: NMR spectra (^1H , ^{13}C , HSQC, H2BC, HMBC, NOESY, COSY) for (+)-14-fluorocopal-8,13-ol (31).....	154
Figure 4.S3: NMR spectra (^1H , ^{13}C , HSQC, H2BC, HMBC, NOESY, COSY) for (+)-11-oxo-copal-8-ol (34)	158
Figure 4.S4: NMR spectra (^1H , ^{13}C , HSQC, H2BC, HMBC, NOESY, COSY) for (+)-11-oxo-copal-8,13-ol (46)	162
Figure 4.S5: NMR spectra (^1H , ^{13}C , HSQC, H2BC, HMBC, NOESY, COSY) for <i>ent</i> -11-oxo-copal-8,13-ol (53)	166
Figure 4.S6: NMR spectra (^1H , ^{13}C , HSQC, H2BC, HMBC, NOESY, COSY) for 13-oxo-casbene (61).....	170
Figure 4.S7: GC-MS screening of class II/class I TPS combinations with hits from GC-FID screening	173

Figure 4.S8: GC-MS screening of single-step class I TPSs with hits from GC-FID screening	176
Figure 4.S9: GC-MS screening of class II TPSs with hits from GC-FID screening	179
Figure 4.S10: Mass spectra for all compounds (21-74) shown in GC-MS screening	183
Figure 4.S11: Initial GC-FID screening of class II TPSs with substrates 1-20	193
Figure 4.S12: Initial GC-FID screening of class II/class I TPS combinations	198
Figure 4.S13: GC-FID chromatograms for second-step class I TPS negative controls.....	201
Figure 4.S14: Initial GC-FID screening of single-step class I TPSs with substrates 1-20	203

LIST OF SCHEMES

Scheme 2.S1: Proposed mechanism for *Lf*TPS1 conversion of NNPP to dihydroserulatene58

KEY TO ABBREVIATIONS

IPP – isopentenyl diphosphate

DMAPP – dimethylallyl diphosphate

TPS – terpene synthase

diTPS – diterpene synthase

P450 or CYP – cytochrome P450

GPP – geranyl diphosphate

FPP – farnesyl diphosphate

GGPP – geranylgeranyl diphosphate

MVA – mevalonate

MEP – methyl-erythritol phosphate

CPP – copalyl diphosphate

BGC – biosynthetic gene cluster

NNPP – nerylneryl diphosphate

CPT – *cis*-prenyl transferase

PT – prenyl transferase

SRA – Sequence Read Archive (NCBI)

GC-MS – gas chromatography-mass spectrometry

NMR – nuclear magnetic resonance

NPP – neryl diphosphate

(*Z-Z*)-FPP - *Z-Z*-farnesyl diphosphate

DXS – 1-deoxy-D-xylulose-5-phosphate synthase

MTBE – methyl *tertiary*-butyl ether

BLAST – basic local alignment search tool

CDCl_3 - deuteriochloroform

PDB – Protein Data Bank

GGPPS – geranylgeranyl diphosphate synthase

LC-MS – liquid chromatography-mass spectrometry

MR – mutual rank

CHAPTER 1

Plant Terpenoid Biosynthesis and Elucidation of Biosynthetic Pathways

Specialized Metabolism

Plants produce an array of specialized metabolites with widespread native uses such as antifeedants and pesticides¹, pigments², and pollinator attractants³. Estimates of how many unique specialized metabolites exist are in the hundreds of thousands to millions, with likely thousands produced in individual species⁴. Specialized metabolites are distinct from central metabolites in that they are not present in every plant species and may offer selective advantages, but are not part of the metabolic processes common to all plants⁵. They are often referred to as “secondary metabolites,” however the latter was a name assigned primarily when the utility of these compounds was unknown and were simply regarded as byproducts^{4,6}. Given the sheer number of specialized metabolites found in nature, it is perhaps unsurprising that humans have found widespread uses for these compounds such as flavors and fragrances⁷, cosmetics⁸, biofuels⁹, poisons¹⁰, and medicines¹¹.

Prominent examples of specialized metabolites in medicine are paclitaxel from *Taxus brevifolia* (anti-cancer; sold under the name Taxol)^{12,13} and artemisinin from *Artemisia annua* (anti-malaria)¹⁴. Given the demand for these drugs, methods to produce them at a relevant scale and cost are highly sought after. Total chemical syntheses for both paclitaxel^{15,16} and artemisinin¹⁷ have been solved nearly thirty years ago, however methods to produce both have heavily relied on semi-biosynthesis^{18,19} given the complexity of their structures. Another prominent example are the rebaudiosides—low calorie sweeteners found in *Stevia rebaudiana*—which have gained popularity over use of whole *S. rebaudiana* extracts that contain a mixture of sweet and bitter compounds²⁰. These are distinct by different numbers and linkages of glycosyl subunits (i.e. difficult to isolate from an extract or synthesize specifically) and are commercially produced through biosynthesis by biomanufacturing companies such as Amyris (Purecane)²¹, Conagen

(Bestevia)²², and Manus Bio (NutraSweet)²³.

Terpenoids

Each of the above examples is a terpenoid: the largest class of specialized metabolites with more than 65,000 known compounds in plants estimated as of 2019²⁴. Terpenoids trace their biosynthetic origin back to the five-carbon precursors isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). IPP and DMAPP are converted by prenyl transferases into prenyl diphosphate substrates with varying lengths, arrangements, and stereochemistry^{25,26}. These are then converted to terpene “scaffolds” through carbocation cascade reactions catalyzed by terpene synthases (TPSs)²⁶. Following this initial scaffold formation, subsequent “tailoring” steps convert these terpene scaffolds into terpenoids, usually involving oxidation by enzymes such as P450s and 2-oxoglutarate-dependent dioxygenases and further modification by methyl-, acetyl-, and glycosyl transferases, among many others²⁶.

TPSs can further be classified by their mechanism into class II or class I enzymes, depending on the initial formation of a cation through protonation of an alkene (class II) or dephosphorylation (class I)²⁷. They are also grouped into subfamilies from a-h depending on their evolutionary origin. Those which will be discussed throughout this thesis are primarily the subfamilies: TPS-a (typically cytosolic sesquiterpene synthases; Chapters 2 and 4), TPS-c (typically plastidial class II diterpene synthases; Chapters 3 and 4) and TPS-e (typically plastidial class I diterpene synthases; Chapters 3 and 4). The biosynthesis of most mono- (C₁₀) and sesquiterpenoids (C₁₅) follow a single-step class I mechanism, where a prenyl diphosphate substrate—geranyl diphosphate (GPP) or farnesyl diphosphate (FPP), respectively—are cyclized by a class I TPS directly into a terpene scaffold. Diterpenoid (C₂₀) biosynthesis can follow two

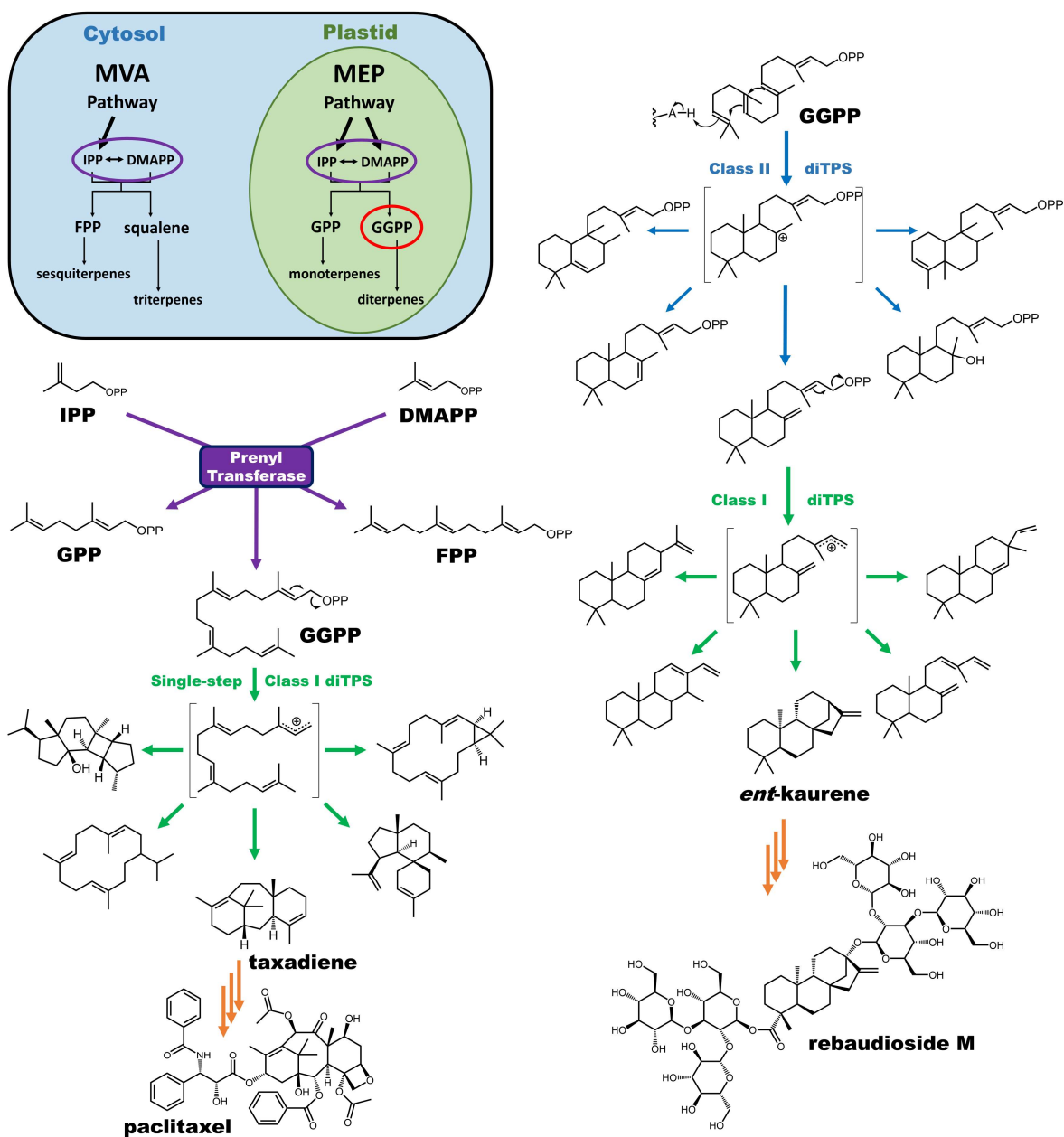


Figure 1.1: Biosynthetic origin of terpenoids and diversity in diterpenoid biosynthesis. (Top Left) Two unique pathways produce IPP and DMAPP in plant cells: the mevalonate (MVA) pathway in the cytosol and the methyl-erythritol phosphate (MEP) pathway in the plastid. Distinct prenyl transferases make FPP in the cytosol (which also condense to form the triterpenoid precursor squalene—not discussed in this thesis), and GPP and GGPP in the plastid. TPSs are colocalized based on the prenyl diphosphate substrates which they natively convert to terpene scaffolds. (Middle Left) IPP and DMAPP are condensed into prenyl diphosphate substrates of varying lengths and stereochemistry; only the three most common are shown here. (Bottom Left) Examples of single step class I conversion of GGPP to diterpene scaffolds; highlighted is the scaffold formation of taxadiene, precursor to paclitaxel. (Right) Examples of multi-step class II/ class I conversion of GGPP to diterpene scaffolds; highlighted is the scaffold formation of *ent*-kaurene, precursor to rebaudioside M.

mechanisms for the cyclization of geranylgeranyl diphosphate (GGPP): single step conversion by a class I enzyme, or multi-step conversion first involving cyclization by a class II TPS followed by further modification by a class I TPS. The latter is most common, and gives rise to the labdane-type diterpenoids with a characteristic decalin core derived from the class II TPS cyclization²⁷, while the former can give rise to a greater diversity of structures. Examples of both, including a small sample of the diversity of structures that can be made, are shown in Figure 1.1.

While a vast range of enzyme families which carry out tailoring steps can follow the initial terpene scaffold formation, cytochromes P450 (P450s) typically carry out initial modifications in the form of site- and stereospecific oxidations²⁸. The number of P450s in a given plant species is enormous, as they typically make up about one percent of all protein coding genes in a genome, leading to hundreds of genes per plant²⁹. Terpenoid-related P450s are typically members of the CYP71 clan—the largest group of plant P450s³⁰—although exceptions exist^{30–32}. Modifications by P450s (and other monooxygenases) enable the addition of other groups from various transferase enzymes (methyl-, acetyl-, etc.) and can facilitate further rearrangement of the initial scaffold.

Evolutionary Considerations in Specialized Metabolism

As highlighted above for terpenoids, the biosynthesis of a vast array of specialized metabolites can stem from only a handful of starting substrates. Different classes of specialized metabolites trace their origin back to different places in central metabolism³³. IPP and DMAPP for terpenoids, for example, are central metabolites. Another major class are the alkaloids, which are generally described as specialized metabolites containing nitrogen. An alkaloid scaffold is typically formed through the condensation of an aldehyde and amine and subsequent iminium quenching to form an alkaloid scaffold, with common examples utilizing the central metabolites

putrescine and various amino acids³⁴. Other examples include phenylpropanoids which also contain building blocks derived from aromatic amino acids in the shikimate pathway³⁵, and polyketides which utilize various CoA substrates (acetyl-, malonyl-, etc.) and are usually distinguishable by alternating carbonyl and methylene groups³⁶.

The enzymatic conversion of these central metabolites into specialized metabolites similarly stems from enzymes involved in central metabolism. The TPS family described above, for example, evolved from a single bifunctional *ent*-copalyl diphosphate (*ent*-CPP)/*ent*-kaurene synthase²⁶, an extant example of which can be found in the bryophyte *Physcomitrium patens*³⁷. Gene duplication and neofunctionalization subsequently led to distinct *ent*-CPP and *ent*-kaurene synthases²⁶, which were likely the founding members of the TPS-c and TPS-e subfamilies, respectively. *Ent*-kaurene is a central metabolite involved in the biosynthesis of plant growth hormones (gibberellins)³⁸, and so continuation of this duplication/functionalization process allowed for the retention *ent*-kaurene biosynthesis while expanding terpenoid pathways towards specialized metabolism. Such duplications can occur through a variety of mechanisms, including tandem duplications (duplication of a portion of DNA due to unequal crossover) and whole genome duplications (e.g. through mis segregation of chromosomes during meiosis)^{39,40}.

Another major driving force in the evolution of specialized metabolism is enzyme promiscuity⁴¹. Retaining a high degree of substrate and product specificity would render gene duplication a useless process with respect to the generation of new metabolites. Promiscuous enzymes can serve as the basis for new functions, both in the case of one involved in primary metabolism which can maintain its central role while quickly being able to take on a new one following duplication⁴², or one already involved in specialized metabolism which can make multiple products that may infer different selective advantages and adaptability to a changing

environment. Similarly, intracellular compartment-switching of enzymes has been demonstrated to be another major driver in the evolution of specialized metabolism^{43,44}, as the substrate availability (i.e. FPP in the cytosol versus GGPP in the plastid) changes with localization. Substrate promiscuity in TPSs is common, with many examples of those which can convert a range of substrates, even when those substrates are not native to the enzyme's species of origin^{45,46}.

Signatures of Biosynthetic Pathways and Strategies to Find Them

The first and most important step in elucidating a biosynthetic pathway is to propose a pathway to begin with, involving starting points in central metabolism, necessary chemical transformations and intermediates, and enzyme families that could carry out such transformations. The initial scaffold-forming steps are typically where central metabolism precursors are most obvious, for example a cyclized ten, fifteen, or twenty carbon scaffold is likely derived from IPP and DMAPP (terpenoid), or a pair of amino acids linked together by a nitrogen (alkaloid). Hydroxylations of this scaffold are typically carried out by P450s or other monooxygenases, and modifications of these hydroxyl groups can be carried out by other enzyme families such as acetyl-, methyl-, or glycosyltransferases. Identification of enzymes in families predicted to be involved is typically done through mining transcriptomic or genomic data, and once a list of each possible candidate is curated, filtering based on properties of the specific pathway can be done to narrow down candidates to functionally characterize.

The lineage-specific nature of specialized metabolites can provide one of the simplest and most effective filters for narrowing down candidates. In the Solanaceae family, for example, separate specialized metabolites derived from N-methylpyrrolinium exist in separate lineages (e.g. nicotine in *Nicotiana*, scopolamine in *Atropa*), presumably through gene duplication leading to N-

methylepyrrolinium accumulation in their common ancestor, and evolution of distinct pathways to form distinct metabolites following speciation³⁴. In this case, orthologous genes are present in each lineage which are responsible for the accumulation on N-methylepyrrolinium, while the following steps in the respective pathways are present in one lineage but *not* the other. Both of these cases have been utilized to discover biosynthetic enzymes. In the case of glucosinolates (defense compounds in Brassicaceae), metabolites present in the *Brassica* genus are not present in the neighboring *Arabidopsis* genus, and pathway enzymes were found through filtering by those which were present in *Brassica* but *not* in *Arabidopsis*^{47,48}. Conversely, both the *Leucophyllum* and neighboring *Eremophila* genera produce antimicrobial serrulatane diterpenoids, and elucidation of this pathway involved filtering by candidates which were present in *both* genera (see Chapter 2)⁴⁹.

Specialized metabolites often accumulate in particular tissue types as well, which can serve as a similar type of filter. The elucidation of the forskolin pathway took advantage of its tissue-specific accumulation in the root cork, and the entire biosynthetic pathway (six enzymes) was found through tissue-specific RNA sequencing^{50,51}. A similar approach was taken for finding enzymes involved in the tropane alkaloid⁵² and serrulatane⁴⁹ pathways.

Specialized metabolites often accumulate to different levels in response to environmental stimuli or throughout different stages of development. Because of this, the biosynthetic genes responsible for making them often have differential expression levels throughout these different conditions as well. Elucidation of biosynthetic pathways can take advantage of this differential expression by identifying genes that have varying expression correlated with varying accumulation of metabolites⁵³ or with other genes already known to be involved in the pathway⁵⁴, and selecting differential conditions for RNA sequencing such as developmental stages⁵⁵, or pre- and post-wounding⁵⁶ or treatment with methyl jasmonate^{57,58}. This can be especially helpful when the

families of enzymes to search for a given step are not obvious, as prior knowledge of enzyme families and their functions is not required when they fall into a coexpressed cluster⁵⁹ along with previously identified genes⁶⁰. Nett et al., for example, chemically synthesized a known intermediate in the colchicine pathway because a methylation step in the middle of their proposed pathway was the most obvious in terms of what enzyme family to search. Once they identified the methyltransferase that could convert this synthetic intermediate, they found enzymes coexpressed with this “anchor gene” to identify the remainder of the pathway⁵⁴. Similarly, Jozwiak et al. used differential expression between tissue types to identify three genes involved in triterpenoid saponin biosynthesis, and subsequently carried out coexpression analysis with these three genes as “anchors” to find an unexpected gene typically implicated in cell wall biosynthesis which was actually involved in this pathway⁶¹.

While not as common as in bacteria or fungi, genes involved in plant specialized metabolism can form biosynthetic gene clusters (BGCs), which can aid in the co-regulation of pathway genes to prevent buildup of toxic intermediates and inheritance of incomplete pathways⁶². The momilactone biosynthetic pathway, for example, has been found in three species of plants, and has emerged independently at least twice⁶³. The presence of these pathways within BGCs is hypothesized to be a method to reduce intermediate toxicity from an incomplete pathway or one that is not sufficiently co-regulated⁶⁴. Similar to the use of coexpression analysis, searching for candidate genes colocalized with known pathway genes in a genome may reveal other enzymes active in a pathway. A miltiradiene BGC found throughout the Lamiaceae contains several P450s which were found to be active in oxidizing the miltiradiene scaffold, and were selected for characterization simply because of their colocalization in this cluster in *Callicarpa americana*⁶⁵.

Biochemical Considerations in Elucidating Biosynthetic Pathways

Although enzymes are often framed as “cheaters” when explaining basic principles in organic chemistry, it’s important to consider mechanistic details of the types of reactions these enzymes actually carry out when seeking out a biochemical pathway. Thinking of this from less of a “black box” perspective is useful as it helps to understand some of the exceptions, uncommon reactions, and unexpected origins with enzymes implicated in specialized metabolism.

Terpene synthases bind a prenyl diphosphate substrate (or labdane intermediate), make a cation (by dephosphorylation or protonation), and mediate a carbocation cascade rearrangement. It would be an oversimplification to say, for example, that a TPS-b enzyme takes GPP and converts it to a monoterpene. Thinking of an enzyme in terms of its mechanism rather than its predicted function helps to not inherently exclude specific candidates when looking for a pathway—i.e. a given TPS sequence is not a monoterpene synthase simply because it belongs to the TPS-b subfamily. Examples of diterpene synthases in the TPS-b subfamily have been seen which use (+)-CPP⁶⁶ (a labdane-type intermediate) and NNPP⁶⁷ (neryleryl diphosphate; all *cis*-prenyl isomer of GGPP). Likewise, some compartment-switching, plastidial members of the TPS-a (typically cytosolic sesquiterpene synthase) subfamily are diterpene synthases which use GGPP^{44,68,69} or NNPP^{49,67}, and there are TPS-e enzymes (typically labdane-type diterpene synthases) in the Solanaceae family which convert *cis*-prenyl substrates^{70–72}. While predicting a TPS’s function by subfamily is typically better than a random guess, there are enough exceptions that one should be careful in making these assumptions; the evolution of a TPS to use a different substrate⁴⁴ or give a different product⁷³ happens frequently.

Cytochrome P450s have an inherently complicated reaction mechanism, including the involvement of a reductase partner enzyme and nine steps in a catalytic cycle⁷⁴ towards site and

stereo-specific oxidations of a given substrate. While this typically involves the addition of a single hydroxyl group, exceptions exist such as those that act as desaturases⁷⁵, epoxidases⁴⁷, or carry out sequential oxidations of a single position⁷⁶. Often times, P450s are implicated in modification of the initial scaffold of a specialized metabolite in ways that the scaffold-forming step could not carry out. Examples include additional ring formation⁴⁷, expansion⁵⁴, contraction³¹, or C-C bond cleavage⁷⁷. Scaffolds for a specialized metabolite that appear modified beyond what may be possible for the respective scaffold-forming step (e.g. a TPS) may involve a P450.

In contrast to the examples above, sometimes the mechanism of a given chemical transformation is not obvious in terms of which family of enzymes to even begin searching in the first place. Take norcoclaurine biosynthesis for example: dopamine (a primary amine) and 4-hydroxyphenyl-acetaldehyde (an aldehyde) condense to form an iminium cation, which is quenched through an electrophilic aromatic substitution (known as a Pictet-Spengler reaction) with the *meta* hydroxyl of dopamine acting as a directing group³⁴. This can happen spontaneously in solution⁷⁸, although the reaction is enzyme-catalyzed in instances of norcoclaurine biosynthesis throughout the Ranunculales^{79,80}. In *Coptis japonica*, there are two norcoclaurine synthases from two distinct enzyme families which do not have any evolutionary history of carrying out Pictet-Spengler reactions⁷⁹. One of these enzymes (*Cj*NCS2) is most closely related to the 2-oxoglutarate-dependent dioxygenase family, however despite what the name would imply, *Cj*NCS2 is neither 2-oxoglutarate dependent nor a dioxygenase⁷⁹. In contrast to the reactions carried out by enzymes like TPSs or P450s, searching for enzymes which catalyze reactions that can happen spontaneously can present a challenge in terms of where to search, and may benefit from mechanism-agnostic search strategies like coexpression analysis or searching for genomic clustering.

Work Presented in this Dissertation

The research detailed here seeks both to discover new terpenoid pathways and repurpose known pathways for the conversion of non-native substrates. Chapter 2 details the elucidation of the full biosynthetic pathway to leubethanol, a diterpenoid active against multidrug-resistant TB and metabolic precursor to the entire class of antimicrobial serrulatane diterpenoids from the Scrophulariaceae family. This involved an uncommon prenyl diphosphate substrate—only seen in one previous instance—and highlighted the importance of the consideration of where a pathway truly begins with respect to central metabolism. Chapter 3 details the identification of seven enzymes active in the pathway towards diterpenoid alkaloids found in the Ranunculaceae family, which are a class of hundreds of structurally complex metabolites with a range of applications. This work highlighted the utility of incorporating public data and cross-referencing datasets to elucidate biochemical pathways in instances of species with hundreds of candidate enzymes to filter through. Finally, Chapter 4 details how the substrate promiscuity of many TPSs can be utilized in the conversion of synthetic substrates not found in nature. More than 500 enzyme-substrate (and enzyme-enzyme-substrate, in the case of class II/class I biosynthesis) combinations were tested, resulting in a range of promiscuous activities and more than fifty novel products. This work highlights the ability of TPSs to catalyze like-nature reactions with modified substrates, leading to chemically-derivatized products in cases where derivatization of a substrate is feasible, but of the target product is not.

REFERENCES

REFERENCES

- (1) Kortbeek, R. W. J.; van der Gragt, M.; Bleeker, P. M. Endogenous Plant Metabolites against Insects. *Eur J Plant Pathol* **2019**, *154* (1), 67–90. <https://doi.org/10.1007/s10658-018-1540-6>.
- (2) Moghe, G. D.; Smith, S. D. The Push and Pull of Plant Specialized Metabolism Underlies a Long-Standing, Colorful Mystery. *The New Phytologist* **2018**, *217* (2), 471–473.
- (3) Wink, M. Plant Secondary Metabolites Modulate Insect Behavior-Steps Toward Addiction? *Frontiers in Physiology* **2018**, *9*.
- (4) Pichersky, E.; Lewinsohn, E. Convergent Evolution in Plant Specialized Metabolism. *Annu Rev Plant Biol* **2011**, *62*, 549–566. <https://doi.org/10.1146/annurev-arplant-042110-103814>.
- (5) Hartmann, T. From Waste Products to Ecochemicals: Fifty Years Research of Plant Secondary Metabolism. *Phytochemistry* **2007**, *68* (22), 2831–2846. <https://doi.org/10.1016/j.phytochem.2007.09.017>.
- (6) Fernie, A. R.; Pichersky, E. Focus Issue on Metabolism: Metabolites, Metabolites Everywhere. *Plant Physiology* **2015**, *169* (3), 1421–1423. <https://doi.org/10.1104/pp.15.01499>.
- (7) Gang, D. R. Evolution of Flavors and Scents. *Annu Rev Plant Biol* **2005**, *56*, 301–325. <https://doi.org/10.1146/annurev-arplant.56.032604.144128>.
- (8) Faccio, G. Plant Complexity and Cosmetic Innovation. *iScience* **2020**, *23* (8), 101358. <https://doi.org/10.1016/j.isci.2020.101358>.
- (9) Voloshin, R. A.; Rodionova, M. V.; Zharmukhamedov, S. K.; Nejat Veziroglu, T.; Allakhverdiev, S. I. Review: Biofuel Production from Plant and Algal Biomass. *International Journal of Hydrogen Energy* **2016**, *41* (39), 17257–17273. <https://doi.org/10.1016/j.ijhydene.2016.07.084>.
- (10) Chan, T. Y. K. Aconite Poisoning. *Clinical Toxicology* **2009**, *47* (4), 279–285. <https://doi.org/10.1080/15563650902904407>.
- (11) Balunas, M. J.; Kinghorn, A. D. Drug Discovery from Medicinal Plants. *Life Sciences* **2005**, *78* (5), 431–441. <https://doi.org/10.1016/j.lfs.2005.09.012>.
- (12) Schiff, P. B.; Fant, J.; Horwitz, S. B. Promotion of Microtubule Assembly in Vitro by Taxol. *Nature* **1979**, *277* (5698), 665–667. <https://doi.org/10.1038/277665a0>.
- (13) Wani, M. C.; Taylor, H. L.; Wall, M. E.; Coggon, P.; McPhail, A. T. Plant Antitumor Agents. VI. The Isolation and Structure of Taxol, a Novel Antileukemic and Antitumor Agent

from *Taxus Brevifolia*. *J Am Chem Soc* **1971**, 93 (9), 2325–2327. <https://doi.org/10.1021/ja00738a045>.

(14) Tu, Y. The Discovery of Artemisinin (Qinghaosu) and Gifts from Chinese Medicine. *Nat Med* **2011**, 17 (10), 1217–1220. <https://doi.org/10.1038/nm.2471>.

(15) Nicolaou, K. C.; Yang, Z.; Liu, J. J.; Ueno, H.; Nantermet, P. G.; Guy, R. K.; Claiborne, C. F.; Renaud, J.; Couladouros, E. A.; Paulvannan, K.; Sorensen, E. J. Total Synthesis of Taxol. *Nature* **1994**, 367 (6464), 630–634. <https://doi.org/10.1038/367630a0>.

(16) Holton, R. A.; Somoza, C.; Kim, H. B.; Liang, F.; Biediger, R. J.; Boatman, P. D.; Shindo, M.; Smith, C. C.; Kim, S. First Total Synthesis of Taxol. 1. Functionalization of the B Ring. *J. Am. Chem. Soc.* **1994**, 116 (4), 1597–1598. <https://doi.org/10.1021/ja00083a066>.

(17) Avery, M. A.; Chong, W. K. M.; Jennings-White, C. Stereoselective Total Synthesis of (+)-Artemisinin, the Antimalarial Constituent of *Artemisia Annua* L. *J. Am. Chem. Soc.* **1992**, 114 (3), 974–979. <https://doi.org/10.1021/ja00029a028>.

(18) Paddon, C. J.; Keasling, J. D. Semi-Synthetic Artemisinin: A Model for the Use of Synthetic Biology in Pharmaceutical Development. *Nat Rev Microbiol* **2014**, 12 (5), 355–367. <https://doi.org/10.1038/nrmicro3240>.

(19) Prince, C. L.; Schubmehl, B. F.; Kane, E. J.; Roach, B.; Bringi, V.; Kadkade, P. G. Enhanced Production of Taxol and Taxanes by Cell Cultures of *Taxus* Species. EP0960944A1, December 1, 1999.

(20) Prakash, I.; Markosyan, A.; Bunders, C. Development of Next Generation Stevia Sweetener: Rebaudioside M. *Foods* **2014**, 3 (1), 162–175. <https://doi.org/10.3390/foods3010162>.

(21) Zhao, L.; Li, W.; Wichmann, G.; Khankhoje, A.; Garcia, D. G. C.; Mahatdejkul-Meadows, T.; Jackson, S.; Leavell, M.; Platt, D. Udp-Dependent Glycosyltransferase for High Efficiency Production of Rebaudiosides. SG11201900930UA, February 27, 2019.

(22) Mao, G.; Yu, X. Non-Caloric Sweeteners and Methods for Synthesizing. US9567619B2, February 14, 2017.

(23) PHILIPPE, R.; KUMARAN, A. P.; Donald, J.; PATEL, K.; Gupta, S.; LIM, R.; Li, L. Microbial Production of Steviol Glycosides. US10463062B2, November 5, 2019.

(24) Zeng, T.; Liu, Z.; Liu, H.; He, W.; Tang, X.; Xie, L.; Wu, R. Exploring Chemical and Biological Space of Terpenoids. *J. Chem. Inf. Model.* **2019**, 59 (9), 3667–3678. <https://doi.org/10.1021/acs.jcim.9b00443>.

(25) Zhou, F.; Pichersky, E. More Is Better: The Diversity of Terpene Metabolism in Plants. *Current Opinion in Plant Biology* **2020**, 55, 1–10. <https://doi.org/10.1016/j.pbi.2020.01.005>.

- (26) Karunanithi, P. S.; Zerbe, P. Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity. *Frontiers in Plant Science* **2019**, *10*.
- (27) Peters, R. J. Two Rings in Them All: The Labdane-Related Diterpenoids. *Nat. Prod. Rep.* **2010**, *27* (11), 1521–1530. <https://doi.org/10.1039/C0NP00019A>.
- (28) Hamberger, B.; Bak, S. Plant P450s as Versatile Drivers for Evolution of Species-Specific Chemical Diversity. *Philos Trans R Soc Lond B Biol Sci* **2013**, *368* (1612), 20120426. <https://doi.org/10.1098/rstb.2012.0426>.
- (29) Nelson, D.; Werck-Reichhart, D. A P450-Centric View of Plant Evolution. *The Plant Journal* **2011**, *66* (1), 194–211. <https://doi.org/10.1111/j.1365-313X.2011.04529.x>.
- (30) Bathe, U.; Tissier, A. Cytochrome P450 Enzymes: A Driving Force of Plant Diterpene Diversity. *Phytochemistry* **2019**, *161*, 149–162. <https://doi.org/10.1016/j.phytochem.2018.12.003>.
- (31) Helliwell, C. A.; Chandler, P. M.; Poole, A.; Dennis, E. S.; Peacock, W. J. The CYP88A Cytochrome P450, Ent-Kaurenoic Acid Oxidase, Catalyzes Three Steps of the Gibberellin Biosynthesis Pathway. *Proceedings of the National Academy of Sciences* **2001**, *98* (4), 2065–2070. <https://doi.org/10.1073/pnas.98.4.2065>.
- (32) Nomura, T.; Magome, H.; Hanada, A.; Takeda-Kamiya, N.; Mander, L. N.; Kamiya, Y.; Yamaguchi, S. Functional Analysis of *Arabidopsis* CYP714A1 and CYP714A2 Reveals That They Are Distinct Gibberellin Modification Enzymes. *Plant Cell Physiol* **2013**, *54* (11), 1837–1851. <https://doi.org/10.1093/pcp/pct125>.
- (33) Pott, D. M.; Osorio, S.; Vallarino, J. G. From Central to Specialized Metabolism: An Overview of Some Secondary Compounds Derived From the Primary Metabolism for Their Role in Conferring Nutritional and Organoleptic Characteristics to Fruit. *Frontiers in Plant Science* **2019**, *10*.
- (34) Lichman, B. R. The Scaffold-Forming Steps of Plant Alkaloid Biosynthesis. *Nat. Prod. Rep.* **2021**, *38* (1), 103–129. <https://doi.org/10.1039/D0NP00031K>.
- (35) Vogt, T. Phenylpropanoid Biosynthesis. *Molecular Plant* **2010**, *3* (1), 2–20. <https://doi.org/10.1093/mp/ssp106>.
- (36) Nivina, A.; Yuet, K. P.; Hsu, J.; Khosla, C. Evolution and Diversity of Assembly-Line Polyketide Synthases. *Chem. Rev.* **2019**, *119* (24), 12524–12547. <https://doi.org/10.1021/acs.chemrev.9b00525>.
- (37) Hayashi, K.; Kawaide, H.; Notomi, M.; Sakigi, Y.; Matsuo, A.; Nozaki, H. Identification and Functional Analysis of Bifunctional Ent-Kaurene Synthase from the Moss *Physcomitrella Patens*. *FEBS Letters* **2006**, *580* (26), 6175–6181. <https://doi.org/10.1016/j.febslet.2006.10.018>.

- (38) Grennan, A. K. Gibberellin Metabolism Enzymes in Rice. *Plant Physiology* **2006**, *141* (2), 524–526. <https://doi.org/10.1104/pp.104.900192>.
- (39) Panchy, N.; Lehti-Shiu, M.; Shiu, S.-H. Evolution of Gene Duplication in Plants. *Plant Physiology* **2016**, *171* (4), 2294–2316. <https://doi.org/10.1104/pp.16.00523>.
- (40) Qiao, X.; Li, Q.; Yin, H.; Qi, K.; Li, L.; Wang, R.; Zhang, S.; Paterson, A. H. Gene Duplication and Evolution in Recurring Polyploidization–Diploidization Cycles in Plants. *Genome Biology* **2019**, *20* (1), 38. <https://doi.org/10.1186/s13059-019-1650-2>.
- (41) O’Brien, P. J.; Herschlag, D. Catalytic Promiscuity and the Evolution of New Enzymatic Activities. *Chem Biol* **1999**, *6* (4), R91–R105. [https://doi.org/10.1016/S1074-5521\(99\)80033-7](https://doi.org/10.1016/S1074-5521(99)80033-7).
- (42) Aharoni, A.; Gaidukov, L.; Khersonsky, O.; Gould, S. M.; Roodveldt, C.; Tawfik, D. S. The “evolvability” of Promiscuous Protein Functions. *Nat Genet* **2005**, *37* (1), 73–76. <https://doi.org/10.1038/ng1482>.
- (43) Schenck, C. A.; Last, R. L. Location, Location! Cellular Relocalization Primes Specialized Metabolic Diversification. *The FEBS Journal* **2020**, *287* (7), 1359–1368. <https://doi.org/10.1111/febs.15097>.
- (44) Johnson, S. R.; Bhat, W. W.; Sadre, R.; Miller, G. P.; Garcia, A. S.; Hamberger, B. Promiscuous Terpene Synthases from *Prunella Vulgaris* Highlight the Importance of Substrate and Compartment Switching in Terpene Synthase Evolution. *New Phytologist* **2019**, *223* (1), 323–335. <https://doi.org/10.1111/nph.15778>.
- (45) Andersen-Ranberg, J.; Kongstad, K. T.; Nielsen, M. T.; Jensen, N. B.; Pateraki, I.; Bach, S. S.; Hamberger, B.; Zerbe, P.; Staerk, D.; Bohlmann, J.; Møller, B. L.; Hamberger, B. Expanding the Landscape of Diterpene Structural Diversity through Stereochemically Controlled Combinatorial Biosynthesis. *Angewandte Chemie International Edition* **2016**, *55* (6), 2142–2146. <https://doi.org/10.1002/anie.201510650>.
- (46) Jia, M.; Mishra, S. K.; Tufts, S.; Jernigan, R. L.; Peters, R. J. Combinatorial Biosynthesis and the Basis for Substrate Promiscuity in Class I Diterpene Synthases. *Metab Eng* **2019**, *55*, 44–58. <https://doi.org/10.1016/j.ymben.2019.06.008>.
- (47) Klein, A. P.; Sattely, E. S. Two Cytochromes P450 Catalyze S-Heterocyclizations in Cabbage Phytoalexin Biosynthesis. *Nat Chem Biol* **2015**, *11* (11), 837–839. <https://doi.org/10.1038/nchembio.1914>.
- (48) Klein, A. P.; Sattely, E. S. Biosynthesis of Cabbage Phytoalexins from Indole Glucosinolate. *Proceedings of the National Academy of Sciences* **2017**, *114* (8), 1910–1915. <https://doi.org/10.1073/pnas.1615625114>.
- (49) Miller, G. P.; Bhat, W. W.; Lanier, E. R.; Johnson, S. R.; Mathieu, D. T.; Hamberger, B. The Biosynthesis of the Anti-Microbial Diterpenoid Leubethanol in *Leucophyllum Frutescens*

Proceeds via an All-Cis Prenyl Intermediate. *The Plant Journal* **2020**, *104* (3), 693–705. <https://doi.org/10.1111/tpj.14957>.

(50) Pateraki, I.; Andersen-Ranberg, J.; Hamberger, B.; Heskes, A. M.; Martens, H. J.; Zerbe, P.; Bach, S. S.; Møller, B. L.; Bohlmann, J.; Hamberger, B. Manoyl Oxide (13R), the Biosynthetic Precursor of Forskolin, Is Synthesized in Specialized Root Cork Cells in *Coleus Forskohlii*. *Plant Physiology* **2014**, *164* (3), 1222–1236. <https://doi.org/10.1104/pp.113.228429>.

(51) Pateraki, I.; Andersen-Ranberg, J.; Jensen, N. B.; Wubshet, S. G.; Heskes, A. M.; Forman, V.; Hallström, B.; Hamberger, B.; Motawia, M. S.; Olsen, C. E.; Staerk, D.; Hansen, J.; Møller, B. L.; Hamberger, B. Total Biosynthesis of the Cyclic AMP Booster Forskolin from *Coleus Forskohlii*. *eLife* **2017**, *6*, e23001. <https://doi.org/10.7554/eLife.23001>.

(52) Bedewitz, M. A.; Jones, A. D.; D'Auria, J. C.; Barry, C. S. Tropinone Synthesis via an Atypical Polyketide Synthase and P450-Mediated Cyclization. *Nat Commun* **2018**, *9*, 5281. <https://doi.org/10.1038/s41467-018-07671-3>.

(53) Hamilton, J. P.; Godden, G. T.; Lanier, E.; Bhat, W. W.; Kinser, T. J.; Vaillancourt, B.; Wang, H.; Wood, J. C.; Jiang, J.; Soltis, P. S.; Soltis, D. E.; Hamberger, B.; Buell, C. R. Generation of a Chromosome-Scale Genome Assembly of the Insect-Repellent Terpenoid-Producing Lamiaceae Species, *Callicarpa Americana*. *GigaScience* **2020**, *9* (9), giaa093. <https://doi.org/10.1093/gigascience/giaa093>.

(54) Nett, R. S.; Lau, W.; Sattely, E. S. Discovery and Engineering of Colchicine Alkaloid Biosynthesis. *Nature* **2020**, *584* (7819), 148–153. <https://doi.org/10.1038/s41586-020-2546-8>.

(55) Min, Y.; Kramer, E. M. Transcriptome Profiling and Weighted Gene Co-Expression Network Analysis of Early Floral Development in *Aquilegia Coerulea*. *Sci Rep* **2020**, *10* (1), 19637. <https://doi.org/10.1038/s41598-020-76750-7>.

(56) Lau, W.; Sattely, E. S. Six Enzymes from Mayapple That Complete the Biosynthetic Pathway to the Etoposide Aglycone. *Science* **2015**, *349* (6253), 1224–1228. <https://doi.org/10.1126/science.aac7202>.

(57) Shen, Q.; Lu, X.; Yan, T.; Fu, X.; Lv, Z.; Zhang, F.; Pan, Q.; Wang, G.; Sun, X.; Tang, K. The Jasmonate-Responsive AaMYC2 Transcription Factor Positively Regulates Artemisinin Biosynthesis in *Artemisia Annua*. *New Phytologist* **2016**, *210* (4), 1269–1281. <https://doi.org/10.1111/nph.13874>.

(58) Guo, Q.; Yoshida, Y.; Major, I. T.; Wang, K.; Sugimoto, K.; Kapali, G.; Havko, N. E.; Benning, C.; Howe, G. A. JAZ Repressors of Metabolic Defense Promote Growth and Reproductive Fitness in *Arabidopsis*. *Proceedings of the National Academy of Sciences* **2018**, *115* (45), E10768–E10777. <https://doi.org/10.1073/pnas.1811828115>.

- (59) Wisecaver, J. H.; Borowsky, A. T.; Tzin, V.; Jander, G.; Kliebenstein, D. J.; Rokas, A. A Global Coexpression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. *Plant Cell* **2017**, *29* (5), 944–959. <https://doi.org/10.1105/tpc.17.00009>.
- (60) Delli-Ponti, R.; Shivhare, D.; Mutwil, M. Using Gene Expression to Study Specialized Metabolism—A Practical Guide. *Frontiers in Plant Science* **2021**, *11*.
- (61) Jozwiak, A.; Sonawane, P. D.; Panda, S.; Garagounis, C.; Papadopoulou, K. K.; Abebie, B.; Massalha, H.; Almekias-Siegl, E.; Scherf, T.; Aharoni, A. Plant Terpenoid Metabolism Co-opts a Component of the Cell Wall Biosynthesis Machinery. *Nat Chem Biol* **2020**, *16* (7), 740–748. <https://doi.org/10.1038/s41589-020-0541-x>.
- (62) Takos, A. M.; Rook, F. Why Biosynthetic Genes for Chemical Defense Compounds Cluster. *Trends in Plant Science* **2012**, *17* (7), 383–388. <https://doi.org/10.1016/j.tplants.2012.04.004>.
- (63) Mao, L.; Kawaide, H.; Higuchi, T.; Chen, M.; Miyamoto, K.; Hirata, Y.; Kimura, H.; Miyazaki, S.; Teruya, M.; Fujiwara, K.; Tomita, K.; Yamane, H.; Hayashi, K.; Nojiri, H.; Jia, L.; Qiu, J.; Ye, C.; Timko, M. P.; Fan, L.; Okada, K. Genomic Evidence for Convergent Evolution of Gene Clusters for Momilactone Biosynthesis in Land Plants. *Proceedings of the National Academy of Sciences* **2020**, *117* (22), 12472–12480. <https://doi.org/10.1073/pnas.1914373117>.
- (64) Zhang, J.; Peters, R. J. Why Are Momilactones Always Associated with Biosynthetic Gene Clusters in Plants? *Proceedings of the National Academy of Sciences* **2020**, *117* (25), 13867–13869. <https://doi.org/10.1073/pnas.2007934117>.
- (65) *Uncovering a miltiradiene biosynthetic gene cluster in the Lamiaceae reveals a dynamic evolutionary trajectory.* <https://www.researchsquare.com> (accessed 2022-06-02). <https://doi.org/10.21203/rs.3.rs-1535494/v1>.
- (66) Hansen, N. L.; Heskes, A. M.; Hamberger, B.; Olsen, C. E.; Hallström, B. M.; Andersen-Ranberg, J.; Hamberger, B. The Terpene Synthase Gene Family in *Tripterygium Wilfordii* Harbors a Labdane-Type Diterpene Synthase among the Monoterpene Synthase TPS-b Subfamily. *Plant J* **2017**, *89* (3), 429–441. <https://doi.org/10.1111/tpj.13410>.
- (67) Gericke, O.; Hansen, N. L.; Pedersen, G. B.; Kjaerulff, L.; Luo, D.; Staerk, D.; Möller, B. L.; Pateraki, I.; Heskes, A. M. Neryleryl Diphosphate Is the Precursor of Serrulatane, Viscidane and Cembrane-Type Diterpenoids in *Eremophila* Species. *BMC Plant Biology* **2020**, *20* (1), 91. <https://doi.org/10.1186/s12870-020-2293-x>.
- (68) Mau, C. J.; West, C. A. Cloning of Casbene Synthase cDNA: Evidence for Conserved Structural Features among Terpenoid Cyclases in Plants. *Proc Natl Acad Sci U S A* **1994**, *91* (18), 8497–8501. <https://doi.org/10.1073/pnas.91.18.8497>.
- (69) Vaughan, M. M.; Wang, Q.; Webster, F. X.; Kiemle, D.; Hong, Y. J.; Tantillo, D. J.; Coates, R. M.; Wray, A. T.; Askew, W.; O'Donnell, C.; Tokuhisa, J. G.; Tholl, D. Formation of

the Unusual Semivolatile Diterpene Rhizathalene by the *Arabidopsis* Class I Terpene Synthase TPS08 in the Root Stele Is Involved in Defense against Belowground Herbivory[W]. *Plant Cell* **2013**, 25 (3), 1108–1125. <https://doi.org/10.1105/tpc.112.100057>.

(70) Sallaud, C.; Rontein, D.; Onillon, S.; Jabès, F.; Duffé, P.; Giacalone, C.; Thoraval, S.; Escoffier, C.; Herbette, G.; Leonhardt, N.; Causse, M.; Tissier, A. A Novel Pathway for Sesquiterpene Biosynthesis from Z,Z-Farnesyl Pyrophosphate in the Wild Tomato *Solanum Habrochaites*. *The Plant Cell* **2009**, 21 (1), 301–317. <https://doi.org/10.1105/tpc.107.057885>.

(71) Schilmiller, A. L.; Schauvinhold, I.; Larson, M.; Xu, R.; Charbonneau, A. L.; Schmidt, A.; Wilkerson, C.; Last, R. L.; Pichersky, E. Monoterpenes in the Glandular Trichomes of Tomato Are Synthesized from a Neryl Diphosphate Precursor Rather than Geranyl Diphosphate. *Proceedings of the National Academy of Sciences* **2009**, 106 (26), 10865–10870. <https://doi.org/10.1073/pnas.0904113106>.

(72) Zi, J.; Matsuba, Y.; Hong, Y. J.; Jackson, A. J.; Tantillo, D. J.; Pichersky, E.; Peters, R. J. Biosynthesis of Lycosantalanol, a Cis-Prenyl Derived Diterpenoid. *J. Am. Chem. Soc.* **2014**, 136 (49), 16951–16953. <https://doi.org/10.1021/ja508477e>.

(73) Durairaj, J.; Di Girolamo, A.; Bouwmeester, H. J.; de Ridder, D.; Beekwilder, J.; van Dijk, A. DJ. An Analysis of Characterized Plant Sesquiterpene Synthases. *Phytochemistry* **2019**, 158, 157–165. <https://doi.org/10.1016/j.phytochem.2018.10.020>.

(74) Guengerich, F. P. Mechanisms of Cytochrome P450-Catalyzed Oxidations. *ACS Catal.* **2018**, 8 (12), 10964–10976. <https://doi.org/10.1021/acscatal.8b03401>.

(75) Morikawa, T.; Mizutani, M.; Aoki, N.; Watanabe, B.; Saga, H.; Saito, S.; Oikawa, A.; Suzuki, H.; Sakurai, N.; Shibata, D.; Wadano, A.; Sakata, K.; Ohta, D. Cytochrome P450 CYP710A Encodes the Sterol C-22 Desaturase in *Arabidopsis* and Tomato. *The Plant Cell* **2006**, 18 (4), 1008–1022. <https://doi.org/10.1105/tpc.105.036012>.

(76) Morrone, D.; Chen, X.; Coates, R. M.; Peters, R. J. Characterization of the Kaurene Oxidase CYP701A3, a Multifunctional Cytochrome P450 from Gibberellin Biosynthesis. *Biochemical Journal* **2010**, 431 (3), 337–347. <https://doi.org/10.1042/BJ20100597>.

(77) Yamamoto, H.; Katano, N.; Ooi, A.; Inoue, K. Secologanin Synthase Which Catalyzes the Oxidative Cleavage of Loganin into Secologanin Is a Cytochrome P450. *Phytochemistry* **2000**, 53 (1), 7–12. [https://doi.org/10.1016/S0031-9422\(99\)00471-9](https://doi.org/10.1016/S0031-9422(99)00471-9).

(78) Pesnot, T.; Gershtater, M. C.; Ward, J. M.; Hailes, H. C. Phosphate Mediated Biomimetic Synthesis of Tetrahydroisoquinoline Alkaloids. *Chem. Commun.* **2011**, 47 (11), 3242–3244. <https://doi.org/10.1039/C0CC05282E>.

(79) Ilari, A.; Franceschini, S.; Bonamore, A.; Arengi, F.; Botta, B.; Macone, A.; Pasquo, A.; Bellucci, L.; Boffi, A. Structural Basis of Enzymatic (S)-Norcoclaurine Biosynthesis *. *Journal of Biological Chemistry* **2009**, 284 (2), 897–904. <https://doi.org/10.1074/jbc.M803738200>.

(80) Minami, H.; Dubouzet, E.; Iwasa, K.; Sato, F. Functional Analysis of Norcoclaurine Synthase in *Coptis Japonica*. *J Biol Chem* **2007**, *282* (9), 6274–6282. <https://doi.org/10.1074/jbc.M608933200>.

CHAPTER 2

Elucidating the Biosynthetic Pathway to the Antimicrobial Diterpenoid Leubethanol in *Leucophyllum frutescens*

This chapter is adapted from its original publication in *The Plant Journal*:

Miller, G. P.; Bhat, W. W.; Lanier, E. R.; Johnson, S. R.; Mathieu, D. T.; Hamberger, B. The Biosynthesis of the Anti-Microbial Diterpenoid Leubethanol in *Leucophyllum Frutescens* Proceeds via an All-Cis Prenyl Intermediate. *The Plant Journal* **2020**, *104* (3), 693–705. <https://doi.org/10.1111/tpj.14957>.

Abstract

Serrulatane diterpenoids are natural products found in plants from a subset of genera within the figwort family (Scrophulariaceae). Many of these compounds have been characterized as having antimicrobial properties and share a common diterpene backbone. One example, leubethanol from Texas sage (*Leucophyllum frutescens*), has demonstrated activity against multi-drug-resistant tuberculosis. Leubethanol is the only serrulatane diterpenoid identified from this genus, however a range of such compounds have been found throughout the closely related *Eremophila* genus. Despite their potential therapeutic relevance, the biosynthesis of serrulatane diterpenoids has not been previously reported. Here we leverage the simple product profile and high accumulation of leubethanol in the roots of *L. frutescens* and compare tissue-specific transcriptomes with existing data from *Eremophila serrulata* to decipher the biosynthesis of leubethanol. A short-chain *cis*-prenyl transferase (*Lf*CPT1) first produces the rare diterpene precursor nerylneryl diphosphate, which is cyclized by an unusual plastidial terpene synthase (*Lf*TPS1) into the characteristic serrulatane diterpene backbone. Final conversion to leubethanol is catalyzed by a cytochrome P450 (CYP71D616) of the CYP71 clan. This pathway documents the presence of a short-chain *cis*-prenyl diphosphate synthase, previously only found in Solanaceae, which is likely involved in the biosynthesis of other known diterpene backbones in *Eremophila*. *Lf*TPS1 represents neofunctionalization of a compartment-switching terpene synthase accepting a novel substrate in the plastid. Biosynthetic access to leubethanol will enable pathway discovery to more complex serrulatane diterpenoids which share this common starting structure and provide a platform for the production and diversification of this class of promising antimicrobial therapeutics in heterologous systems.

Significance Statement

Serrulatane diterpenoids are natural products known for their antimicrobial activities, and access is currently limited to chemical synthesis or extraction from natural sources. Here we report the full biosynthetic pathway to the serrulatane diterpenoid leubethanol from *Leucophyllum frutescens*, which is active against multi-drug-resistant tuberculosis. The pathway involves an uncommon diterpene precursor, and further steps yield the archetypal diterpenoid structure shared across nearly all serrulatanes, which may enable development of a new class of antimicrobial therapeutics.

Introduction

Terpenoids are a major class of specialized metabolites in plants, with applications ranging from fragrances and cosmetics to pesticides and pharmaceuticals. This wide variety of uses can be attributed to the incredible structural diversity of terpenoid compounds, resulting from sequential and combinatorial modifications of common starting molecules. Typically beginning with three common C10, C15 and C20 *trans*-prenyl diphosphate substrates, hundreds of mono-, sesqui-, and diterpene backbones are cyclized by terpene synthases (TPSs)¹⁻³. These backbones are further diversified to thousands of terpenoids⁴ through successive modification by enzymes such as cytochrome P450 mono-oxygenases, aldehyde dehydrogenases, and acetyl transferases⁵⁻⁷. Diterpenoids (C20) make up more than 13,000 known plant terpenoids³, and the vast majority with known biosynthetic pathways are derived from all-*trans* (*E,E,E*)-geranylgeranyl diphosphate (GGPP). A given diterpene backbone can be the source of anywhere from one to hundreds of

diterpenoids following downstream modification³, and the biosynthetic routes for the majority of these backbones remain unknown³.

The serrulatane diterpenoids are one such example of a range of compounds derived from a single diterpene backbone (Figure 2.1), with more than thirty identified within the Scrophulariaceae family (order Lamiales) (DNP v28.2). Many of these compounds have been shown to be bioactive. Leubethanol from *Leucophyllum frutescens* is active against multi-drug-resistant tuberculosis⁸, biflorin from *Capraria biflora* and *Eremophila neglecta* has both antitumor⁹ and antimicrobial¹⁰ properties, and microthecalin A from *Eremophila microtheca* is active against malaria¹¹. While relatively few have been identified in other genera, the *Eremophila* genus is especially rich in these antimicrobial compounds^{12–16} and at least three are found in the *Myoporum* genus¹⁷. Given their promise in therapeutic applications, there has been a substantial effort to devise total chemical syntheses^{18–23} as an alternative to extraction and purification from natural sources.

Despite the efforts invested into natural product discovery, antimicrobial screens, and total chemical syntheses, the biosynthetic pathway to these serrulatane diterpenoids has remained elusive. Identifying the enzymes responsible would pave the way for production of these serrulatane diterpenoids in heterologous systems, offering an appealing alternative to formal chemical synthesis. To address this, we sought to elucidate the biosynthetic pathway to the serrulatane diterpenoid leubethanol in *L. frutescens*. Three properties of this species made it an ideal target for studying this pathway. First, leubethanol is the only serrulatane diterpenoid known to be produced by this plant and accumulates in high quantities in root tissue⁸. Second, RNA-seq data is publicly available for the closely related species *E. serrulata*²⁴, allowing for comparative transcriptomics between genera. Third, leubethanol shares a common hydroxylation with the

majority of other known serrulatane diterpenoids (Figure 2.1) and is the likely intermediate in their biosynthesis. Beyond these advantages, leubethanol is itself an appealing target, with activity against multi-drug-resistant tuberculosis (minimum inhibitory concentration 22 μ M) comparable to isoniazid (23 μ M) and ethambutol (39 μ M)⁸, two drugs commonly used in combination therapy.

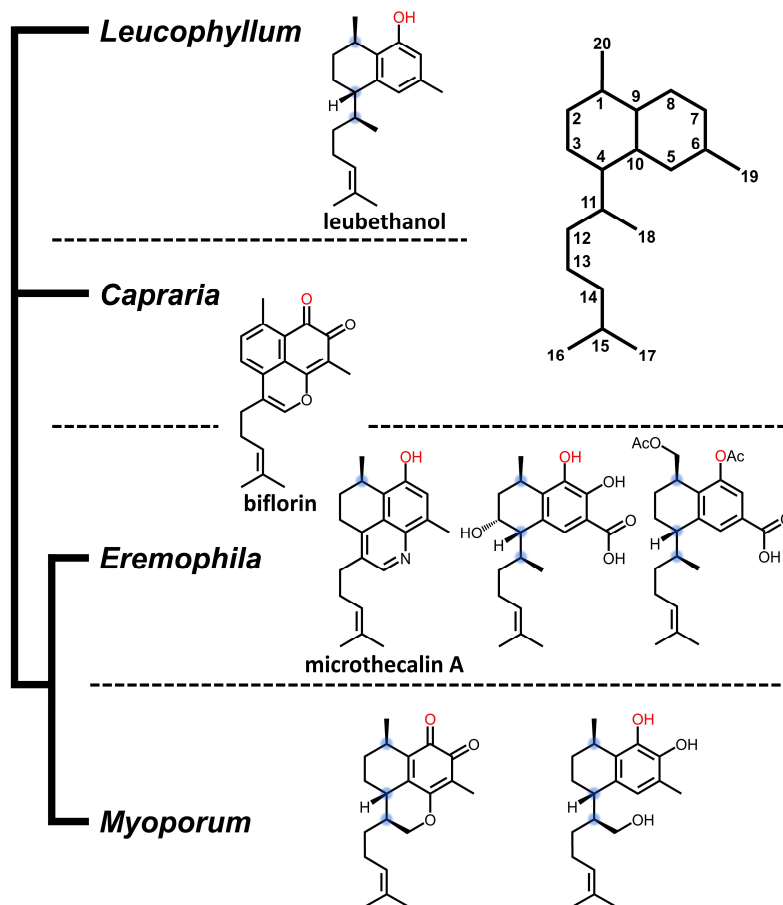


Figure 2.1: Distribution of serrulatane diterpenoids in members of the Scrophulariaceae family. Shared backbone structure in top right. Leubethanol contains common stereochemistry highlighted in blue, and common oxygenation highlighted in red. Only one serrulatane diterpenoid each has been identified in *Leucophyllum* (leubethanol) and *Capraria* (biflorin), while only a few representatives are shown for the *Eremophila* and *Myoporium* genera; *Eremophila* alone harbors more than thirty.

Through comparing transcriptomes between tissue types and genera, we have identified three enzymes which constitute the full biosynthetic pathway to leubethanol in *L. frutescens*. While the vast majority of known diterpenoids originate from GGPP, this pathway involves a short-chain

cis-prenyl transferase (*cis*-PT) which produces the uncommon diterpene precursor (Z,Z,Z)-neryleryldiphosphate (NNPP – the all-*cis* stereoisomer of GGPP). This is then cyclized to the shared serrulatane backbone by a terpene synthase (TPS) which exclusively converts NNPP and is a member of the primarily sesquiterpene (C15) synthase TPS-a subfamily. Finally, this backbone is converted to leubethanol by a cytochrome P450 of the CYP71 clan, which harbors many recently discovered P450s involved in terpene specialized metabolism^{25,26}. The identification of a short-chain *cis*-PT which makes NNPP clarifies the likely origin of other diterpenes identified in the *Eremophila* genus based on the presence of *cis*-double bonds in these backbones. Reconstruction of the full pathway in *Nicotiana benthamiana* allowed for production of leubethanol in a heterologous system and provides access to a plausible key intermediate in the biosynthesis of other serrulatane diterpenoids.

Results

Accumulation of leubethanol guided tissue-specific RNA sequencing.

To begin our search for the biosynthetic pathway to leubethanol, we took advantage of its tissue-specific accumulation in *L. frutescens*. Previous work on the medicinal properties of this species has shown that root extracts were most potent against multi-drug-resistant tuberculosis, while leaves showed some activity and flowers showed none²⁷. To confirm the tissue-specific accumulation of leubethanol, extracts of the leaves, roots, and flowers were analyzed by GC-MS. Leubethanol was found to accumulate in both root and leaf tissue, while none was detected in flower tissue (Figure 2.S1). Consequently, we isolated and sequenced RNA from both the roots and flowers to allow for comparative transcriptomics between tissue types. Serrulatane diterpenoids are also found in the closely related *Eremophila* genus²⁸. RNA-seq data are publicly

available from the leaves of *E. serrulata* (SRA: ERX1321488²⁴), and serrulatanes are known to accumulate in this tissue¹². These data were also included to allow for comparison between genera.

Identification of TPS candidates from L. frutescens.

We began our search by identifying TPS candidates from *L. frutescens* through a homology-based search of our transcriptomic data against a reference set of TPSs. Fifteen candidates were identified, and a phylogenetic tree was constructed to group each candidate by TPS subfamily (Figure 2.2). One candidate (*Lf*TPS13) was not expressed in root tissue and was eliminated from further consideration.

While containing a bicyclic decalin core, the structure of leubethanol is inconsistent with the labdane group of plant diterpenoids, the most common type of backbone which results from cyclization by pairs of class II and class I diTPS²⁹. In contrast, the cyclization pattern of leubethanol indicates activity of a class I enzyme, which catalyzes cyclization via removal of the diphosphate moiety. Out of the fourteen root-expressed candidates, only one was predicted to be a class II TPS (*Lf*TPS4; TPS-c subfamily), and therefore thirteen possibilities remained.

A number of non-labdane diterpenes have been shown previously to be made by TPS-a enzymes which are localized to the plastid^{5,7,30–35}. The majority of TPS-a enzymes are sesquiterpene synthases localized to the cytosol³⁶, and the presence of an N-terminal plastidial transit peptide in the primary amino acid sequence can therefore aid in prediction of diterpene synthase activity in this subfamily⁷. Two *L. frutescens* candidates (*Lf*TPS1 and *Lf*TPS2) in the TPS-a subfamily were found to carry N-terminal extensions. Additionally, both have an ortholog in *E. serrulata* with nearly identical sequence length and homology through these N-terminal

extensions (Figure 2.S2). Of these two candidates, only *Lf*TPS1 is exclusively expressed in root tissue and was therefore considered the more likely candidate, however both were tested.

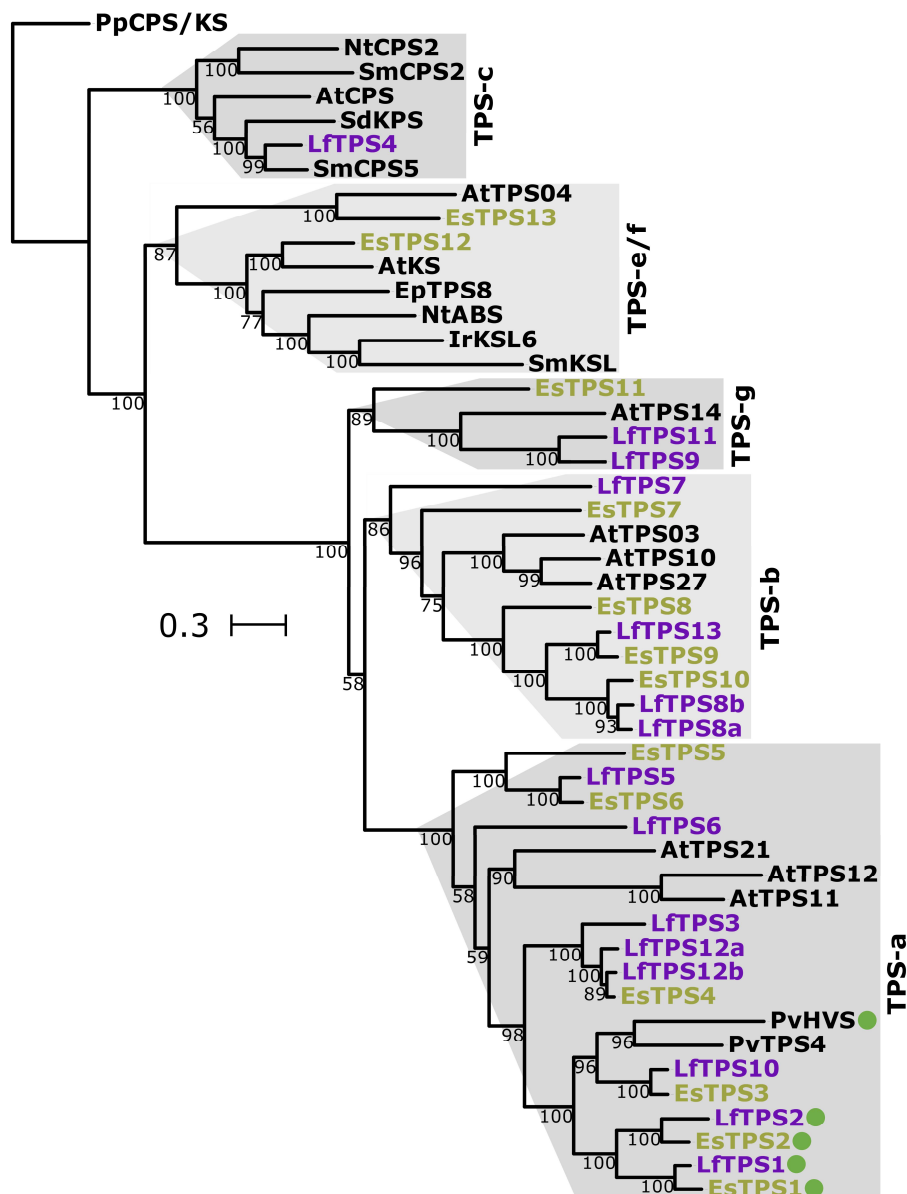


Figure 2.2: Maximum likelihood phylogenetic tree of TPS candidates. Candidate terpene synthases from *L. frutescens* are shown in purple and *E. serrulata* in yellow, with reference TPSs in black. Putative transit peptides, predicted by N-terminal extensions, are denoted within the TPS-a subfamily by green dots. Scale bar represents substitutions per site, and branch numbers represent percent support from 1,000 bootstrap replicates. The bifunctional *ent*-CPP/*ent*-kaurene synthase from *Physcomitrella patens* (PpCPS/KS) is used as an outgroup. Abbreviations: *Pp*, *Physcomitrella patens*; *Nt*, *Nicotiana tabacum*; *Sm*, *Salvia miltiorrhiza*; *At*, *Arabidopsis thaliana*; *Sd*, *Salvia divinorum*; *Ep*, *Euphorbia peplus*; *Ir*, *Isodon rubescens*; *Pv*, *Prunella vulgaris*.

Full-length genes for both *Lf*TPS1 and *Lf*TPS2 were cloned from root cDNA for transient expression in an *N. benthamiana* system engineered for increased levels of the presumed substrate GGPP³⁷. N-terminal truncated constructs, removing the putative transit peptides, were cloned into pET-28b(+) for expression of pseudomature variants in *E. coli*. Assays were extracted with hexane and analyzed by GC-MS.

To account for uncertainty of the predicted plastidial targeting signals, transient expression assays in *N. benthamiana* were carried out separately with co-expression of either plastidial or cytosolic GGPP terpene precursor pathway enzymes. Co-expression of both candidates with either cytosolic or plastidial precursor enzymes did not yield detectable products (Figure 2.S3). To independently verify activity, each enzyme was expressed in *E. coli* with a C-terminal histidine tag and purified through Ni-affinity chromatography. Consistent with the results of the transient *N. benthamiana* assays, incubation of both *Lf*TPS1 and *Lf*TPS2 with GGPP in *in vitro* assays yielded no measurable activity. Additionally, no activity was seen when incubated with farnesyl diphosphate (FPP, precursor for sesquiterpenes) or geranyl diphosphate (GPP, precursor for monoterpenes) (Figure 2.S3).

***Lf*TPS1 exclusively cyclizes neryl neryl diphosphate into the serrulatane backbone.**

Following these results, we considered two routes forward: first, to expand testing to each other class I TPS candidate, and second, to test *Lf*TPS1 and *Lf*TPS2 against uncommon terpene precursors. The former route was considered because even very closely related TPSs can have activities which differ substantially², and there are many examples of TPSs which have different functions than would be predicted by their subfamily⁷. The latter route was considered because of the absence of activity against each common substrate. GPP, FPP, and GGPP contain exclusively

trans double bonds. All-*cis* stereoisomers of each have been reported in members of the nightshade (Solanaceae) family³⁸, together with TPSs which can convert these to terpene products^{39–42}.

The serrulatane backbone is ambiguous with respect to the original stereochemistry of its precursor; however, closer inspection of diterpenoids from the *Eremophila* genus shows that acyclic, bisabolane, and cembrane type diterpenoids (Figure 2.3A) in various *Eremophila* species contain internal *cis* double bonds²⁸. This prompted us to test NNPP (the all-*cis* stereoisomer of GGPP) as the precursor for the serrulatane backbone in *L. frutescens*. Since NNPP is not commercially available, truncated constructs of *Lf*TPS1 and *Lf*TPS2 in pET-28b(+) were used for co-expression with *Sf*CPT2, the plastidial *S. lycopersicum cis*-PT³⁸, in an *E. coli* system engineered to increase terpene precursor availability⁴³. Following hexane extraction and analysis by GC-MS, *Lf*TPS1 was found to convert NNPP (Figure 2.S4). This activity was independently confirmed in *N. benthamiana* (Figure 2.3B and Figure 2.S4).

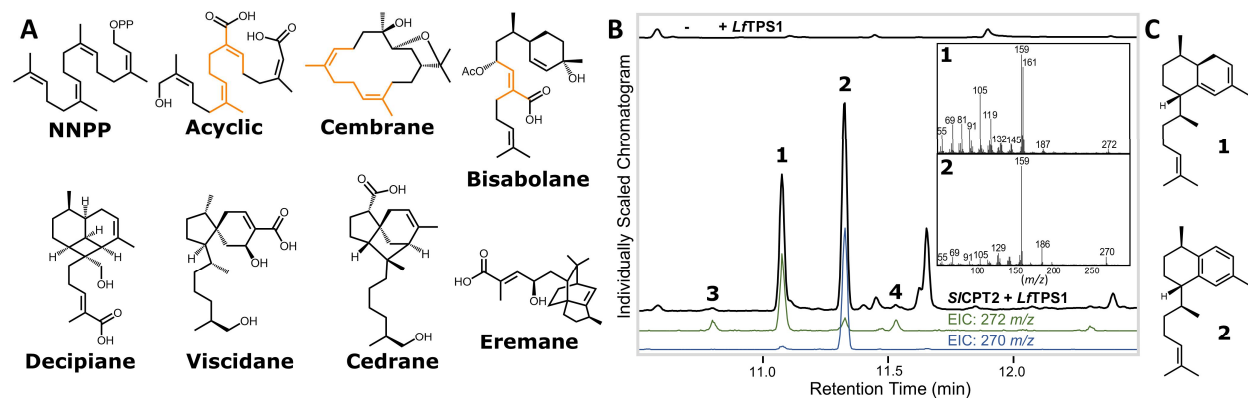


Figure 2.3: Dihydroserrulatene production by *Lf*TPS1. (A) Structure of NNPP and representative non-serrulatane diterpenoids from *Eremophila* with labeled backbone structure types. Isoprenyl subunits found in acyclic, cembrane, and bisabolane type diterpenoids, which can be used to infer the stereochemistry of their prenyl diphosphate precursor, are highlighted in orange. (B) Transient expression of *Lf*TPS1 in *N. benthamiana* with *Sf*CPT2 (a plastidial NNPP synthase), and mass spectra of major products. Each assay has *Cf*DXS co-expressed in addition to those listed. (C) Structures of dihydroserrulatene (**1**) and serrulatene (**2**).

Four diterpene products were observed, with only one major product (**1**) in the *E. coli* system, and a relative amount of **2** exceeding **1** in *N. benthamiana*. Diterpene olefins typically

have a molecular ion of 272 m/z , however **2** has a molecular ion of 270 m/z . The fragmentation pattern for **2** is consistent with an aromatic product, and is similar to that of leubethanol (286 m/z) with major peaks shifted by 16, consistent with a difference of one hydroxylation (Figure 2.S1). Given that TPSs are not known to catalyze redox reactions, **2** is likely derived from spontaneous aromatization of the major product **1**, a phenomenon seen previously in diterpene biosynthesis⁴⁴. To confirm the structure of **1**, production in the *E. coli* system was scaled up for NMR analysis (Table 2.S1 and Figure 2.S5), revealing that *Lf*TPS1 makes dihydroserrulatene (peak **1**), and supporting the identity of peak **2** as serrulatene (Figure 2.3C). A proposed mechanism for *Lf*TPS1 conversion of NNPP to dihydroserrulatene is given in Scheme 2.S1.

In parallel to the testing against NNPP, we began working towards testing the remaining class I candidate TPSs. While we cloned each of these candidates out of *L. frutescens* cDNA, we received the positive results for *Lf*TPS1 conversion of NNPP to dihydroserrulatene before we characterized these other candidates. These were, however, cloned and sequence verified, and are given here with GenBank accession numbers for reference.

LfCPT1, a short chain cis-prenyl transferase, supplies NNPP in serrulatane biosynthesis.

We next sought out the source of NNPP in *L. frutescens* by searching for a *cis*-prenyl transferase. *Cis*-PTs are ubiquitous throughout plants and are typically involved in the synthesis of long chain polyisoprenoids³⁸, although very few which make short chain products (fewer than 35 carbons) have been identified. Three short-chain *cis*-PTs which yield NPP (neryl diphosphate; 10 carbon), (*Z-Z*)-FPP (*Z-Z*-farnesyl diphosphate; 15 carbon), and NNPP (20 carbon) have been identified from *Solanum lycopersicum* through functional characterization of the entire family of *cis*-PTs from this species³⁸. We identified candidate *cis*-PTs from both the *L. frutescens* and *E. serrulata* transcriptomes through a homology-based search against the entire family of *cis*-PTs

from *S. lycopersicum*. Ten candidate *cis*-PTs were identified from *L. frutescens*, and phylogenetic analysis revealed that six are closely related to the short-chain *cis*-PTs from *S. lycopersicum* (Figure 2.4). *LfTPS1* has a predicted plastidial transit peptide, and successfully converts NNPP in *N. benthamiana* assays when co-expressed with *SlCPT2*, which is known to be targeted to the plastid³⁸. Therefore, we looked for a *cis*-PT candidate that is likely targeted to the plastid. Three of these candidates were found to carry predicted plastidial transit peptides and are expressed in root tissue (*LfCPT1-3*). *LfCPT1* was considered to be the most likely candidate as it is the only of these three to have a direct ortholog in our *E. serrulata* transcriptome assembly (*EsCPT1*), however all three were tested.

LfCPT1-3 were cloned from *L. frutescens* root cDNA. Each candidate *cis*-PT was co-expressed in *N. benthamiana* with *LfTPS1*, and products were analyzed by GC-MS following hexane extraction. Co-expression with *LfCPT1* yielded the same diterpene product profile as with the NNPP synthase from *S. lycopersicum* (*SlCPT2*) (Figure 2.4 and Figure 2.S6). In addition, direct comparison of *LfCPT1* with *SlCPT2* without co-expression of a TPS showed the same peak and mass spectrum for dephosphorylated NNPP (Figure 2.4).

A cytochrome P450 converts the serrulatane backbone to leubethanol.

Leubethanol is oxidized twice relative to dihydroserrulatene, presumably through hydroxylation by a cytochrome P450 and aromatization. Given the propensity for dihydroserrulatene to spontaneously aromatize to serrulatene, we set out to identify P450 candidates for the required oxidation at C8. A homology-based search of both the *L. frutescens*

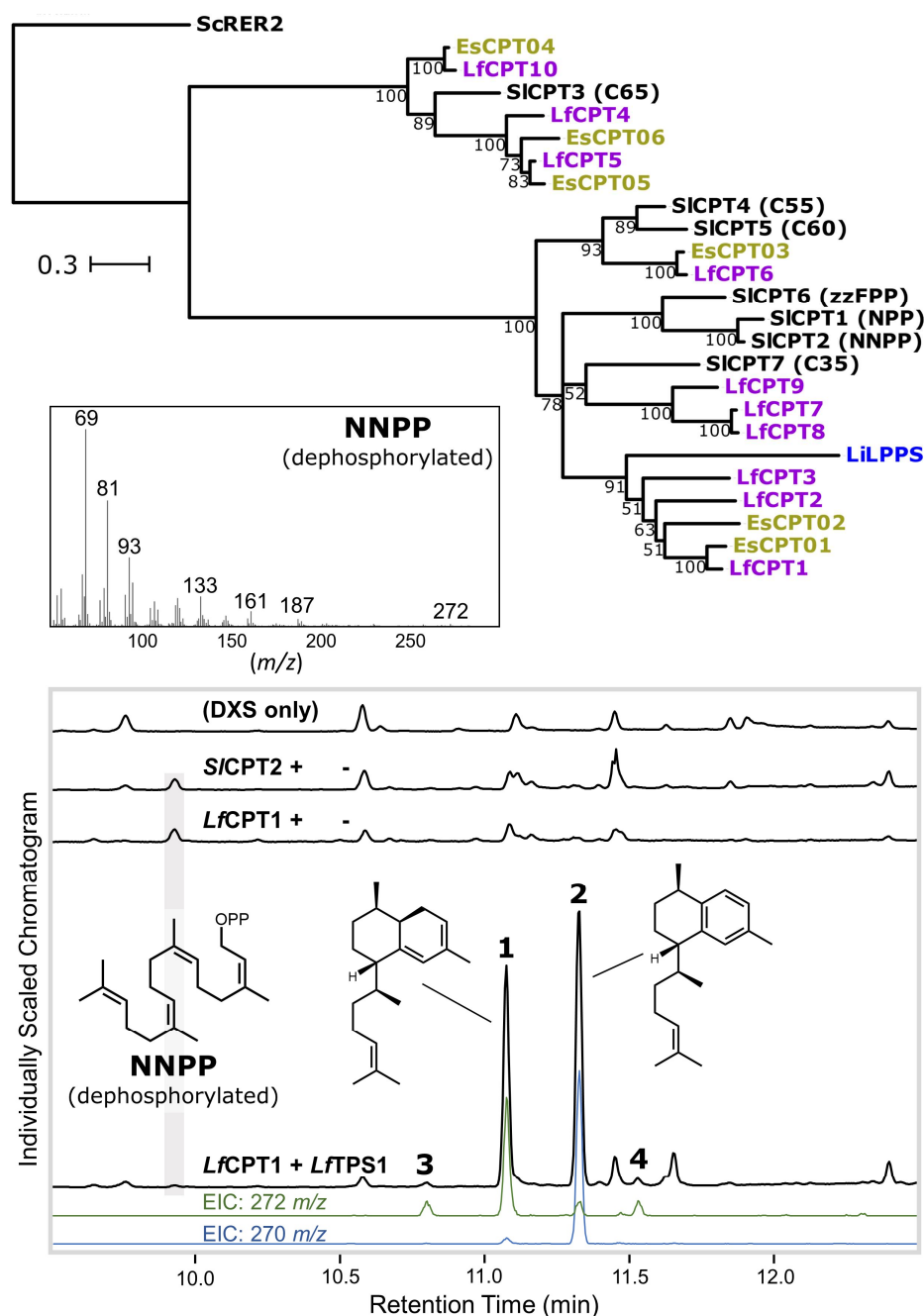


Figure 2.4: *NNPP* production by *LfCPT1*. (Top) Maximum likelihood phylogenetic tree of *cis*-prenyl transferases from *L. frutescens* (purple) and *E. serrulata* (yellow) transcriptome assemblies, and *S. lycopersicum* (black) with products in parentheses. An unusual head-to-middle condensation *cis*-PT from *Lavandula x intermedia* (*LiLPPS*; blue) is included. Scale bar represents substitutions per site, and branch numbers represent percent support from 1,000 bootstrap replicates; branches with less than 50% support have been collapsed. The *cis*-PT *ScRER2* from *Saccharomyces cerevisiae* is used as an outgroup. (Bottom) GC-MS chromatograms for *N. benthamiana* assay of *LfCPT1* function, both alone and in combination with *LfTPS1*, and mass spectrum (70 eV EI) of dephosphorylated NNPP found in the highlighted region (gray) of each sample except DXS control. Each assay has *CfDXS* co-expressed in addition to those listed.

and *E. serrulata* transcriptomes was carried out against a reference set of plant P450s. 165 candidates were identified from *L. frutescens*. We first narrowed our search by focusing on those within the CYP71 clan. While P450s in other clans have been identified in diterpenoid specialized metabolism, we began our search here based on the CYP71 clan containing the majority of previously characterized examples^{25,26}. Clustering each P450 candidate by family and eliminating those outside of the CYP71 clan reduced the list of candidates to 59 (Figure 2.S7). Considering only those that were expressed in root tissue but not flower tissue, and those that had an ortholog in our *E. serrulata* transcriptome assembly, only five candidates remained. One additional candidate (CYP71D615), which did not have a direct ortholog in *E. serrulata*, was included based on its root-exclusive expression and location among a cluster of other *L. frutescens* and *E. serrulata* candidates in the phylogenetic tree (Figure 2.S7).

These six P450 candidates were cloned from *L. frutescens* root cDNA. Co-expression with *LfCPT1* and *LfTPS1* in *N. benthamiana* revealed that CYP71D616 facilitates the conversion of dihydroserrulatene to leubethanol (Figure 2.5 and Figure 2.S8). A relative decrease of dihydroserrulatene over serrulatene indicates that the preferred substrate for CYP71D616 is dihydroserrulatene. The observed minor reduction in serrulatene is plausibly due to P450-mediated turnover of dihydroserrulatene preceding spontaneous aromatization. This is supported by the metabolomic data from root tissue extracts, which shows an accumulation of serrulatene but no detectible quantities of dihydroserrulatene (Figure 2.5, Figure 2.S1).

The interdependence of each enzyme in the pathway is demonstrated (Figures 4 and 5), showing that all three are necessary for leubethanol production when expressed in *N. benthamiana*. To determine whether the TPS activity is conserved in the *Eremophila* genus, we tested a synthetic homolog of *LfTPS1* (*EsTPS1*; 85% amino acid identity) from the *E. serrulata* transcriptome

assembly. Replacing *Lf*TPS1 with *Es*TPS1 yields the same products in each combination (Figure 2.S9), demonstrating orthology between the enzymes and conservation of this pathway in the serrulatane-rich *Eremophila* genus.

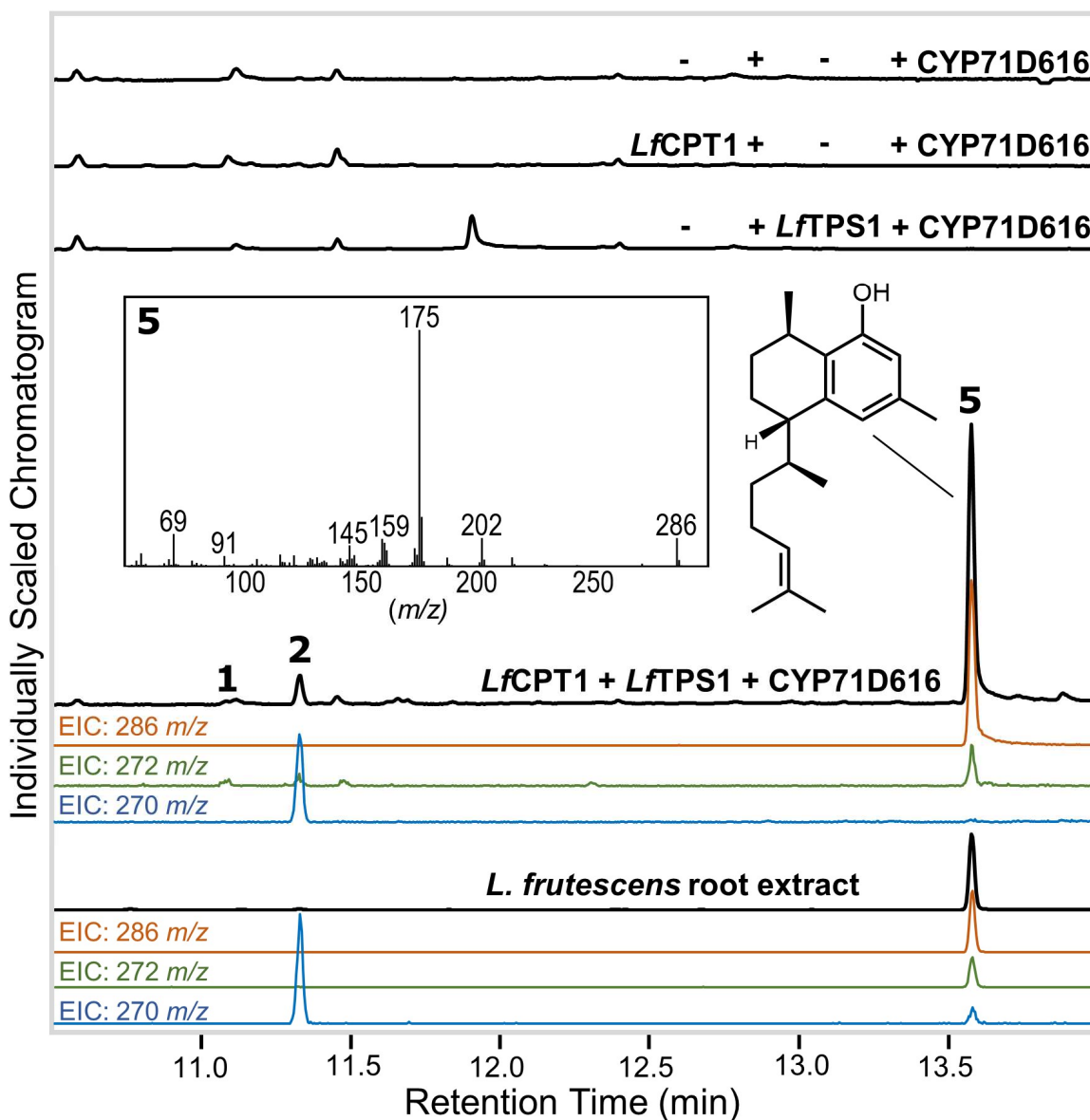


Figure 2.5: Leubethanol production by CYP71D616. GC-MS chromatograms of CYP71D616 assay in *N. benthamiana* and *L. frutescens* root extract, with mass spectrum (70 eV EI) of leubethanol (5) from heterologous expression of all three enzymes in the pathway. Total ion chromatograms are shown in black; each assay has *Cf*DXS co-expressed in addition to those listed.

Discussion

Through comparative transcriptomics between tissue types and genera, we have identified three enzymes responsible for the biosynthesis of the serrulatane diterpenoid leubethanol in *L. frutescens*. The stereochemistry at all three chiral centers in dihydroserrulatene matches that of every serrulatane diterpenoid identified from the Scrophulariaceae family wherever the stereocenter is retained in the final diterpenoid product (See examples in Figure 2.1). This, and the conserved function between *Lf*TPS1 and *Es*TPS1, suggest that dihydroserrulatene is in fact the common precursor to all serrulatanes. In the preparation of this manuscript, Gericke *et al.*⁴⁵ reported a similar pathway to dihydroserrulatene involving a *cis*-PT and plastidial TPS-a in *Eremophila drummondii* and *Eremophila denticulata*, further supporting the conservation of this pathway. Nearly all of the serrulatane diterpenoids in Scrophulariaceae share a common hydroxylation (or derivative thereof) with leubethanol, suggesting that leubethanol itself is a common precursor. Given this commonality, the CYP71D616-catalyzed hydroxylation is likely the entry step between the diterpene backbone and diversification toward other antimicrobial serrulatane diterpenoids from other genera such as biflorin and microthecalin A.

This pathway is unusual in that it involves the all-*cis* prenyl diphosphate precursor NNPP rather than the common diterpene precursor GGPP. Prenyl diphosphate substrates are synthesized by members of either the *trans*- or *cis*-prenyl transferase families, typically in a head-to-tail condensation of the 5-carbon molecules isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP)⁴⁶. These two enzyme families are distinct with no sequence⁴⁷ or structural⁴⁸ homology. The evolution of members of the *cis*-PT family to make uncommon terpene precursors has been found in two other cases, with the series of NPP (*Sl*CPT1), (*Z,Z*)-FPP (*Sl*CPT6), and NNPP (*Sl*CPT2) in *S. lycopersicum* (Solanaceae)³⁸, and lavandulyl diphosphate (head-to-middle

condensation catalyzed by *LiLPPS*) in *Lavandula x intermedia* (Lamiaceae)⁴⁹. *LfCPT1*, *LiLPPS*, and the *S. lycopersicum* short-chain *cis*-PT are phylogenetically closely related when compared to the overall characterized *cis*-PT family in *S. lycopersicum* (Figure 2.4). This may indicate a shared common ancestry of the short-chain *cis*-PTs in Solanaceae, Lamiaceae, and Scrophulariaceae. Scrophulariaceae diverged from Solanaceae between 75 to 88 MYA, and from Lamiaceae between 44 and 67 MYA based on molecular time estimates (timetree.org⁵⁰ and 32 references therein, accessed February 12, 2020), which is consistent with the divergence pattern of the short-chain *cis*-PTs (Figure 2.4): *LfCPT1* appears to be more closely related to *LiLPPS* (Lamiaceae) than any of the *Solanum* *cis*-PTs, despite being closer to the *Solanum* enzymes in product profile. Additionally, it has been suggested that the shorter product length of the *S. lycopersicum* *cis*-PTs may be due in part to a shortened alpha helix not present in the long-chain *cis*-PTs from this species³⁸. This is not present in either *LiLPPS* or *LfCPT1* based on homology modeling (Figure 2.S10) and a sequence alignment (Figure 2.S11), suggesting that the evolution towards smaller precursors is independent and follows different trajectories from an ancestral sequence.

In addition to finding a similar pathway to dihydrosecuratene, Gericke *et al.*⁴⁵ identified TPSs which make the cembrane and viscidane backbones in *Eremophila lucida*, and showed that these exclusively use NNPP over GGPP as well. To identify where the TPSs and *cis*-PTs from these three other *Eremophila* species (*E. denticulata*, *E. drummondii*, and *E. lucida*) lie relative to our candidates, we generated phylogenetic trees including each candidate identified from these species and our sequences (TPS: Figure 2.S12; *cis*-PT: Figure 2.S13). Each other *Eremophila* NNPP synthase is a direct ortholog of *LfCPT1*, while *LfCPT2* and *LfCPT3* have no orthologs in any of these *Eremophila* species (Figure 2.S13). Interestingly, a (Z,Z)-FPP synthase (*EdCPT2*) was found, however a TPS in *Eremophila* which converts (Z,Z)-FPP has yet to be identified⁴⁵. The

cembratrienol synthase (*ElTPS31*) is a member of the TPS-b subfamily, commonly involved in monoterpene synthesis, and *L. frutescens* does not have an ortholog. The hydroxyviscidane synthase (*ElTPS3*) lines up closely with *LfTPS2* and another enzyme from *E. denticulata* (*EdtTPS5*), however neither of these candidates were found to have this same function. Interestingly, more TPS-a candidates which are putatively targeted to the plastid, but do not convert GGPP or NNPP, are present in these three *Eremophila* species. The function of *LfTPS2* and these other plastidial TPS-a enzymes remains to be seen, and may suggest that other precursors that were not taken into account in either study may be present in the plastids of these plants.

The identification of a short-chain *cis*-PT in Scrophulariaceae clarifies the likely origin of other diterpene backbones present in the *Eremophila* genus. Acyclic and bisabolane type diterpenoids identified in this genus contain internal alkenes in *cis* configuration (Figure 2.4). As serrulatanes and viscidanes have now both been shown to be derived from NNPP, it is likely that the decipiane, cycloserrulatane, and cedrane backbones²⁸ are derived from NNPP as well. The backbones for decipianes and cycloserrulatanes resemble a tricyclic serrulatane backbone, and the cedrane backbone resembles a tricyclic viscidane backbone. Beyond Scrophulariaceae, there are hundreds of other diterpene backbones with unknown biosynthetic routes. In Lamiaceae alone there are at least 200³, and in *Salvia sclarea* (Lamiaceae), two previously reported diterpenoids salviatriene A and B⁵¹ resemble a cycloserrulatane and tricyclic viscidane, respectively. Given the independent emergence of *cis*-PTs which yield NNPP in different plant families, it may be that some of these unknown diterpenoid pathways involve NNPP as well.

Numerous diterpene backbones that differ from the more common labdane structure have been shown to be formed by enzymes in the TPS-a subfamily, which is mostly comprised of cytosolic sesquiterpene synthases. *LfTPS1* provides another example of a compartment and

substrate-switching TPS from this subfamily, but differs from these previous examples in that it does not convert GGPP. In contrast to earlier work in *P. vulgaris* (Lamiaceae), where the enzyme *PvHVS* showed acceptance of both GGPP and the presumed non-native NNPP⁷, *LfTPS1* showed a high specificity towards NNPP. *PvTPS5* and *PvTPS2* (both TPS-a) could also convert NNPP to a diterpene product in addition to their native functions as sesquiterpene and diterpene synthases, respectively⁷. This could plausibly arise from negative selection against GGPP, as both substrates are available in *L. frutescens* and presumably only GGPP is available in *P. vulgaris*. The presence of competing substrates in *L. frutescens* may introduce a strong selective pressure for specificity⁵², while the absence of NNPP in *P. vulgaris* means that no such selective pressure exists. Such specificity can also be seen in *Solanum* where these all-*cis* substrates are present, where *PHS1*⁴⁰, *SBS*³⁹, and *SlTPS21*⁴² all showed high specificity towards NPP, (Z,Z)-FPP, and NNPP, respectively compared to their all-*trans* counterparts.

Even some class II diTPSs (TPS-c) have been shown to have promiscuous activities in converting NNPP into irregular labdane structures⁵³. The substrate promiscuity of these TPSs suggests that the evolution of a prenyl transferase to afford an unusual terpene precursor may not require the co-evolution of a TPS, as the ability to convert a novel substrate may already be present in lineages where promiscuity was never selected against. Additionally, the occurrence of TPSs which natively convert *cis*-prenyl substrates is widespread throughout different TPS subfamilies. Examples have now been seen in the TPS e/f (*Solanum* species), TPS-b (*Eremophila lucida*), and TPS-a (*L. frutescens* and three *Eremophila* species) subfamilies (Figure 2.S12), showing that evolution towards specificity for these substrates has happened independently in vastly different lineages of TPSs. Taken together, the presence of uncommon substrates may be more widespread than generally assumed, and the search for biosynthetic routes to new terpene backbones should

involve a consideration of other possible precursors beyond the all-*trans* substrates which are typical.

Materials and Methods

Plant material, RNA Isolation and cDNA synthesis, and metabolite analysis

Leucophyllum frutescens plants were obtained from Stokes Tropicals (Homestead, FL, USA) and grown in a greenhouse under ambient photoperiod and 24°C day/17°C night temperatures. Total RNA from flower, leaf, and root tissues was extracted following methods described in Hamberger *et al.*⁵⁴ using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA). RNA extraction was followed by DNase I digestion using DNA-free™ DNA Removal Kit (ThermoFisher Scientific). Total RNA was assessed for quantity and integrity by Qubit™ (ThermoFisher Scientific) and RNA-nano assays (Agilent Bioanalyzer 2100), prior to whole transcriptome sequencing (Novogene, Sacramento, CA, USA). First-strand cDNA was synthesized from 2 µg of root total RNA using SuperScript III (Invitrogen). For GC-MS-based metabolomics, approximately 1 g of root, leaf, or flower tissue was extracted in 1 mL MTBE for 3 hours and analyzed by GC-MS with the same method described below for analysis of enzyme assays.

L. frutescens and E. serrulata de novo transcriptome assembly and analysis

RNA-seq data were obtained through tissue-specific RNA sequencing on an Illumina HiSeq 4000 for *L. frutescens* and the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra> (ERX1321488)) for *E. serrulata*²⁴. Quality of sequencing data was checked with FastQC (v0.11.4), and adapters were trimmed with Trimmomatic (v0.39⁵⁵). A

transcriptome was assembled with Trinity (v2.8.4⁵⁶), expression levels calculated with Salmon (v0.11.2⁵⁷), and open reading frames picked out with TransDecoder (v5.5.0⁵⁸). A BLAST (v2.7.1+) search against reference databases of respective enzyme families was done to pick out candidates. Phylogenetic trees were made with Clustal Omega (v1.2.4⁵⁹) and RAxML (v8.0.0⁶⁰) and visualized with Interactive Tree of Life⁶¹. Plastidial transit peptides were predicted between TargetP (v 1.1⁶²) and sequence alignments with Clustal Omega (v1.2.4⁵⁹).

Cloning and sources of genes used

Candidate enzymes were PCR-amplified from root cDNA, and coding sequences were cloned through In-Fusion cloning into the plant expression vector pEAQ-HT⁶³ for transient expression assays in *N. benthamiana*, or into pET-28b(+) for expression in *E. coli*. *LfTPS1* and *LfTPS2* were cloned into pET-28b(+) as N-terminal truncated constructs omitting the first 23 amino acid residues, removing their putative transit peptides. For *in vitro* assays, constructs for *PvTPS4*, *PvTPS5*, and *PvHVS*(Δ 43) in pET-28b(+) made in Johnson *et al.*⁷ were used as positive controls. For *in vivo E. coli* assays, the same truncated *LfTPS* constructs described above were used. TPS constructs were co-transformed with pIRS⁶⁴ and pNN⁷.

For all assays in *N. benthamiana*, full-length candidates were cloned into pEAQ-HT. For cytosolic tests, TPS candidates were co-expressed with *Euphorbia lathyris* HMGR and *Methanothermobacter thermautotrophicus* GGPPS⁶⁵ in the pEarlygate vector⁶⁶. As a positive control for cytosolic tests, an N-terminal truncated construct of *PvHVS* (*PvHVS*(Δ 43)) was cloned into pEAQ-HT in this study. For plastidial tests, each candidate was co-expressed with *Coleus forskohlii* DXS³⁷ in pEarlygate. TPS candidate tests involved either co-expression of *C. forskohlii*

GGPPS³⁷ in pEarlygate or *Solanum lycopersicum* CPT2 in pEAQ-HT⁷, with a full-length construct of PvHVS in pEAQ-HT as a positive control⁷.

***In vitro* assays**

TPS expression and purification was carried out as described in Johnson *et al.*⁷. *Lf*TPS1 and *Lf*TPS2 constructs in pET-28b(+) were transformed into the *E.coli* C41 OverExpress strain. Primary cultures (5 mL LB plus 50 µg/mL kanamycin) were grown overnight 37°C, and 1 mL was used to inoculate a bulk culture (100 mL TB plus 50 µg/mL kanamycin). This culture was grown to an OD₆₀₀ of 0.6 at 37°C, and expression was induced with 0.2 mM IPTG. Expression was carried out overnight at 17°C, cells were collected by centrifugation, and resuspended in Buffer A (20 mM HEPES, pH 7.2, 25 mM imidazole, 500 mM NaCl, 5% (v/v) glycerol) plus 10 µL/ml protease inhibitor cocktail (Sigma) and 0.1 mg/ml lysozyme (VWR). Cells were lysed by sonication and centrifuged at 11,000 xg for 30 min. Supernatants were loaded onto Ni-NTA columns (His SpinTrap; GE Healthcare) preequilibrated with Buffer A, washed with two column volumes of Buffer A, and protein was eluted with Buffer B (Buffer A with 350 mM imidazole). Samples were desalted with a PD MidiTrap G-25 column (GE Healthcare) preequilibrated with Buffer C (20 mM HEPES, pH 7.2, 1 mM MgCl₂, 350 mM NaCl, and 5% (v/v) glycerol). Purified enzymes were frozen in liquid nitrogen and stored at - 80°C prior to *in vitro* assays.

In vitro assays were carried out with 1 µM enzyme and 30 µM substrate (GPP, FPP, or GGPP; Cayman Chemical) in 750 µL Buffer D (50 mM HEPES, pH 7.2, 7.5 mM MgCl₂, and 5% (v/v) glycerol), with 500 µL hexane overlay. Reactions were carried out for 16 hours at 30°C, vortexed to extract products, and centrifuged to re-separate the aqueous and organic layers. The organic layer was directly removed for GC-MS analysis.

Transient expression in N. benthamiana

Transient expression assays in *N. benthamiana* were carried out as described earlier³. *N. benthamiana* plants were grown for 5 weeks in a controlled growth room under 16 h light (24°C) and 8 h dark (17°C) cycle before infiltration. Constructs of candidates in pEAQ and others used for co-expression were separately transformed into *Agrobacterium tumefaciens* strain LBA4404. Cultures were grown overnight at 30°C in 10 mL LB plus 50 µg/mL kanamycin and 50 µg/mL rifampicin, collected by centrifugation, and washed with 10 mL water twice. Cells were resuspended and diluted to an OD₆₀₀ of 1.0 in water plus 200 µM acetosyringone and incubated at 30°C for 2-3 hours. Separate cultures were mixed in a 1:1 ratio for each combination of enzyme tested (e.g. for leubethanol production, equal volumes of cultures were mixed harboring *CfDXS*, *LfCPT1*, *LfTPS1*, and *CYP71D616*). Mixed cultures were infiltrated with a syringe into the abaxial side of *N. benthamiana* leaves, and plants were returned to the controlled growth room for 5 days. Approximately 200 mg fresh weight from infiltrated leaves was extracted with 1 mL hexane overnight at 18°C, plant material was collected by centrifugation, and the organic phase was removed for GC-MS analysis.

E. coli in vivo assays

For *in vivo E. coli* assays, an engineered *E. coli* system⁴³ was used. *LfTPS1*(Δ23) and *LfTPS2*(Δ23) were co-transformed with pIRS and pNN and grown overnight at 37°C in 5 mL LB plus 25 µg/mL kanamycin, 17 µg/mL chloramphenicol, and 25 µg/mL streptomycin. A culture of 10 mL TB including the same antibiotics (same concentrations) was inoculated with 100 µL of the overnight culture and grown to an OD₆₀₀ of 0.6 at 37°C. The incubation temperature was lowered to 16°C for 1 hour, expression was induced with 0.5 mM IPTG, and cultures were supplemented

with 1 mM MgCl₂ and 40 mM pyruvate. Cultures were incubated at 16°C for an additional 60 hours before extraction with an equal volume of hexane and 2% (v/v) EtOH. The organic phase was separated by centrifugation and analyzed by GC-MS.

Dihydroserrulatene production scale-up and NMR

To generate enough of the major *Lf*TPS1 product (dihydroserrulatene) for NMR analysis, production in the *E. coli* system was carried out as detailed above, scaled up to 1 L. Following extraction, the organic layer was separated by centrifugation, concentrated under N₂ gas, and analyzed by GC-MS to confirm the presence of the *Lf*TPS1 product. This product was purified by silica gel flash column chromatography with a mobile phase of 10% ethyl acetate in hexane. NMR spectra were measured on an Agilent DirectDrive2 500 MHz spectrometer using CDCl₃ as the solvent. CDCl₃ peaks were referenced to 7.26 and 77.00 ppm for ¹H and ¹³C spectra, respectively.

GC-MS

All GC-MS analyses were performed on an Agilent 7890A GC with an Agilent VF-5ms column (30 m x 250 μm x 0.25 μm, with 10m EZ-Guard) and an Agilent 5975C detector. The inlet was set to 250°C splitless injection of 1 μL, He carrier gas (1 ml/min), and the detector was activated following a 3 min solvent delay. All assays and tissue analysis, with the exception of *in vitro* assays against GPP, used the following method: temperature ramp start 40°C, hold 1 min, 40°C/min to 200°C, hold 4.5 min, 20°C/min to 240°C, 10°C/min to 280°C; 40°C/min to 320°C; hold 5 min (3 min hold for *in vitro* assays). For *in vitro* assays against GPP, the following method was used: temperature ramp start 40°C; 10°C/min to 180°C; 40°C/min to 320°C; hold 3 min.

Homology Modeling

Homology models for LfCPT1 (Figure 2.S10) were generated using I-TASSER (v. 5.1⁶⁷) with either *Solanum habrochaites* (Z-Z)-FPPS (PDB ID: 5HXN⁶⁸) or LiLPPS (PDB ID: 5HC6⁶⁸) as the template structure. Figures were generated in PyMOL (v2.3).

Data Availability

RNA-seq data for *L. frutescens* has been submitted to the NCBI Sequence Read Archive (SRA) under the accession numbers SRX8371655 (root) and SRX8371656 (flower). GenBank accession numbers for nucleotide sequences of all enzymes tested in this study are as follows: LfTPS1: MT136608; LfTPS2: MT136609; LfCPT1: MT136610; LfCPT2: MT136611; LfCPT3: MT136612; CYP706G22: MT136613; CYP76A112: MT136614; CYP736A294: MT136615; CYP736A295: MT136616; CYP71D615: MT136617; CYP71D616: MT136618 EsTPS1: MT136619. Additional *L. frutescens* class I TPS candidates which were cloned but not characterized: LfTPS3: MT521506; LfTPS5: MT521507; LfTPS6: MT521505; LfTPS7: MT521508; LfTPS8a: MT521515; LfTPS8b: MT521516; LfTPS9: MT521509; LfTPS10: MT521511; LfTPS11: MT521510; LfTPS12a: MT521512; LfTPS12b: MT521513; LfTPS13: MT521514.

Acknowledgments

We are grateful for assistance from facilities at Michigan State University including the Institute for Cyber-Enabled Research, the Mass Spectrometry and Metabolomics Core, and the Max T. Rogers NMR Facility. We thank Thorben Höltkemeier, Matt Hofmeister, and Lars Bostelmann-Arp for excellent technical assistance, and Daniel Holmes for assistance with NMR. This work

was supported by the Michigan State University Strategic Partnership Grant program (“Evolutionary-Driven Genome Mining of Plant Biosynthetic Pathways” and “Plant-Inspired Chemical Diversity”). B.H. gratefully acknowledges the U.S. Department of Energy Great Lakes Bioenergy Research Center Cooperative Agreement DE-FC02-07ER64494 and DE-SC0018409, startup funding from the Department of Biochemistry and Molecular Biology, and support from AgBioResearch (MICL02454). G. M. is supported by a fellowship from Michigan State University under the Training Program in Plant Biotechnology for Health and Sustainability (T32-GM110523), and E. L. is supported by the NSF Graduate Research Fellowship Program (DGE-1848739). B.H. is in part supported by the National Science Foundation under Grant Number 1737898. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

APPENDIX

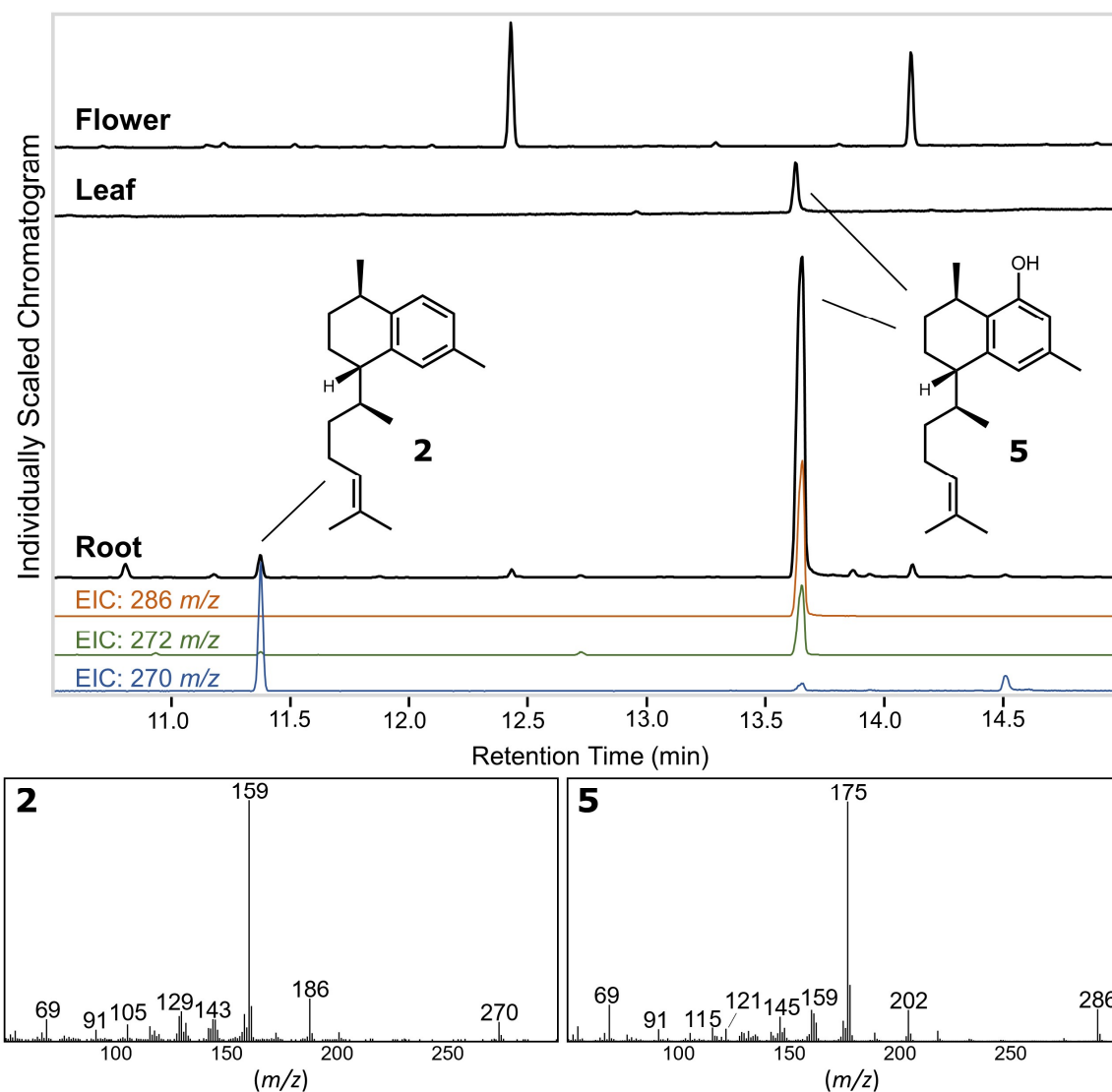


Figure 2.S1: GC-MS analysis of MTBE extracts of *L. frutescens* flower, leaf, and root tissue. Leubethanol (5) is present in both root and leaf tissue. Extracted and total ion chromatograms for root extract show accumulation of serrulatene (2) but no detectable quantities of dihydroserrulatene. Mass spectra for serrulatene (2) and leubethanol (5) (both from root sample) are shown.

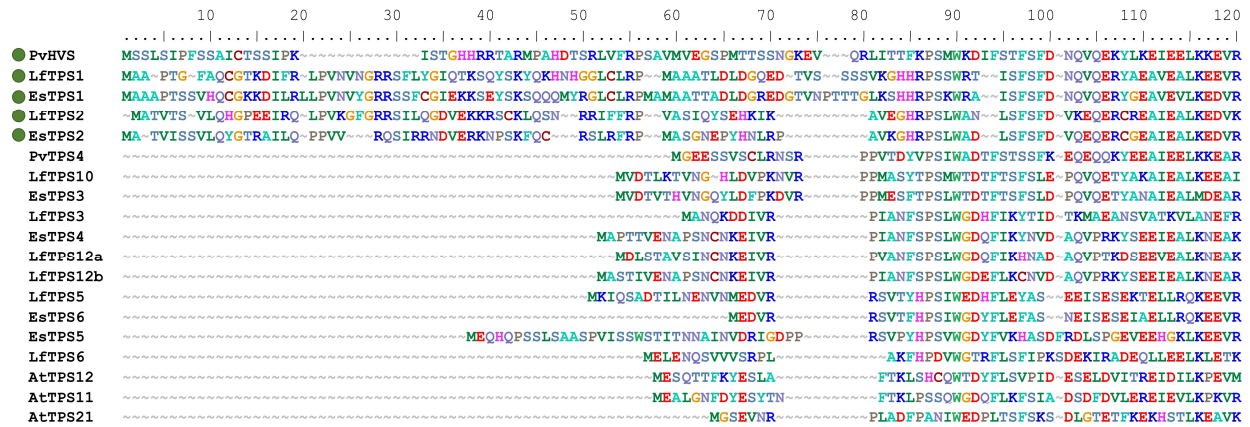


Figure 2.S2: First 120 positions of a sequence alignment of each reference and candidate TPS-a from Figure 2.2 in the main text. Green dots indicate predicted transit peptides based on presence of an N-terminal extension. PvHVS and PvTPS4 are known to localize to the plastid and cytosol, respectively⁷.

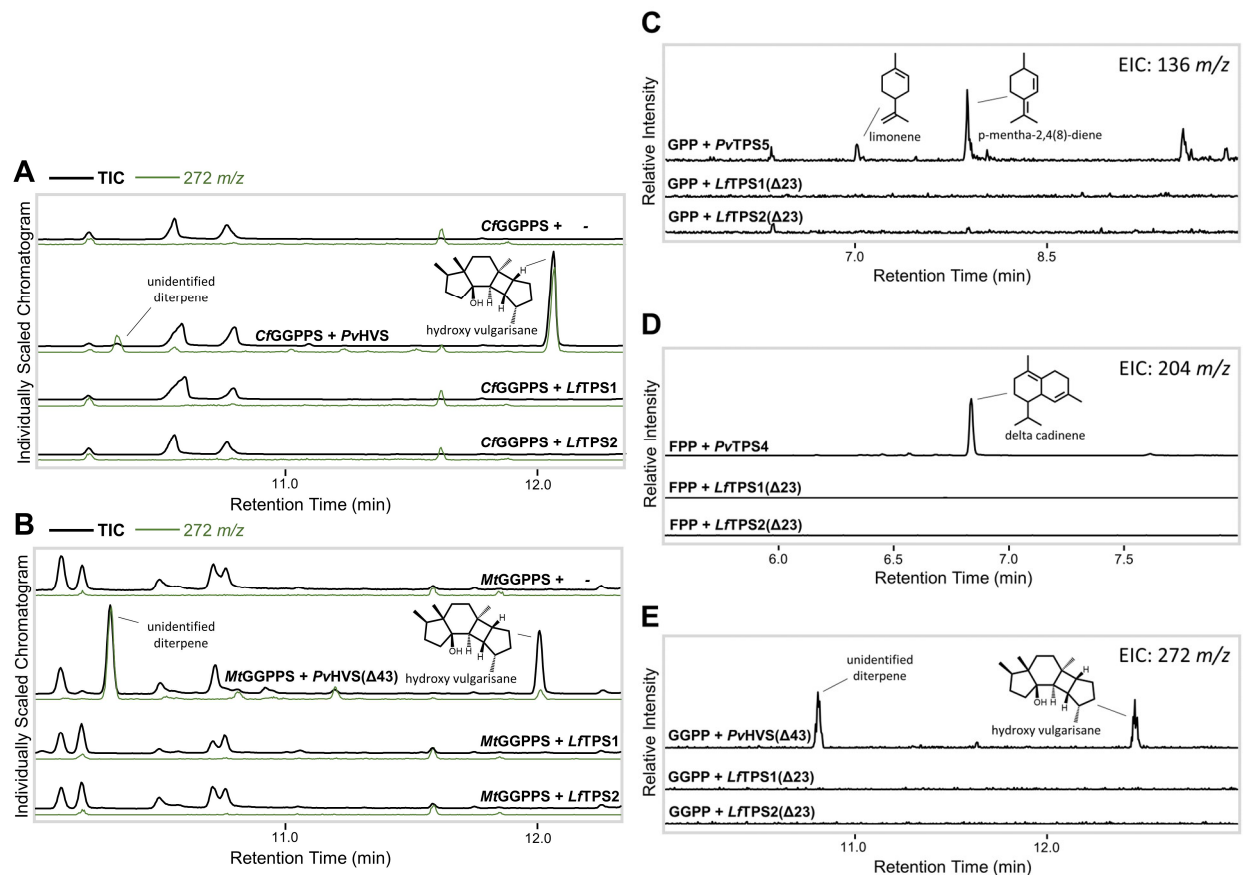


Figure 2.S3: CG-MS chromatograms of initial screening of LfTPS1 and LfTPS2. (A) Transient expression in *Nicotiana benthamiana* with plastidial precursors co-expressed. Each assay has CfDXS co-expressed in addition to those listed. (B) Transient expression in *Nicotiana benthamiana* with cytosolic precursors co-expressed. Each assay has E/HMGR co-expressed in addition to those listed. (C-E) Activity of purified LfTPS1 and LfTPS2 *in vitro* with (C) GPP, (D) FPP, and (E) GGPP. Putative transit peptides for both enzymes were removed (LfTPS1(Δ23) and LfTPS2(Δ23)). PvTPS5, PvTPS4, and PvHVS are TPSs in the same subfamily (TPS-a) from *Prunella vulgaris* which serve as positive controls for monoterpene, sesquiterpene, and diterpene synthesis, respectively, and products are shown as assigned in Johnson *et al.*⁷. PvHVS is natively targeted to the plastid, and PvHVS(Δ43) had its transit peptide removed for cytosolic expression and *in vitro* assays.

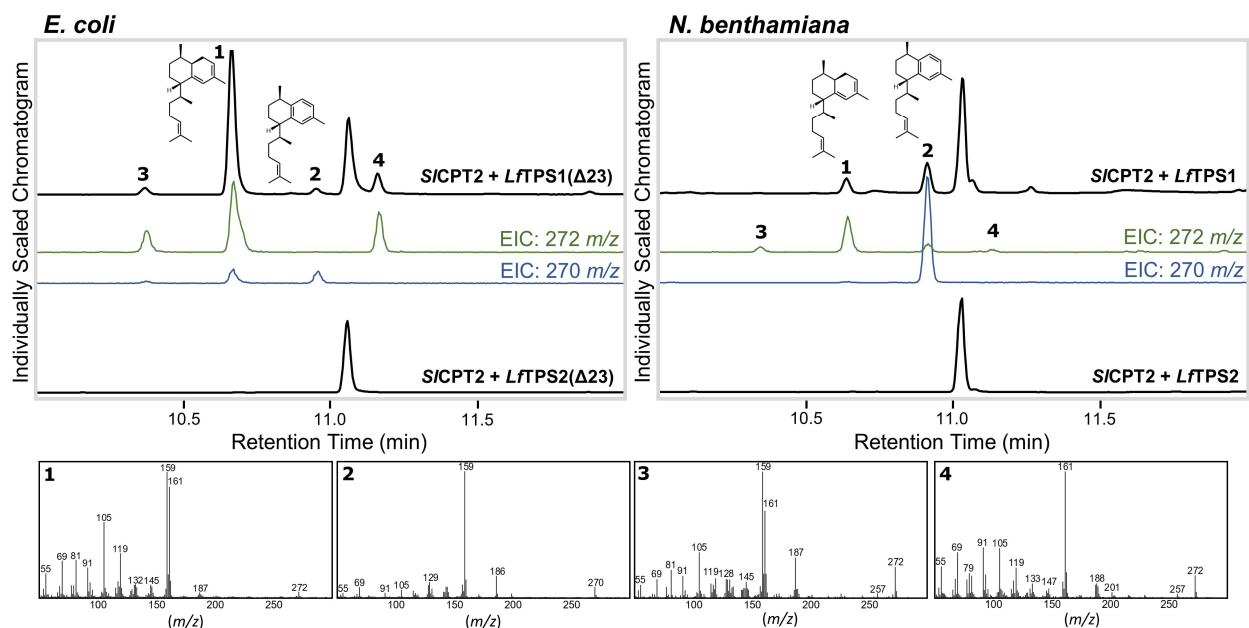
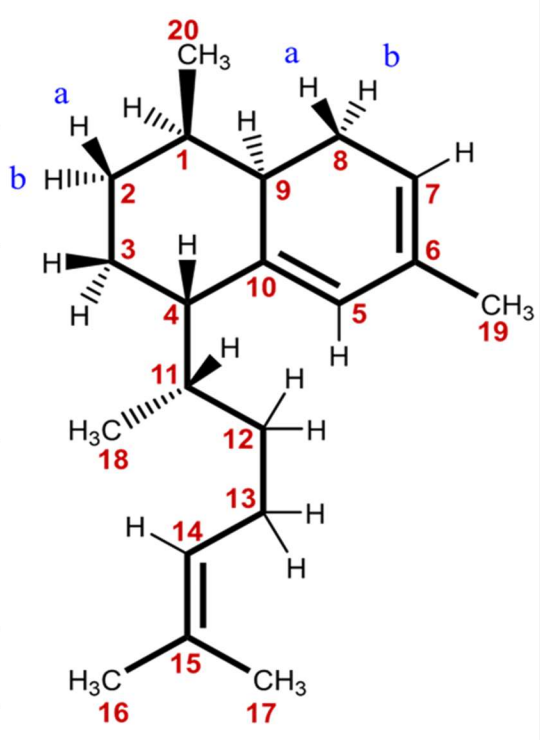


Figure 2.S4: GC-MS chromatograms for initial screening of LfTPS1 and LfTPS2 activity against NNPP in independent systems. Mass spectra for products 1-4 are given below. Both *N. benthamiana* assays have CfDXS co-expressed. Result of LfTPS1 conversion in *N. benthamiana* was replicated and is shown in Figure 2.3 in the main text.

Table 2.S1: ^{13}C and ^1H chemical shifts for NMR spectra of dihydroserrulatene. Structure and numbering given on the right. CDCl_3 was used as the solvent, and CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively.

^{13}C NMR	δ	^1H NMR	δ
C1	39.13	m, 2H	1.68
C2	29.45	m, 1H(a) m, 1H(b)	1.25 1.80
C3	25.52	m, 2H	1.73
C4	50.27	m, 1H	1.78
C5	122.43	d, $J = 1.5$ Hz, 1H	5.50
C6	129.18	-	-
C7	116.75	m, 1H	5.12
C8	28.24	m, 1H(a) m, 1H(b)	1.99 2.60
C9	35.00	m, 1H	2.58
C10	141.83	-	-
C11	30.80	m, 1H	1.80
C12	34.82	m, 1H m, 1H	1.04 1.58
C13	25.44	m, 1H m, 1H	1.93 2.08
C14	125.03	m, 1H	5.14
C15	131.22	-	-
C16	25.75	s, 3H	1.71
C17	17.70	s, 3H	1.64
C18	18.13	d, $J = 6.2$ Hz, 3H	0.84
C19	21.70	m, 3H	1.65
C20	12.63	d, $J = 7.0$ Hz, 3H	0.97



dihydroserrulatene

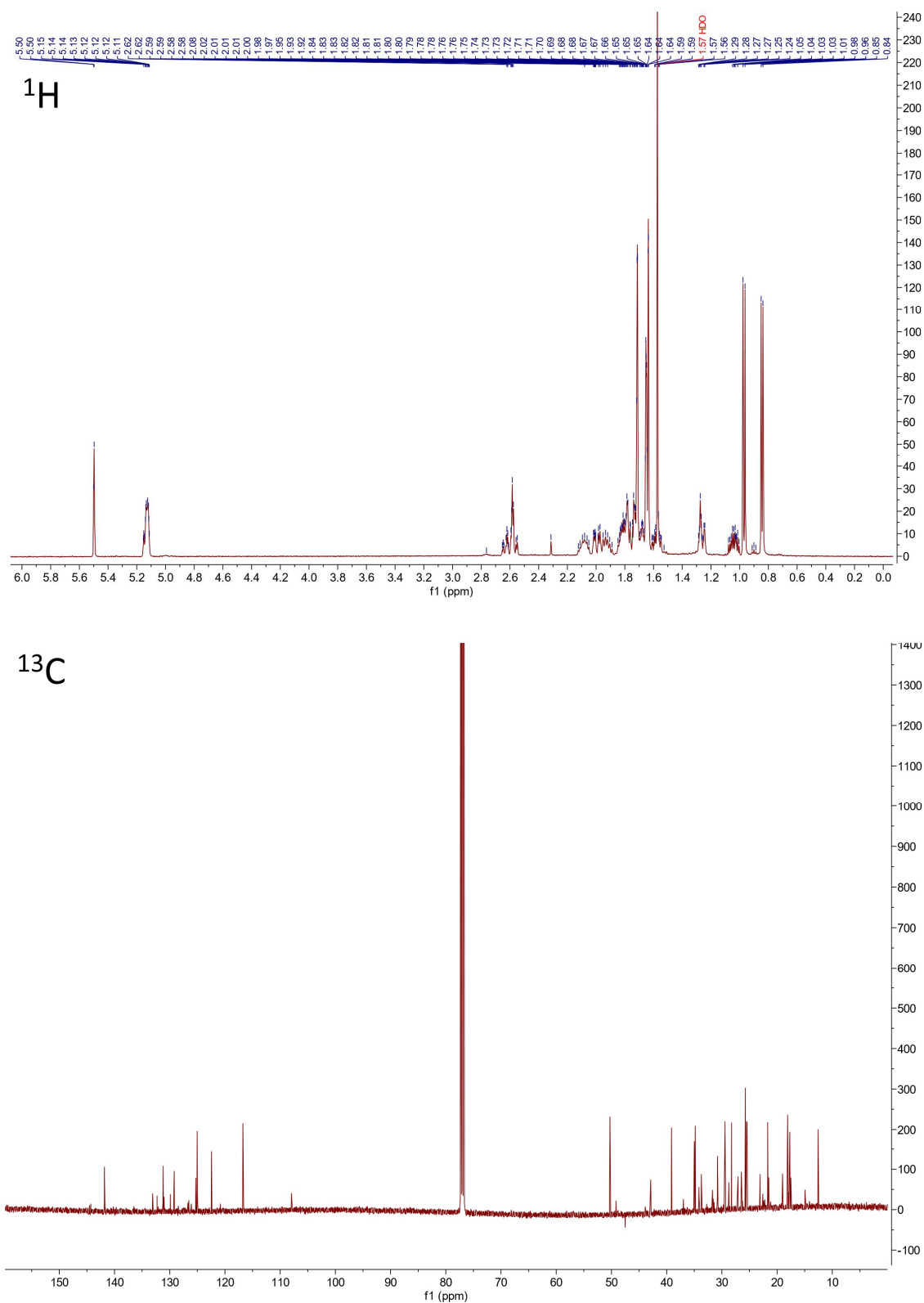
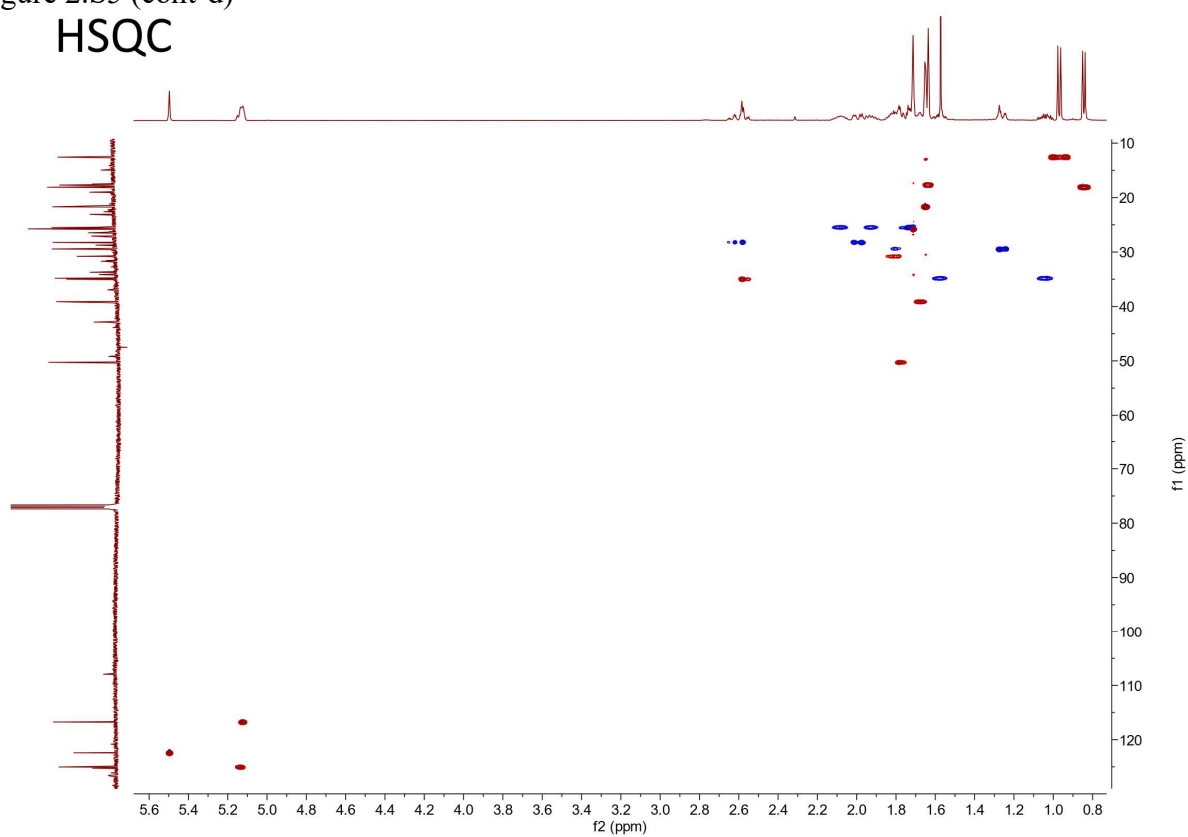


Figure 2.S5: ^1H , ^{13}C , HSQC, H2BC, HMBC, COSY, and NOESY NMR spectra for dihydroserrulatene. Relative stereochemistry matching serrulatane diterpenoids (including leubethanol), and stereochemistry at carbon 9, is supported by NOESY correlations.

Figure 2.S5 (cont'd)

HSQC



H2BC

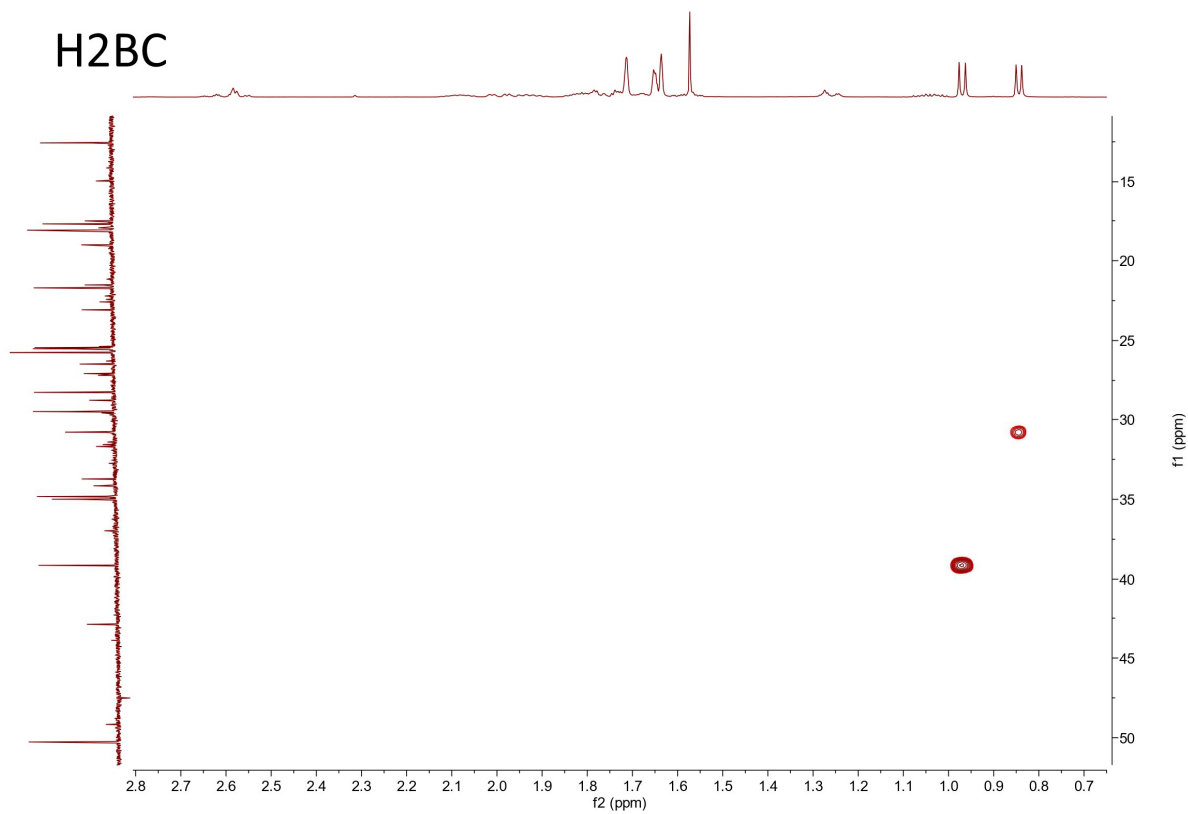
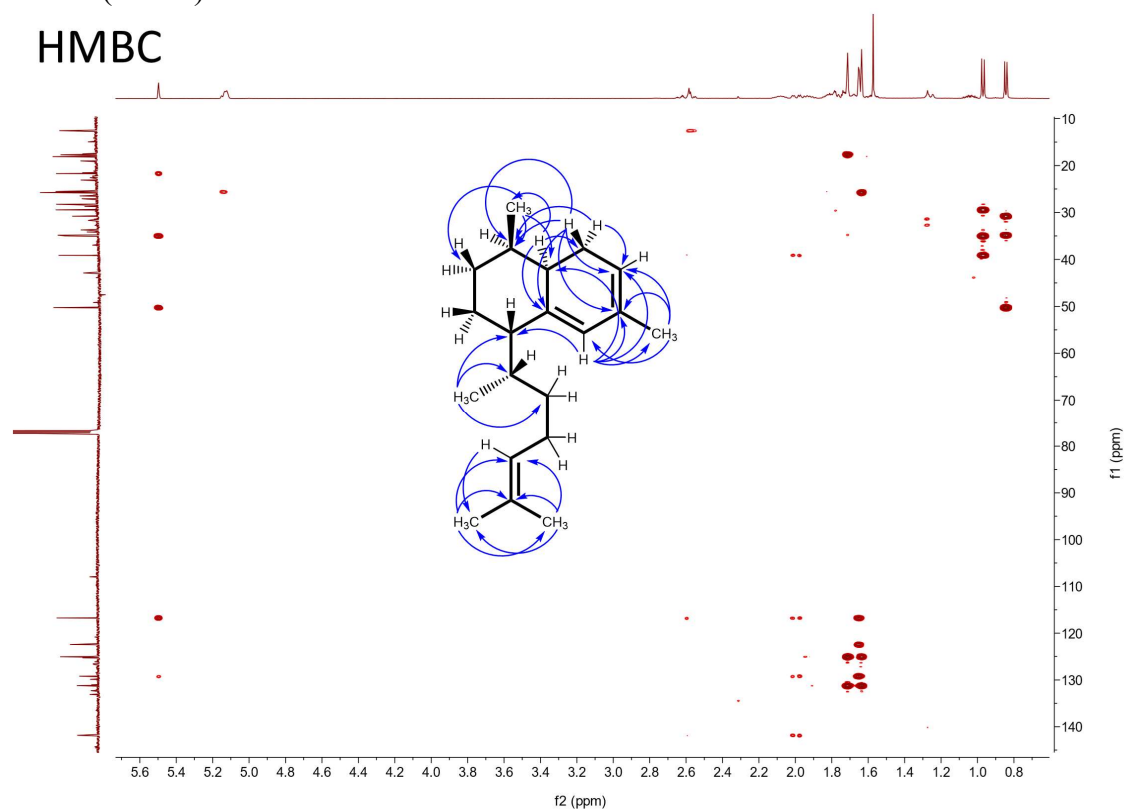


Figure 2.S5 (cont'd)

HMBC



NOESY

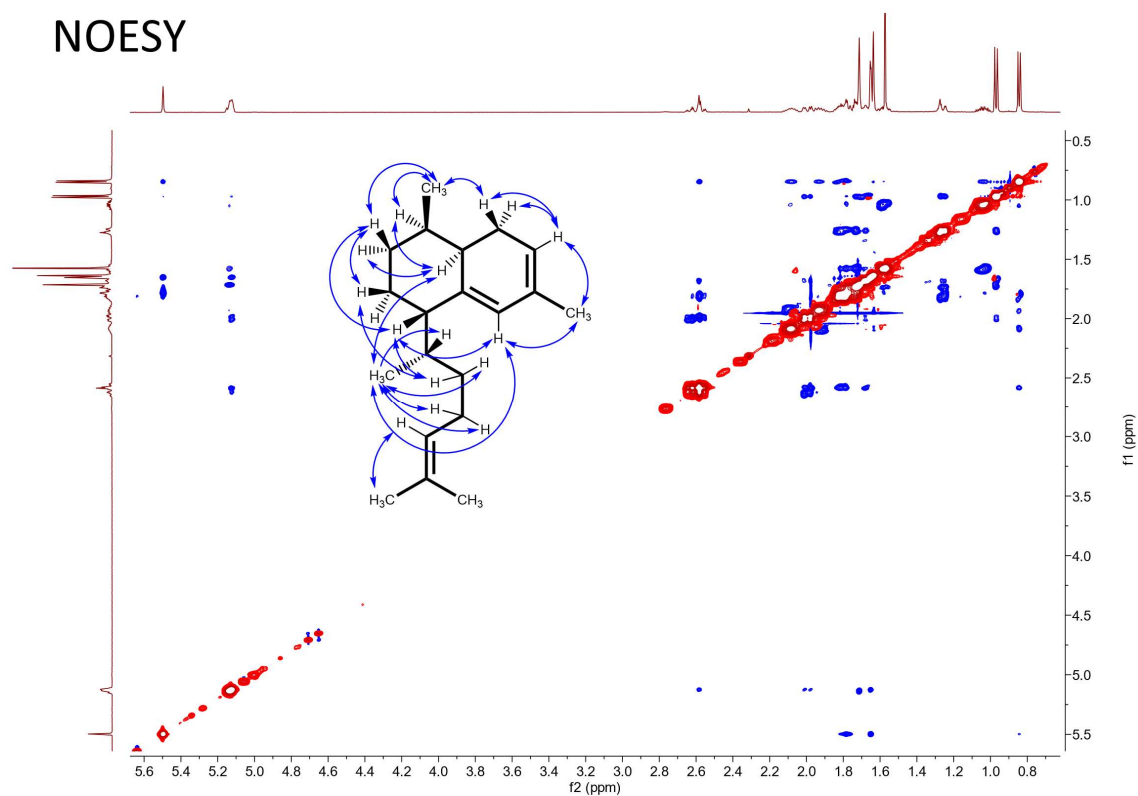
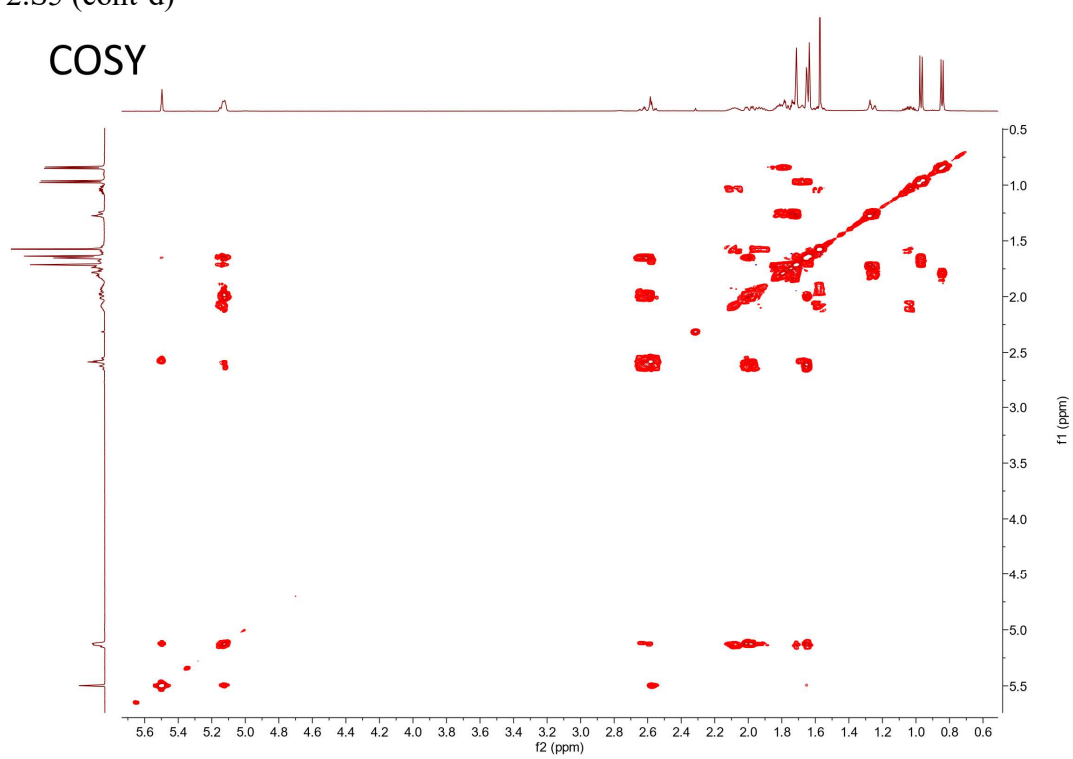
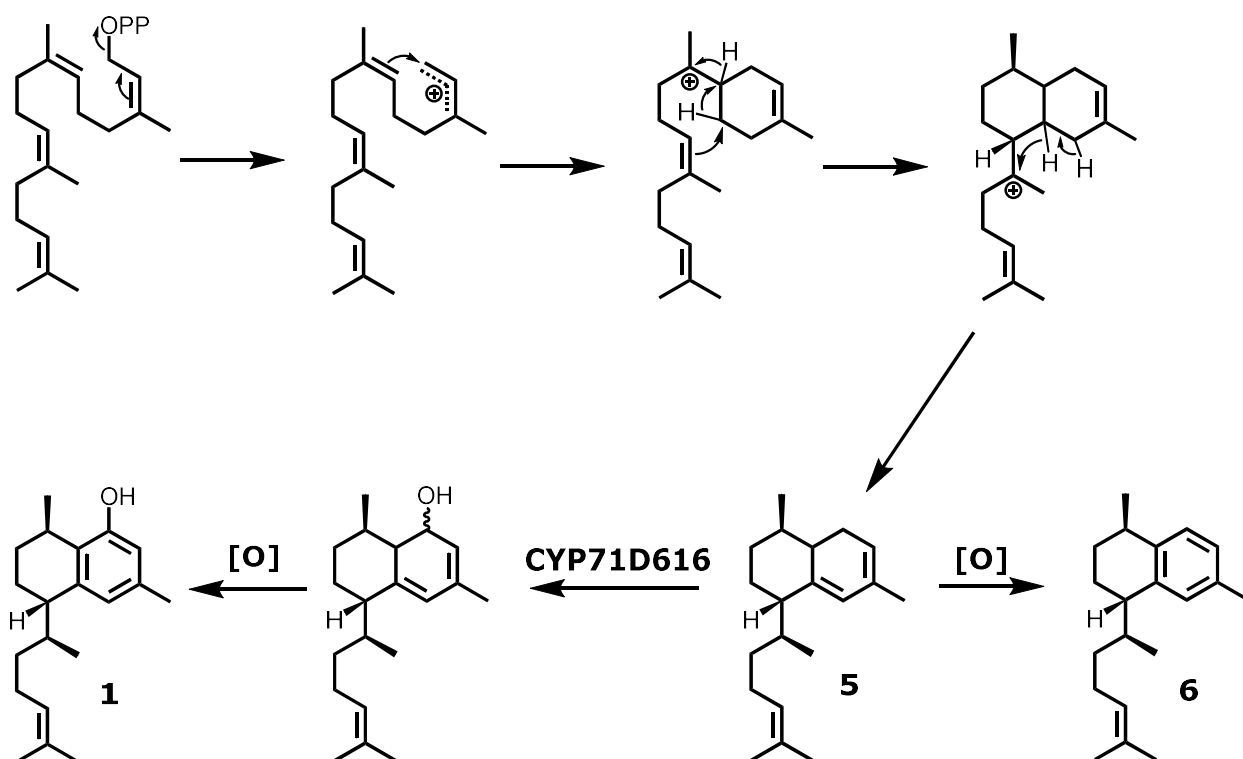


Figure 2.S5 (cont'd)





Scheme 2.S1: *Proposed mechanism for LfTPS1 conversion of NNPP to dihydroserulatene.* Subsequent aromatization to serulatene or conversion to leubethanol follows. [O] indicates spontaneous oxidation to an aromatic product. Shown in brackets is a putative intermediate between dihydroserulatene and leubethanol, not isolated in this study.

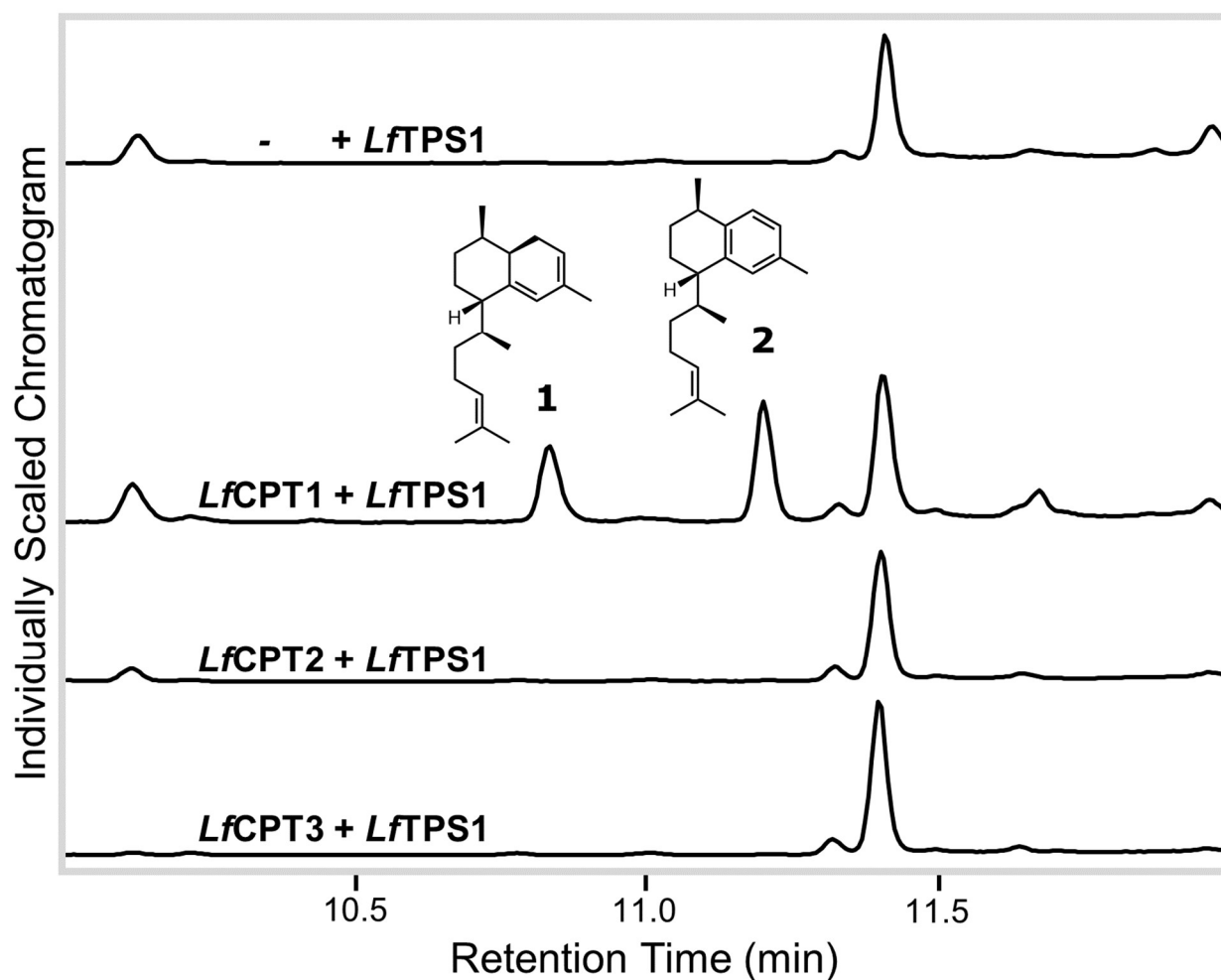


Figure 2.S6: GC-MS chromatograms for initial screening of LfCPT1-3 co-expressed with LfTPS1 in *N. benthamiana*. This initial screen used production of dihydroseerrulatene (1) and serrulatene (2) in coupled assays with LfTPS1 as a proxy for showing NNPP production of the *cis*-PT candidates. Direct characterization LfCPT1 showing NNPP production, and replication of this coupled assay, is shown in Figure 2.4 in the main text. Each assay has CfDXS co-expressed in addition to those listed.

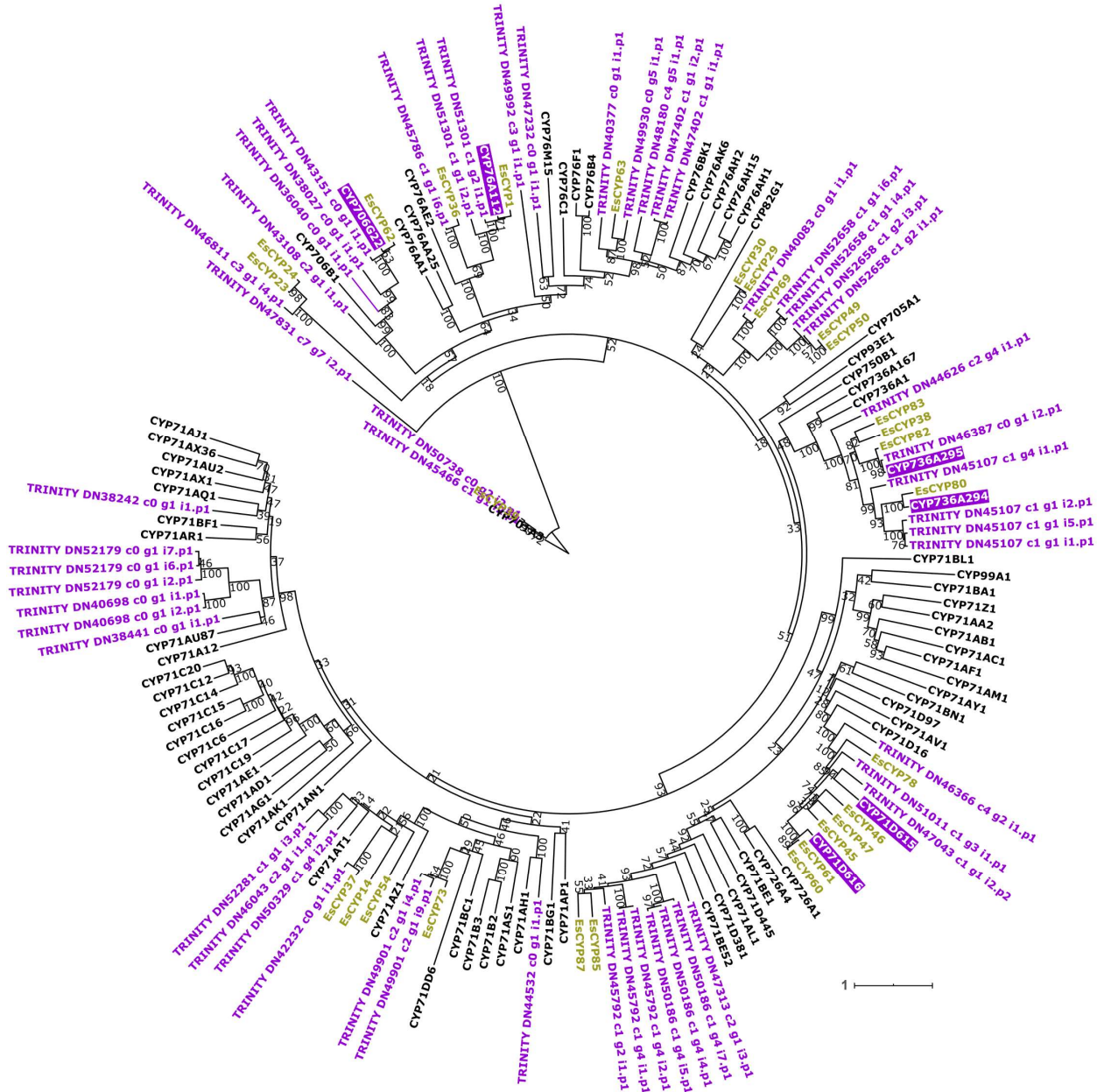


Figure 2.S7: Maximum likelihood phylogenetic tree of candidate cytochrome P450s. *L. frutescens* candidates are in purple, *E. serrulata* candidates are in yellow, and reference P450s are in black. Only candidates and references from the CYP71 clan are included. *L. frutescens* candidates are left as Trinity assembly codes, except for the six which were cloned and tested in this study (highlighted). Note that *E. serrulata* candidates are numbered 1-87, and the number has no correlation to CYP family. Scale bar represents substitutions per site, and branch numbers represent percent support from 1,000 bootstrap replicates. CYP701A3 from *Arabidopsis*, involved in gibberellic acid central metabolism, is used as an outgroup.

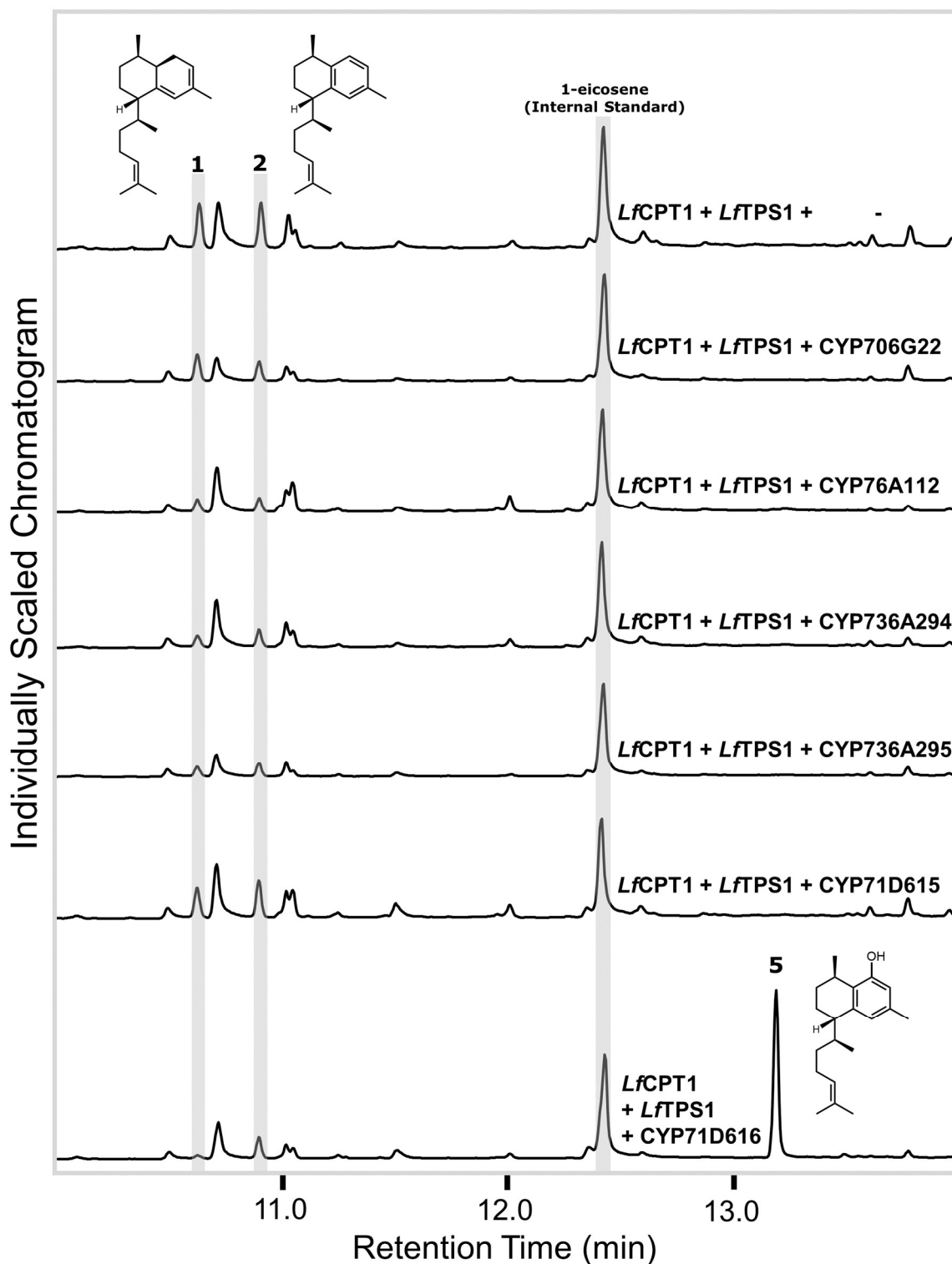


Figure 2.S8: GC-MS chromatograms for initial screening of *L. frutescens* P450 candidates co-expressed with *LfCPT1* and *LfTPS1* in *N. benthamiana*. Each assay has *CfDXS* co-expressed in addition to those listed. Result of CYP71D616 formation of leubethanol was replicated and is shown in Figure 2.5 in the main text.

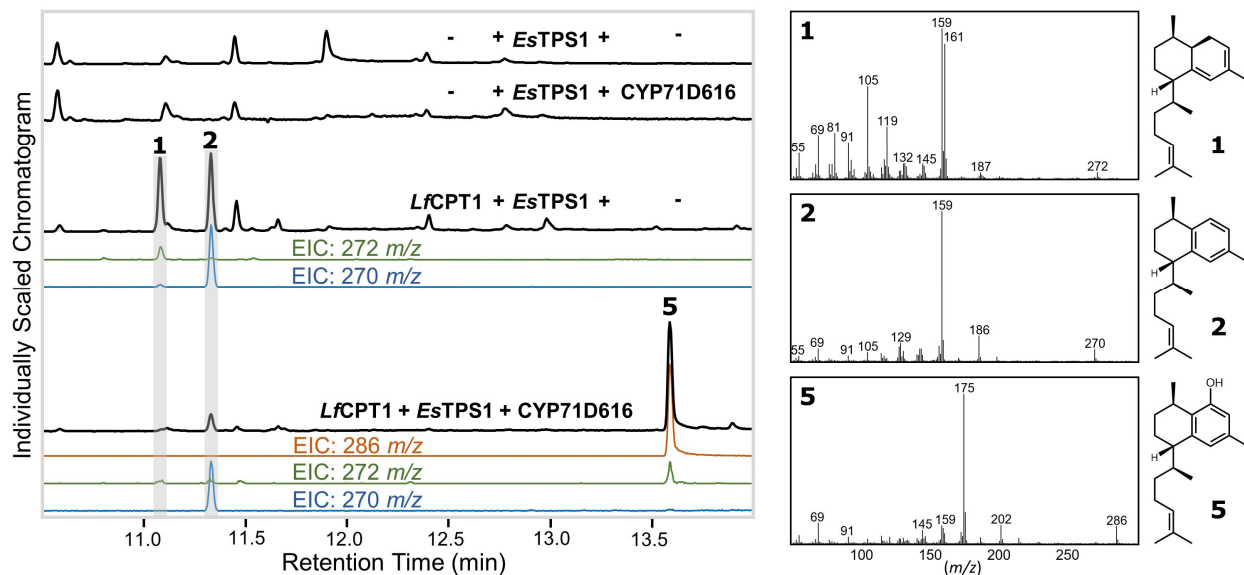


Figure 2.S9: GC-MS chromatograms of transient expression of the *Eremophila serrulata* orthologue to *LfTPS1* (*EsTPS1*). Screening was conducted in *N. benthamiana* with the other two enzymes from *L. frutescens* in the pathway for leubethanol. Each assay has *CfDXS* co-expressed in addition to those listed. Replacing *LfTPS1* with its orthologue *EsTPS1* yields the same results in each combination.

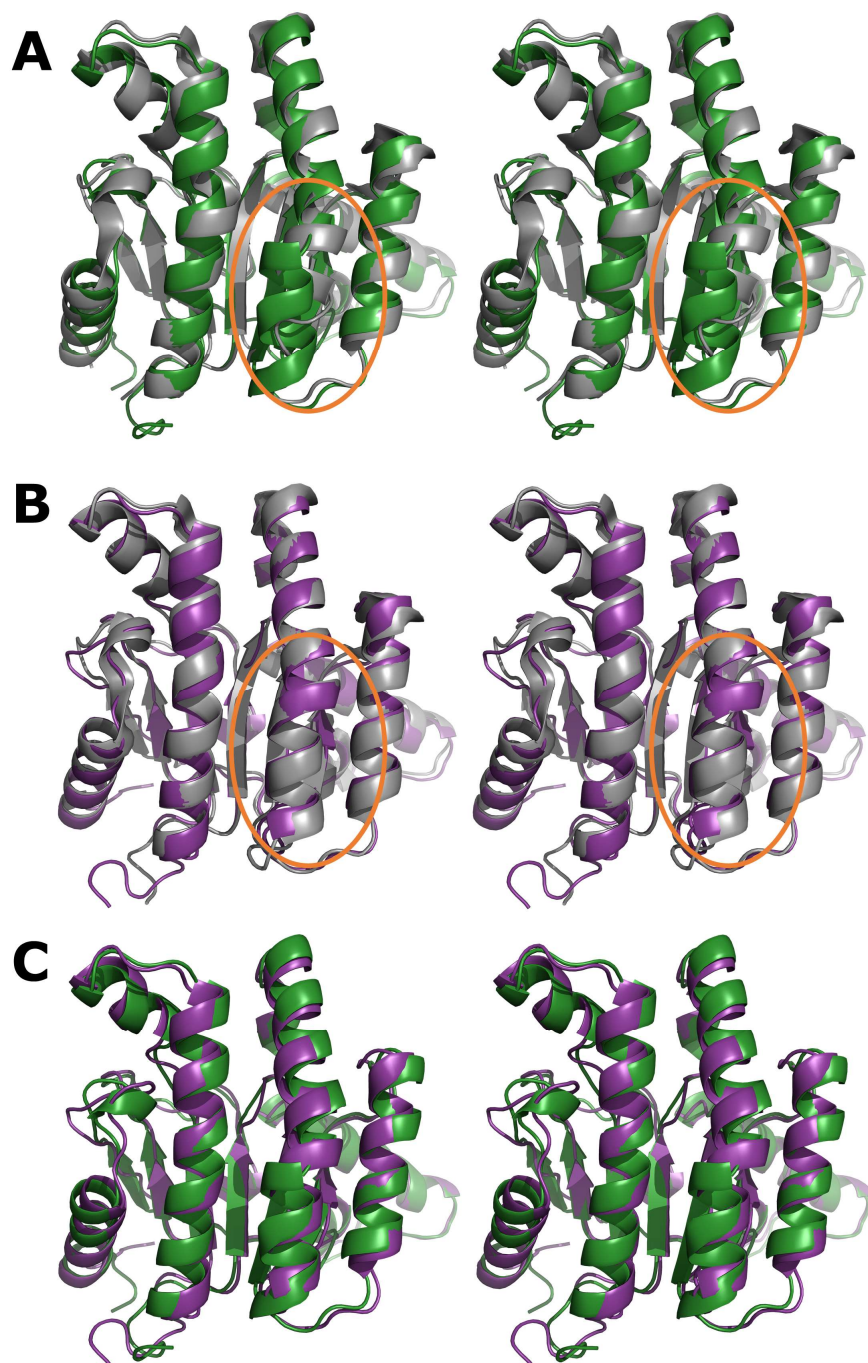


Figure 2.S10: Stereo-view of two different homology models of LfCPT1 aligned with model templates. (A) LfCPT1 model (green) aligned with its template *Solanum habrochaites* (Z-Z)-FPP synthase (closely related to *S/CPT2*; gray; PDB ID: 5HXN⁶⁸). (B) LfCPT1 model (purple) aligned with its template *Lavandula x intermedia* lavandulyl diphosphate synthase (gray; PDB ID: 5HC6⁶⁹). (C) LfCPT1 modeled off of 5HXN (green) and 5HC6 (purple) aligned to each other. Circled in orange is the shortened alpha helix present in each *Solanum* short-chain *cis*-PTs previously suggested to play a role in formation of shorter chain lengths, which is not present in either LfCPT1 or the unusual head-to-middle prenyl-transferase LiLPPS from the same *cis*-PT family.

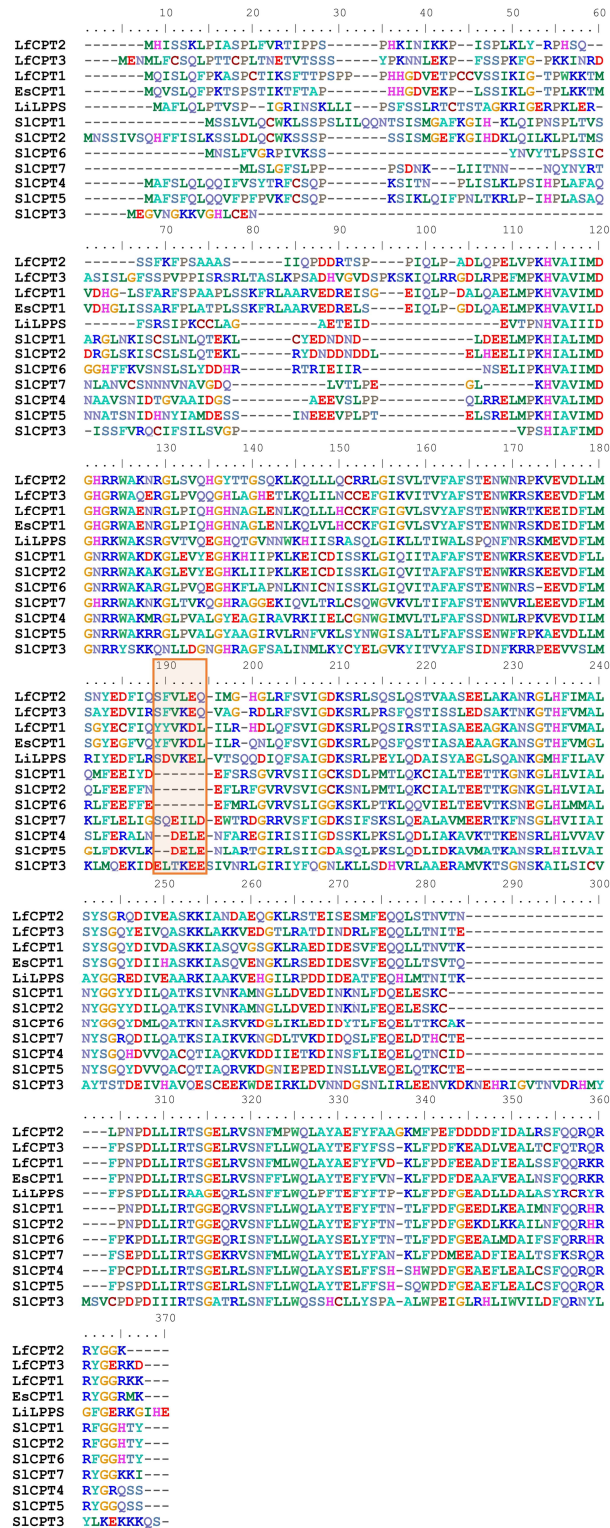


Figure 2.S11: Sequence alignment of reference cis-PTs from *S. lycopersicum*, *LiLPPS*, each *L. frutescens* cis-PT tested in this study, and the *LfCPT1* orthologue from *E. serrulata* (*EsCPT1*). Highlighted in orange is the shortened alpha helix present in *S/CPT1* (NPP synthase), *S/CPT2* (NNPP synthase), and *S/CPT6* ((Z-Z)-FPP synthase) as described in Figure 2.S10 and in the main text.

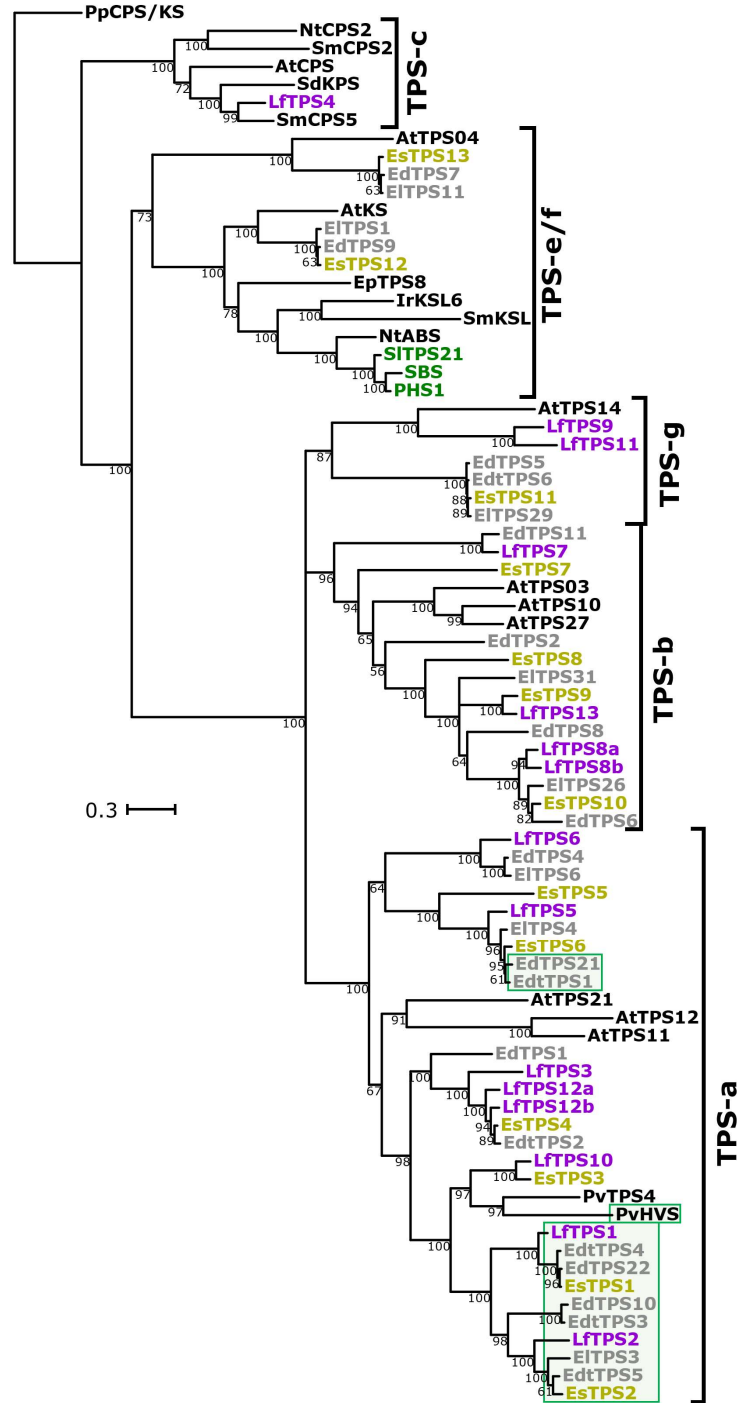


Figure 2.S12: Maximum likelihood phylogenetic tree of TPSs shown in Figure 2.2 in the main text, with recently identified TPSs⁴⁵ added from three other *Eremophila* species. Recently identified sequences are in gray, and also included are three enzymes from *Solanum* which exclusively convert *cis*-prenyl diphosphate precursors (green). Putative transit peptides are denoted within the TPS-a subfamily by green boxes. Branch numbers represent support from 1,000 bootstrap replicates, and branches with less than 50% support have been collapsed. Scale bar represents substitutions per site. Abbreviations for added sequences: *El*: *Eremophila lucida*; *Ed*: *Eremophila drummondii*; *Edt*: *Eremophila denticulata*; *Sl*: *Solanum lycopersicum*.

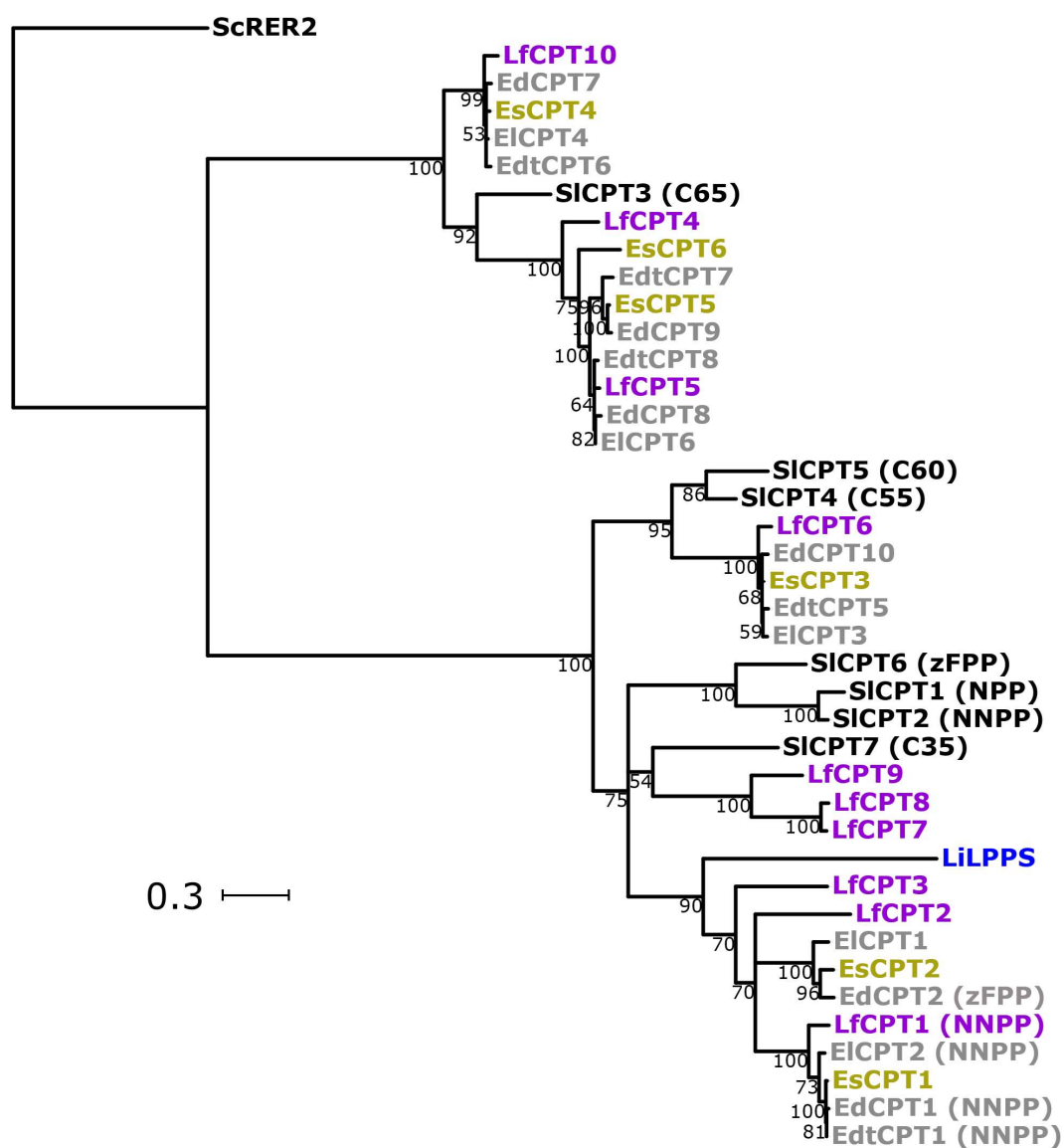


Figure 2.S13: Maximum likelihood phylogenetic tree of cis-PTs shown in Figure 2.4 in the main text, with recently identified cis-PTs⁴⁵ added from three other *Eremophila* species. Recently identified sequences are in gray. Shown in parenthesis are major products of each enzyme (if known). Branch numbers represent support from 1,000 bootstrap replicates, and branches with less than 50% support have been collapsed. Scale bar represents substitutions per site.

REFERENCES

REFERENCES

- (1) Degenhardt, J.; Köllner, T. G.; Gershenzon, J. Monoterpene and Sesquiterpene Synthases and the Origin of Terpene Skeletal Diversity in Plants. *Phytochemistry* **2009**, *70* (15–16), 1621–1637. <https://doi.org/10.1016/j.phytochem.2009.07.030>.
- (2) Durairaj, J.; Di Girolamo, A.; Bouwmeester, H. J.; de Ridder, D.; Beekwilder, J.; van Dijk, A. DJ. An Analysis of Characterized Plant Sesquiterpene Synthases. *Phytochemistry* **2019**, *158*, 157–165. <https://doi.org/10.1016/j.phytochem.2018.10.020>.
- (3) Johnson, S. R.; Bhat, W. W.; Bibik, J.; Turmo, A.; Hamberger, B.; Hamberger, B. A Database-Driven Approach Identifies Additional Diterpene Synthase Activities in the Mint Family (Lamiaceae). *J Biol Chem* **2019**, *294* (4), 1349–1362. <https://doi.org/10.1074/jbc.RA118.006025>.
- (4) Zeng, T.; Liu, Z.; Liu, H.; He, W.; Tang, X.; Xie, L.; Wu, R. Exploring Chemical and Biological Space of Terpenoids. *J. Chem. Inf. Model.* **2019**, *59* (9), 3667–3678. <https://doi.org/10.1021/acs.jcim.9b00443>.
- (5) Luo, D.; Callari, R.; Hamberger, B.; Wubshet, S. G.; Nielsen, M. T.; Andersen-Ranberg, J.; Hallström, B. M.; Cozzi, F.; Heider, H.; Lindberg Møller, B.; Staerk, D.; Hamberger, B. Oxidation and Cyclization of Casbene in the Biosynthesis of Euphorbia Factors from Mature Seeds of Euphorbia Lathyris L. *Proceedings of the National Academy of Sciences* **2016**, *113* (34), E5082–E5089. <https://doi.org/10.1073/pnas.1607504113>.
- (6) Pateraki, I.; Andersen-Ranberg, J.; Jensen, N. B.; Wubshet, S. G.; Heskes, A. M.; Forman, V.; Hallström, B.; Hamberger, B.; Motawia, M. S.; Olsen, C. E.; Staerk, D.; Hansen, J.; Møller, B. L.; Hamberger, B. Total Biosynthesis of the Cyclic AMP Booster Forskolin from Coleus Forskohlii. *eLife* **2017**, *6*, e23001. <https://doi.org/10.7554/eLife.23001>.
- (7) Johnson, S. R.; Bhat, W. W.; Sadre, R.; Miller, G. P.; Garcia, A. S.; Hamberger, B. Promiscuous Terpene Synthases from Prunella Vulgaris Highlight the Importance of Substrate and Compartment Switching in Terpene Synthase Evolution. *New Phytologist* **2019**, *223* (1), 323–335. <https://doi.org/10.1111/nph.15778>.
- (8) Molina-Salinas, G. M.; Rivas-Galindo, V. M.; Said-Fernández, S.; Lankin, D. C.; Muñoz, M. A.; Joseph-Nathan, P.; Pauli, G. F.; Waksman, N. Stereochemical Analysis of Leubethanol, an Anti-TB-Active Serrulatane, from Leucophyllum Frutescens. *J. Nat. Prod.* **2011**, *74* (9), 1842–1850. <https://doi.org/10.1021/np2000667>.
- (9) Andrade Carvalho, A.; da Costa, P. M.; Da Silva Souza, L. G.; Lemos, T. L. G.; Alves, A. P. N. N.; Pessoa, C.; de Moraes, M. O. Inhibition of Metastatic Potential of B16-F10 Melanoma Cell Line in Vivo and in Vitro by Biflorin. *Life Sciences* **2013**, *93* (5), 201–207. <https://doi.org/10.1016/j.lfs.2013.05.018>.

- (10) Ndi, C. P.; Semple, S. J.; Griesser, H. J.; Pyke, S. M.; Barton, M. D. Antimicrobial Compounds from the Australian Desert Plant *Eremophila Neglecta*. *J. Nat. Prod.* **2007**, *70* (9), 1439–1443. <https://doi.org/10.1021/np070180r>.
- (11) Kumar, R.; Duffy, S.; Avery, V. M.; Carroll, A. R.; Davis, R. A. Microthecaline A, a Quinoline Serrulatane Alkaloid from the Roots of the Australian Desert Plant *Eremophila Microtheca*. *J. Nat. Prod.* **2018**, *81* (4), 1079–1083. <https://doi.org/10.1021/acs.jnatprod.7b00992>.
- (12) Ndi, C. P.; Semple, S. J.; Griesser, H. J.; Pyke, S. M.; Barton, M. D. Antimicrobial Compounds from *Eremophila Serrulata*. *Phytochemistry* **2007**, *68* (21), 2684–2690. <https://doi.org/10.1016/j.phytochem.2007.05.039>.
- (13) Anakok, O. f.; Ndi, C. p.; Barton, M. d.; Griesser, H. j.; Semple, S. j. Antibacterial Spectrum and Cytotoxic Activities of Serrulatane Compounds from the Australian Medicinal Plant *Eremophila Neglecta*. *Journal of Applied Microbiology* **2012**, *112* (1), 197–204. <https://doi.org/10.1111/j.1365-2672.2011.05174.x>.
- (14) Barnes, E. C.; Kavanagh, A. M.; Ramu, S.; Blaskovich, M. A.; Cooper, M. A.; Davis, R. A. Antibacterial Serrulatane Diterpenes from the Australian Native Plant *Eremophila Microtheca*. *Phytochemistry* **2013**, *93*, 162–169. <https://doi.org/10.1016/j.phytochem.2013.02.021>.
- (15) Mon, H. H.; Christo, S. N.; Ndi, C. P.; Jasieniak, M.; Rickard, H.; Hayball, J. D.; Griesser, H. J.; Semple, S. J. Serrulatane Diterpenoid from *Eremophila Neglecta* Exhibits Bacterial Biofilm Dispersion and Inhibits Release of Pro-Inflammatory Cytokines from Activated Macrophages. *J. Nat. Prod.* **2015**, *78* (12), 3031–3040. <https://doi.org/10.1021/acs.jnatprod.5b00833>.
- (16) Hossain, M. A.; Biva, I. J.; Kidd, S. E.; Whittle, J. D.; Griesser, H. J.; Coad, B. R. Antifungal Activity in Compounds from the Australian Desert Plant *Eremophila Alternifolia* with Potency Against *Cryptococcus* Spp. *Antibiotics* **2019**, *8* (2), 34. <https://doi.org/10.3390/antibiotics8020034>.
- (17) Aminimoghadamfarouj, N.; Nematollahi, A. Structure Elucidation and Botanical Characterization of Diterpenes from a Specific Type of Bee Glue. *Molecules* **2017**, *22* (7), 1185. <https://doi.org/10.3390/molecules22071185>.
- (18) Best, W. M.; Wege, D. Intramolecular Diels-Alder Additions of Benzyne to Furans. Exploratory Studies. *Aust. J. Chem.* **1986**, *39* (4), 635–645. <https://doi.org/10.1071/ch9860635>.
- (19) Lu, J. M. H.; Perkins, M. V.; Griesser, H. J. Total Synthesis and Structural Confirmation of the Antibacterial Diterpene Leubethanol. *Tetrahedron* **2013**, *69* (31), 6468–6473. <https://doi.org/10.1016/j.tet.2013.05.082>.
- (20) Yu, X.; Su, F.; Liu, C.; Yuan, H.; Zhao, S.; Zhou, Z.; Quan, T.; Luo, T. Enantioselective Total Syntheses of Various Amphilectane and Serrulatane Diterpenoids via Cope Rearrangements. *J. Am. Chem. Soc.* **2016**, *138* (19), 6261–6270. <https://doi.org/10.1021/jacs.6b02624>.

- (21) Kumar, R.; Duffy, S.; Avery, V. M.; Davis, R. A. Synthesis of Antimalarial Amide Analogues Based on the Plant Serrulatane Diterpenoid 3,7,8-Trihydroxyserrulat-14-En-19-Oic Acid. *Bioorganic & Medicinal Chemistry Letters* **2017**, *27* (17), 4091–4095. <https://doi.org/10.1016/j.bmcl.2017.07.039>.
- (22) Tenneti, S.; Biswas, S.; Cox, G. A.; Mans, D. J.; Lim, H. J.; RajanBabu, T. V. Broadly Applicable Stereoselective Syntheses of Serrulatane, Amphilectane Diterpenes, and Their Diastereoisomeric Congeners Using Asymmetric Hydrovinylation for Absolute Stereochemical Control. *J. Am. Chem. Soc.* **2018**, *140* (31), 9868–9881. <https://doi.org/10.1021/jacs.8b03549>.
- (23) Reddy Penjarla, T.; Kundarapu, M.; Mohd. Baquer, S.; Bhattacharya, A. Total Synthesis of the Plant Alkaloid Racemic Microthecaline A: First Example of a Natural Product Bearing a Tricyclic Quinoline-Serrulatane Scaffold. *RSC Advances* **2019**, *9* (40), 23289–23294. <https://doi.org/10.1039/C9RA04675E>.
- (24) Kracht, O. N.; Ammann, A.-C.; Stockmann, J.; Wibberg, D.; Kalinowski, J.; Piotrowski, M.; Kerr, R.; Brück, T.; Kourist, R. Transcriptome Profiling of the Australian Arid-Land Plant *Eremophila Serrulata* (A.DC.) Druce (Scrophulariaceae) for the Identification of Monoterpene Synthases. *Phytochemistry* **2017**, *136*, 15–22. <https://doi.org/10.1016/j.phytochem.2017.01.016>.
- (25) Hamberger, B.; Bak, S. Plant P450s as Versatile Drivers for Evolution of Species-Specific Chemical Diversity. *Philos Trans R Soc Lond B Biol Sci* **2013**, *368* (1612), 20120426. <https://doi.org/10.1098/rstb.2012.0426>.
- (26) Bathe, U.; Tissier, A. Cytochrome P450 Enzymes: A Driving Force of Plant Diterpene Diversity. *Phytochemistry* **2019**, *161*, 149–162. <https://doi.org/10.1016/j.phytochem.2018.12.003>.
- (27) Molina-Salinas, G. M.; Pérez-López, A.; Becerril-Montes, P.; Salazar-Aranda, R.; Said-Fernández, S.; Torres, N. W. de. Evaluation of the Flora of Northern Mexico for in Vitro Antimicrobial and Antituberculosis Activity. *Journal of Ethnopharmacology* **2007**, *109* (3), 435–441. <https://doi.org/10.1016/j.jep.2006.08.014>.
- (28) Ghisalberti, E. L. The Phytochemistry of the Myoporaceae. *Phytochemistry* **1993**, *35* (1), 7–33. [https://doi.org/10.1016/S0031-9422\(00\)90503-X](https://doi.org/10.1016/S0031-9422(00)90503-X).
- (29) Peters, R. J. Two Rings in Them All: The Labdane-Related Diterpenoids. *Nat. Prod. Rep.* **2010**, *27* (11), 1521–1530. <https://doi.org/10.1039/C0NP00019A>.
- (30) Mau, C. J.; West, C. A. Cloning of Casbene Synthase cDNA: Evidence for Conserved Structural Features among Terpenoid Cyclases in Plants. *Proc Natl Acad Sci U S A* **1994**, *91* (18), 8497–8501. <https://doi.org/10.1073/pnas.91.18.8497>.
- (31) Ennajdaoui, H.; Vachon, G.; Giacalone, C.; Besse, I.; Sallaud, C.; Herzog, M.; Tissier, A. Trichome Specific Expression of the Tobacco (*Nicotiana glauca*) Cembratrien-Ol Synthase Genes Is Controlled by Both Activating and Repressing Cis-Regions. *Plant Mol Biol* **2010**, *73* (6), 673–685. <https://doi.org/10.1007/s11103-010-9648-x>.

- (32) Kirby, J.; Nishimoto, M.; Park, J. G.; Withers, S. T.; Nowroozi, F.; Behrendt, D.; Rutledge, E. J. G.; Fortman, J. L.; Johnson, H. E.; Anderson, J. V.; Keasling, J. D. Cloning of Casbene and Neocembrene Synthases from Euphorbiaceae Plants and Expression in *Saccharomyces Cerevisiae*. *Phytochemistry* **2010**, *71* (13), 1466–1473. <https://doi.org/10.1016/j.phytochem.2010.06.001>.
- (33) Vaughan, M. M.; Wang, Q.; Webster, F. X.; Kiemle, D.; Hong, Y. J.; Tantillo, D. J.; Coates, R. M.; Wray, A. T.; Askew, W.; O'Donnell, C.; Tokuhisa, J. G.; Tholl, D. Formation of the Unusual Semivolatile Diterpene Rhizathalene by the Arabidopsis Class I Terpene Synthase TPS08 in the Root Stele Is Involved in Defense against Belowground Herbivory[W]. *Plant Cell* **2013**, *25* (3), 1108–1125. <https://doi.org/10.1105/tpc.112.100057>.
- (34) Vaughan, M. M.; Wang, Q.; Webster, F. X.; Kiemle, D.; Hong, Y. J.; Tantillo, D. J.; Coates, R. M.; Wray, A. T.; Askew, W.; O'Donnell, C.; Tokuhisa, J. G.; Tholl, D. Formation of the Unusual Semivolatile Diterpene Rhizathalene by the Arabidopsis Class I Terpene Synthase TPS08 in the Root Stele Is Involved in Defense against Belowground Herbivory. *The Plant Cell* **2013**, *25* (3), 1108–1125. <https://doi.org/10.1105/tpc.112.100057>.
- (35) Zerbe, P.; Hamberger, B.; Yuen, M. M. S.; Chiang, A.; Sandhu, H. K.; Madilao, L. L.; Nguyen, A.; Hamberger, B.; Bach, S. S.; Bohlmann, J. Gene Discovery of Modular Diterpene Metabolism in Nonmodel Systems. *Plant Physiology* **2013**, *162* (2), 1073–1091. <https://doi.org/10.1104/pp.113.218347>.
- (36) Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. The Family of Terpene Synthases in Plants: A Mid-Size Family of Genes for Specialized Metabolism That Is Highly Diversified throughout the Kingdom. *The Plant Journal* **2011**, *66* (1), 212–229. <https://doi.org/10.1111/j.1365-313X.2011.04520.x>.
- (37) Andersen-Ranberg, J.; Kongstad, K. T.; Nielsen, M. T.; Jensen, N. B.; Pateraki, I.; Bach, S. S.; Hamberger, B.; Zerbe, P.; Staerk, D.; Bohlmann, J.; Møller, B. L.; Hamberger, B. Expanding the Landscape of Diterpene Structural Diversity through Stereochemically Controlled Combinatorial Biosynthesis. *Angewandte Chemie International Edition* **2016**, *55* (6), 2142–2146. <https://doi.org/10.1002/anie.201510650>.
- (38) Akhtar, T. A.; Matsuba, Y.; Schauvinhold, I.; Yu, G.; Lees, H. A.; Klein, S. E.; Pichersky, E. The Tomato Cis–Prenyltransferase Gene Family. *The Plant Journal* **2013**, *73* (4), 640–652. <https://doi.org/10.1111/tpj.12063>.
- (39) Sallaud, C.; Rontein, D.; Onillon, S.; Jabès, F.; Duffé, P.; Giacalone, C.; Thoraval, S.; Escoffier, C.; Herbette, G.; Leonhardt, N.; Causse, M.; Tissier, A. A Novel Pathway for Sesquiterpene Biosynthesis from Z,Z-Farnesyl Pyrophosphate in the Wild Tomato *Solanum Habrochaites*. *The Plant Cell* **2009**, *21* (1), 301–317. <https://doi.org/10.1105/tpc.107.057885>.
- (40) Schilmiller, A. L.; Schauvinhold, I.; Larson, M.; Xu, R.; Charbonneau, A. L.; Schmidt, A.; Wilkerson, C.; Last, R. L.; Pichersky, E. Monoterpenes in the Glandular Trichomes of Tomato Are Synthesized from a Neryl Diphosphate Precursor Rather than Geranyl Diphosphate.

Proceedings of the National Academy of Sciences **2009**, *106* (26), 10865–10870. <https://doi.org/10.1073/pnas.0904113106>.

(41) Zi, J.; Matsuba, Y.; Hong, Y. J.; Jackson, A. J.; Tantillo, D. J.; Pichersky, E.; Peters, R. J. Biosynthesis of Lycosantalanol, a Cis-Prenyl Derived Diterpenoid. *J. Am. Chem. Soc.* **2014**, *136* (49), 16951–16953. <https://doi.org/10.1021/ja508477e>.

(42) Matsuba, Y.; Zi, J.; Jones, A. D.; Peters, R. J.; Pichersky, E. Biosynthesis of the Diterpenoid Lycosantalanol via Nerylneryl Diphosphate in *Solanum Lycopersicum*. *PLOS ONE* **2015**, *10* (3), e0119302. <https://doi.org/10.1371/journal.pone.0119302>.

(43) Cyr, A.; Wilderman, P. R.; Determan, M.; Peters, R. J. A Modular Approach for Facile Biosynthesis of Labdane-Related Diterpenes. *J Am Chem Soc* **2007**, *129* (21), 6684–6685. <https://doi.org/10.1021/ja071158n>.

(44) Zi, J.; J. Peters, R. Characterization of CYP76AH4 Clarifies Phenolic Diterpenoid Biosynthesis in the Lamiaceae. *Organic & Biomolecular Chemistry* **2013**, *11* (44), 7650–7652. <https://doi.org/10.1039/C3OB41885E>.

(45) Gericke, O.; Hansen, N. L.; Pedersen, G. B.; Kjaerulff, L.; Luo, D.; Staerk, D.; Møller, B. L.; Pateraki, I.; Heskes, A. M. Nerylneryl Diphosphate Is the Precursor of Serrulatane, Viscidane and Cembrane-Type Diterpenoids in *Eremophila* Species. *BMC Plant Biology* **2020**, *20* (1), 91. <https://doi.org/10.1186/s12870-020-2293-x>.

(46) Zhou, F.; Pichersky, E. More Is Better: The Diversity of Terpene Metabolism in Plants. *Current Opinion in Plant Biology* **2020**, *55*, 1–10. <https://doi.org/10.1016/j.pbi.2020.01.005>.

(47) Shimizu, N.; Koyama, T.; Ogura, K. Molecular Cloning, Expression, and Purification of Undecaprenyl Diphosphate Synthase: NO SEQUENCE SIMILARITY BETWEEN E- AND Z-PRENYL DIPHOSPHATE SYNTHASES *. *Journal of Biological Chemistry* **1998**, *273* (31), 19476–19481. <https://doi.org/10.1074/jbc.273.31.19476>.

(48) Fujihashi, M.; Zhang, Y.-W.; Higuchi, Y.; Li, X.-Y.; Koyama, T.; Miki, K. Crystal Structure of Cis-Prenyl Chain Elongating Enzyme, Undecaprenyl Diphosphate Synthase. *Proceedings of the National Academy of Sciences* **2001**, *98* (8), 4337–4342. <https://doi.org/10.1073/pnas.071514398>.

(49) Demissie, Z. A.; Erland, L. A. E.; Rheault, M. R.; Mahmoud, S. S. The Biosynthetic Origin of Irregular Monoterpenes in *Lavandula*. *J Biol Chem* **2013**, *288* (9), 6333–6341. <https://doi.org/10.1074/jbc.M112.431171>.

(50) Kumar, S.; Stecher, G.; Suleski, M.; Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* **2017**, *34* (7), 1812–1819. <https://doi.org/10.1093/molbev/msx116>.

- (51) Laville, R.; Castel, C.; Filippi, J.-J.; Delbecq, C.; Audran, A.; Garry, P.-P.; Legendre, L.; Fernandez, X. Amphilectane Diterpenes from *Salvia Sclarea*: Biosynthetic Considerations. *J. Nat. Prod.* **2012**, *75* (2), 121–126. <https://doi.org/10.1021/np2004177>.
- (52) Tawfik, D. S. Accuracy-Rate Tradeoffs: How Do Enzymes Meet Demands of Selectivity and Catalytic Efficiency? *Current Opinion in Chemical Biology* **2014**, *21*, 73–80. <https://doi.org/10.1016/j.cbpa.2014.05.008>.
- (53) Jia, M.; Peters, R. J. Cis or Trans with Class II Diterpene Cyclases. *Org. Biomol. Chem.* **2017**, *15* (15), 3158–3160. <https://doi.org/10.1039/c7ob00510e>.
- (54) Hamberger, B.; Ohnishi, T.; Hamberger, B.; Séguin, A.; Bohlmann, J. Evolution of Diterpene Metabolism: Sitka Spruce CYP720B4 Catalyzes Multiple Oxidations in Resin Acid Biosynthesis of Conifer Defense against Insects. *Plant Physiol.* **2011**, *157* (4), 1677–1695. <https://doi.org/10.1104/pp.111.185843>.
- (55) Bolger, A. M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30* (15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- (56) Grabherr, M. G.; Haas, B. J.; Yassour, M.; Levin, J. Z.; Thompson, D. A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; Chen, Z.; Mauceli, E.; Hacohen, N.; Gnirke, A.; Rhind, N.; di Palma, F.; Birren, B. W.; Nusbaum, C.; Lindblad-Toh, K.; Friedman, N.; Regev, A. Trinity: Reconstructing a Full-Length Transcriptome without a Genome from RNA-Seq Data. *Nat. Biotechnol.* **2011**, *29* (7), 644–652. <https://doi.org/10.1038/nbt.1883>.
- (57) Patro, R.; Duggal, G.; Love, M. I.; Irizarry, R. A.; Kingsford, C. Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression. *Nat. Methods* **2017**, *14* (4), 417–419. <https://doi.org/10.1038/nmeth.4197>.
- (58) Haas, B. J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P. D.; Bowden, J.; Couger, M. B.; Eccles, D.; Li, B.; Lieber, M.; MacManes, M. D.; Ott, M.; Orvis, J.; Pochet, N.; Strozzi, F.; Weeks, N.; Westerman, R.; William, T.; Dewey, C. N.; Henschel, R.; LeDuc, R. D.; Friedman, N.; Regev, A. De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis. *Nat. Protoc.* **2013**, *8* (8), 1494–1512. <https://doi.org/10.1038/nprot.2013.084>.
- (59) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. <https://doi.org/10.1038/msb.2011.75>.
- (60) Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, *30* (9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.

- (61) Letunic, I.; Bork, P. Interactive Tree Of Life (ITOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res* **2021**, *49* (W1), W293–W296. <https://doi.org/10.1093/nar/gkab301>.
- (62) Emanuelsson, O.; Nielsen, H.; Brunak, S.; von Heijne, G. Predicting Subcellular Localization of Proteins Based on Their N-Terminal Amino Acid Sequence. *Journal of Molecular Biology* **2000**, *300* (4), 1005–1016. <https://doi.org/10.1006/jmbi.2000.3903>.
- (63) Sainsbury, F.; Thuenemann, E. C.; Lomonossoff, G. P. PEAQ: Versatile Expression Vectors for Easy and Quick Transient Expression of Heterologous Proteins in Plants. *Plant Biotechnology Journal* **2009**, *7* (7), 682–693. <https://doi.org/10.1111/j.1467-7652.2009.00434.x>.
- (64) Morrone, D.; Chen, X.; Coates, R. M.; Peters, R. J. Characterization of the Kaurene Oxidase CYP701A3, a Multifunctional Cytochrome P450 from Gibberellin Biosynthesis. *Biochemical Journal* **2010**, *431* (3), 337–347. <https://doi.org/10.1042/BJ20100597>.
- (65) Sadre, R.; Kuo, P.; Chen, J.; Yang, Y.; Banerjee, A.; Benning, C.; Hamberger, B. Cytosolic Lipid Droplets as Engineered Organelles for Production and Accumulation of Terpenoid Biomaterials in Leaves. *Nat Commun* **2019**, *10* (1), 853. <https://doi.org/10.1038/s41467-019-08515-4>.
- (66) Earley, K. W.; Haag, J. R.; Pontes, O.; Opper, K.; Juehne, T.; Song, K.; Pikaard, C. S. Gateway-Compatible Vectors for Plant Functional Genomics and Proteomics. *Plant J* **2006**, *45* (4), 616–629. <https://doi.org/10.1111/j.1365-313X.2005.02617.x>.
- (67) Yang, J.; Yan, R.; Roy, A.; Xu, D.; Poisson, J.; Zhang, Y. The I-TASSER Suite: Protein Structure and Function Prediction. *Nat Methods* **2015**, *12* (1), 7–8. <https://doi.org/10.1038/nmeth.3213>.
- (68) Chan, Y.-T.; Ko, T.-P.; Yao, S.-H.; Chen, Y.-W.; Lee, C.-C.; Wang, A. H.-J. Crystal Structure and Potential Head-to-Middle Condensation Function of a Z,Z-Farnesyl Diphosphate Synthase. *ACS Omega* **2017**, *2* (3), 930–936. <https://doi.org/10.1021/acsomega.6b00562>.
- (69) Liu, M.; Chen, C.-C.; Chen, L.; Xiao, X.; Zheng, Y.; Huang, J.-W.; Liu, W.; Ko, T.-P.; Cheng, Y.-S.; Feng, X.; Oldfield, E.; Guo, R.-T.; Ma, Y. Structure and Function of a “Head-to-Middle” Prenyltransferase: Lavandulyl Diphosphate Synthase. *Angewandte Chemie International Edition* **2016**, *55* (15), 4721–4724. <https://doi.org/10.1002/anie.201600656>.

CHAPTER 3

Identifying Entry Steps in the Biosynthetic Pathway to Diterpenoid Alkaloids in *Delphinium grandiflorum*

Abstract

The roots from the *Aconitum* (Wolf's-Bane) and *Delphinium* (Larkspur) genera have been widely used in traditional medicine owing to the abundance of bioactive diterpenoid alkaloids that they produce. Hundreds of these compounds have been identified and characterized from both genera, and despite a wealth of studies on different medicinal properties of these metabolites, efforts towards total chemical synthesis, and publicly available transcriptomic data, very little progress has been made towards elucidation of the biosynthetic pathways for these compounds. The research described in this chapter presents the entry steps in the biosynthesis of these compounds, constituting seven enzymes identified from the Siberian Larkspur (*Delphinium grandiflorum*) through a combination of comparative transcriptomics between tissue types and genera and coexpression analysis. This pathway includes a pair of terpene synthases, four cytochrome P450s—three of which are the founding members of new subfamilies with one belonging to the poorly characterized CYP729 family—and a putative reductase with little homology to other characterized enzymes. Identification of these enzymes and production of a key intermediate in a heterologous host paves the way for biosynthetic production of this group of metabolites with promise for medicinal applications.

Introduction

Alkaloids are a diverse class of compounds broadly defined as nitrogen-containing specialized metabolites. Many examples of plant alkaloids have received considerable attention for their medicinal applications, leading to a wealth of research into elucidating their biosynthetic pathways and production in heterologous hosts. Prominent examples include alkaloids such as morphine¹ (analgesic), colchicine² (anti-inflammatory), scopolamine³⁻⁵ (anti-nausea), and vinblastine⁶⁻⁸ (anti-cancer). Much like terpenoids, the entry steps to the biosynthesis of many of these compounds involve an initial scaffold formation followed by modifications by enzymes such as P450s and methyl- and acetyltransferases. Rather than a carbocation-mediated cyclization of a single molecule as in terpenoid biosynthesis, the scaffold-forming step in alkaloid biosynthesis typically involves the accumulation and condensation of an amine and aldehyde precursor, followed by resolution of the resulting iminium cation to form an alkaloid scaffold⁹. Given the unique pathways towards initial scaffold formation, there is little overlap between the terpenoid and alkaloid classes of specialized metabolites.

One notable exception is the monoterpenoid indole alkaloids, derived from tryptophan and geranyl diphosphate (GPP). Decarboxylation of tryptophan into tryptamine leads to the accumulation of a primary amine, and conversion of GPP to secologanin leads to the accumulation of an aldehyde, which condense to form the initial scaffold towards monoterpenoid indole alkaloid metabolites⁸. Another exception are the diterpenoid alkaloids, which are found in at least 4 independent plant lineages¹⁰⁻¹²—most notably within the Ranunculaceae family^{13,14}. The biosynthesis of this class of metabolites has not been elucidated, however it is apparent from their structure that it involves the initial formation of a diterpene scaffold and nitrogen incorporation

follows, in contrast to the monoterpenoid indole alkaloids where the terpene precursor is not first cyclized by a terpene synthase and does not make up the majority of the scaffold⁸.

Plants from the *Aconitum* and *Delphinium* genera have long been used in traditional medicine due to the bioactivity of these diterpenoid alkaloids. The use of “Fuzi”—the processed lateral root of *A. carmichaelii* (more commonly known as Wolf’s Bane or Aconite)—has been documented for at least two thousand years¹⁴. The diterpenoid alkaloids have a wide range of applications from antifeedants to anti-cancer, choline esterase inhibitors, and analgesics^{13–16}. The therapeutic properties of many of these metabolites has prompted an extensive amount of research into total chemical synthesis of specific compounds^{17–21}, however the structural complexity of these compounds presents an enormous challenge in chemical synthesis. Aconitine (one such compound which is a potent neurotoxin), for example, contains six interconnected rings and fifteen stereocenters.

Elucidating the biosynthesis of these compounds would ameliorate some of the challenge in their production given the complexity of their scaffolds and number of required stereospecific oxidations. The lack of current knowledge in their biosynthesis is not for a lack of effort, as many previous attempts have been made to elucidate biosynthetic genes through transcriptomic analysis in various *Aconitum* species^{22–26}, with only one case published recently which characterized a pair of TPSs²⁷. As a result of all of these efforts, a wealth of public transcriptomic data is available for the *Aconitum* genus.

Similar to the strategy employed in Miller et al. 2020²⁸ (see Chapter 2), we carried out transcriptome sequencing on *Delphinium grandiflorum*, a plant from a neighboring genus to *Aconitum*. Transcriptome assembly both for *D. grandiflorum* and for three other *Aconitum* species (*A. carmichaelii*, *A. japonicum*, and *A. vilmorinianum*)—all of which accumulate diterpenoid

alkaloids—allowed for comparative transcriptomics across tissue types and genera, leading to the identification of six enzymes active in this pathway. Furthermore, the public data for *A. vilmorinianum*—a root tissue timecourse study²²—allowed for coexpression analysis, where top hits were simply searched back against our own *D. grandiflorum* transcriptome for cloning and characterization. This resulted in the identification of a seventh enzyme active in the pathway which has little homology to previously characterized enzymes.

This work demonstrates the utility of analyzing public data to augment the analysis of a single transcriptome, as the availability of these data were key to the identification of five out of the seven enzymes discovered (and likely necessary for at least three). Identification of these entry steps will serve as the basis for further pathway discovery towards diterpenoid alkaloid natural products and biosynthetic production in heterologous hosts.

Results

Proposal of an Initial Biosynthetic Pathway

The majority of diterpenoid alkaloids in the Ranunculaceae family can be divided into two major groups based on the number of carbons in their backbone structure (20 or 19) and ring structure (6/6/6/6 or 6/7/5/6, respectively)^{13,14}. Despite these differences, we proposed that both major groups are derived from the same diterpene starting scaffold. Two examples—the complex structure aconitine and a simple C20 hetidine-type diterpenoid alkaloid—are shown in Figure 3.1, and three structural features of these metabolites suggest a common origin. First, the cyclization pattern matches that of a class II TPS mechanism, with identical stereochemistry at three chiral centers indicated in green in Figure 3.1, suggesting the involvement of an *ent*-copalyl diphosphate (*ent*-CPP) synthase. Second, tracing from the same carbon in both examples shows two three-

carbon bridges making up two sides of a six-membered ring, similar to the structure of *ent*-atiserene²⁹. Third, the nitrogen is covalently bonded to the same methyl groups of the *ent*-atiserene backbone, indicating oxidative functionalization of the same two methyl groups—likely carried out by a pair of cytochrome P450s.

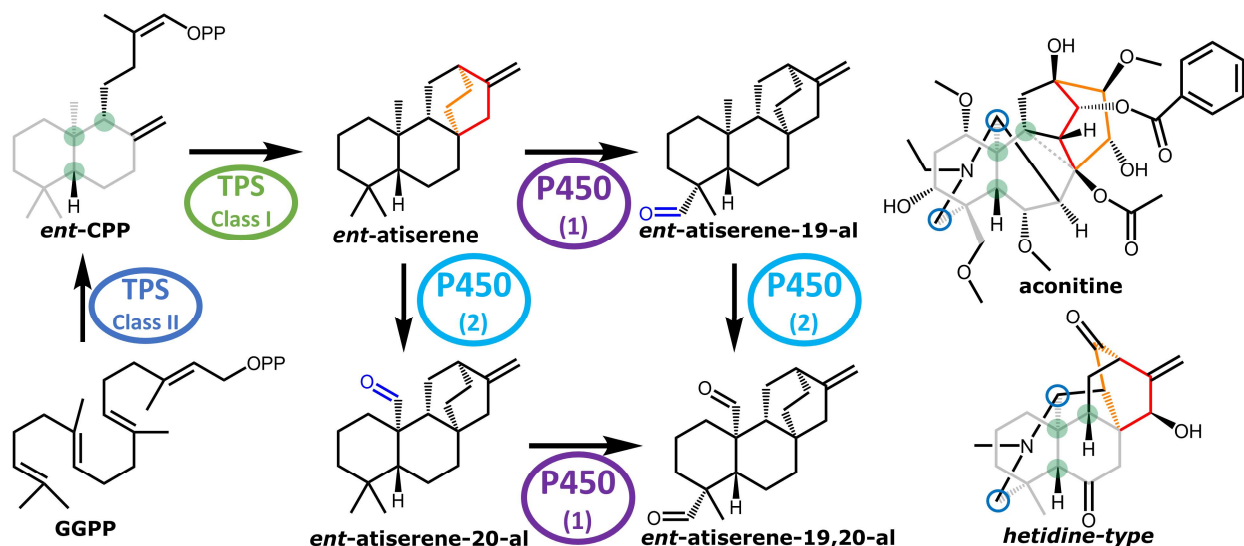


Figure 3.1: Common structural features of diterpenoid alkaloids and proposed biosynthetic pathway. Bonds shaded in gray highlight a common labdane structure likely derived from activity of a class II TPS (shown as a dotted line in aconitine due to a ring expansion proposed to happen further in the pathway). Carbons highlighted in green have common stereochemistry. Bonds highlighted in red/orange show the same three-carbon bridges that make up either side of a six-membered ring. Carbons circled in blue represent methyl groups on *ent*-atiserene which are likely converted to aldehydes to allow for nitrogen incorporation.

The proposed intermediate *ent*-atiserene-19-al closely resembles the central metabolite *ent*-kaurenoic acid—a key intermediate in the central metabolic pathway towards gibberellins³⁰—which is synthesized from GGPP through the activity of a class II/class I TPS pair and a cytochrome P450³⁰. Given these similarities, it is plausible that the genes responsible for making *ent*-atiserene-19-al are recent duplicates of these central metabolism enzymes, especially given the occurrence of polyploidization within the Delphinieae tribe (containing *Aconitum* and *Delphinium*) of the Ranunculaceae family^{31–33}.

RNA Sequencing and Transcriptome Assembly

Diterpenoid alkaloids primarily accumulate in root tissue throughout species in *Aconitum* and *Delphinium*^{34–37}. We isolated and sequenced RNA from *D. grandiflorum* from the roots, leaves, and flowers to allow for comparative transcriptomics across tissue types. Furthermore, a wealth of public RNA sequencing data has been submitted to the NCBI Sequence Read Archive (SRA) for the *Aconitum* genus, and three datasets from *A. carmichaelii* (root, leaf, flower, bud; PRJNA415989)²⁴, *A. japonicum* (root, root tuber, leaf, flower, stem; PRJDB4889), and *A. vilmorinianum* (root timecourse; PRJNA667080)²² were included as well. Transcriptomes for each species were assembled, allowing for multiple cross-tissue and cross-species comparisons to search for genes involved in diterpenoid alkaloid metabolism.

A Pair of TPSs Cyclizes GGPP to ent-atiserene

The first two steps in this pathway were proposed to be a pair of TPSs; first a class II TPS that converts GGPP to *ent*-CPP, and second a class I TPS which converts *ent*-CPP to *ent*-atiserene. At this stage, only the *D. grandiflorum* transcriptome had been assembled, and following analysis of this transcriptome, we began characterizing candidates without the need for data from the three other *Aconitum* species. A BLAST search of the *D. grandiflorum* transcriptome against a reference set of plant TPSs revealed fifteen putative TPS genes. Only three of these were exclusively expressed in root tissue, matching the tissue-specific accumulation of diterpenoid alkaloids. Phylogenetic analysis revealed that these belonged to the TPS-c, TPS-e, and TPS-b subfamilies (Figure 3.2). *Dgr*TPS1 (TPS-c) and *Dgr*TPS7 (TPS-e) appeared to be the most likely candidates, as they belong to the pair of subfamilies typically implicated in labdane-related diterpene biosynthesis. Furthermore, their closest paralogs (*Dgr*TPS2 and *Dgr*TPS5/6, respectively) have

low expression across all three tissues, as would be expected for the pair of TPSs involved in central metabolism for gibberellin biosynthesis.

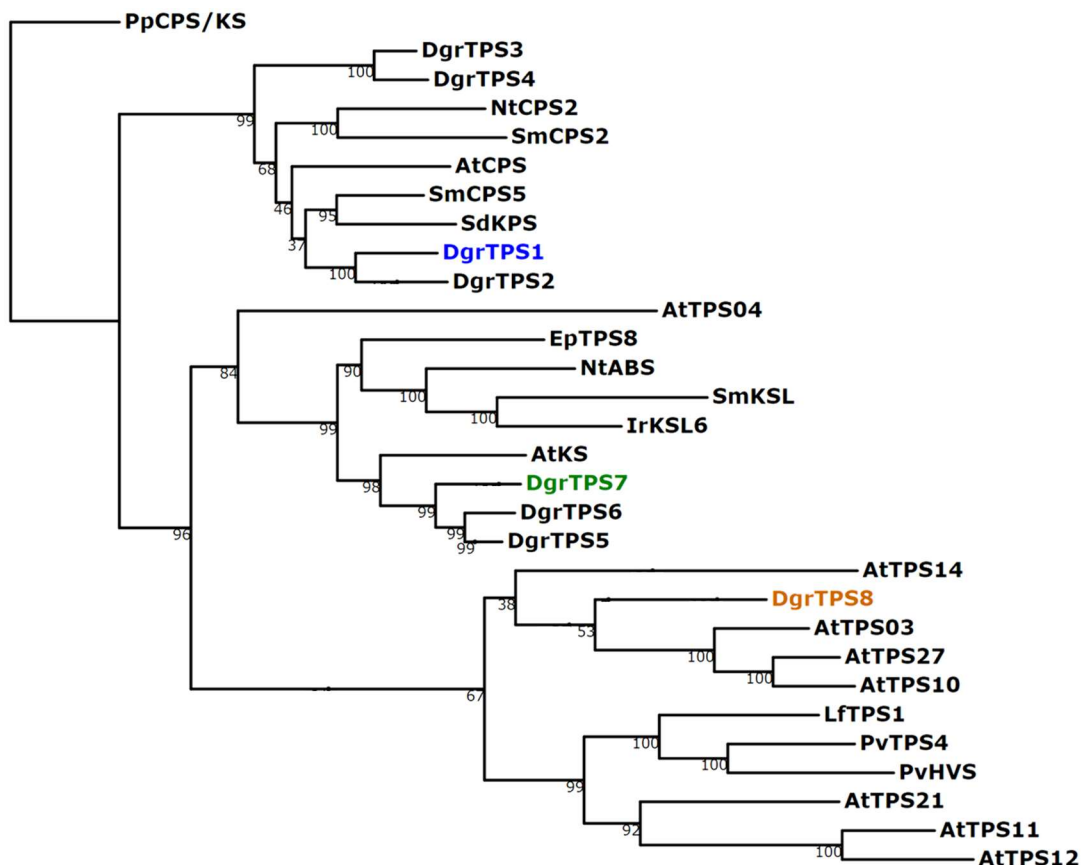


Figure 3.2: Maximum likelihood phylogenetic tree of predicted *D. grandiflorum* TPS sequences. Only eight out of the fifteen predicted sequences are shown, as many resulted in only partial transcripts with low coverage against reference sequences. Labels at branch points indicate percent bootstrap support from 1,000 replicates. Names in a color other than black had root-exclusive expression, and *DgrTPS1* and *DgrTPS7* were functionally characterized.

Full-length genes for *DgrTPS1* and *DgrTPS7* were cloned from *D. grandiflorum* root cDNA into pEAQ for transient expression in *N. benthamiana*. Two isoforms of *DgrTPS7*, not distinct in our transcriptome assembly, were cloned from cDNA, and both were tested (named *DgrTPS7a/7b*). All screening through transient expression in *N. benthamiana* throughout this chapter included coexpression with *CfDXS* and *CfGGPPS* (to increase precursor supply of GGPP³⁸). GC-MS analysis on hexane extracts revealed that of *DgrTPS1* acts as a copalyl

diphosphate (CPP) synthase. Coexpression of an enantioselective *ent*-kaurene synthase (*NmTPS2*) suggests an absolute stereochemistry consistent with *ent*-CPP (Figure 3.3).

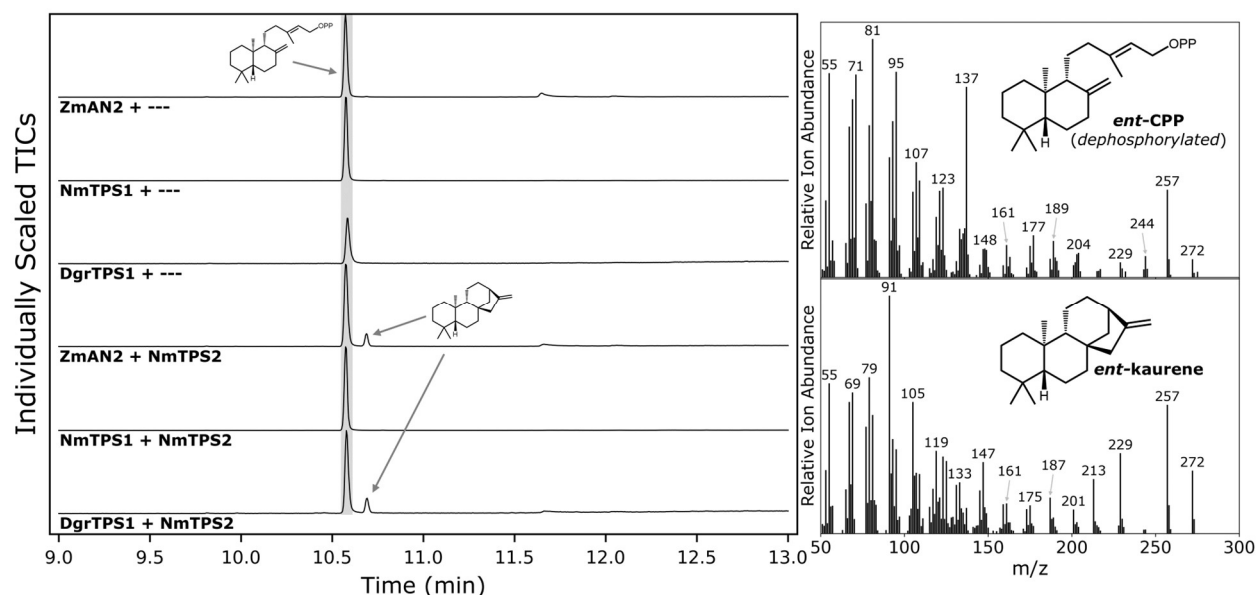


Figure 3.3: *DgrTPS1* is an *ent*-CPP synthase. (Left) Transient expression of *DgrTPS1* in *N. benthamiana* yields a product with the same retention time and mass spectrum as *ZmAN2* (*ent*-CPP synthase) and *NmTPS1* ((+)-CPP synthase; (+)-CPP is the enantiomer of the structure drawn for the highlighted region). The absolute stereochemistry of *DgrTPS1*'s product was determined through coexpression of an enantioselective *ent*-kaurene synthase (*NmTPS2*), which converts only the *ent* enantiomer of CPP to *ent*-kaurene. Each assay has *CfDXS* and *CfGGPPS* coexpressed in addition to those listed. (Right) Mass spectra (70 eV EI) of dephosphorylated *ent*-CPP and *ent*-kaurene.

Following this result, we tested *DgrTPS7a/7b* and showed conversion of *ent*-CPP to a new product with a fragmentation pattern matching that of *ent*-atiserene²⁹ for both isoforms (Figure 3.4). To confirm the identity of this new product as *ent*-atiserene, transient expression in *N. benthamiana* was scaled up with *DgrTPS1* and *DgrTPS7a*, and the product was purified through silica chromatography and confirmed through NMR (Table 3.S1 and Figure 3.S1). Since both isoforms of *DgrTPS7* were shown to have the same function, *DgrTPS7a* was used for further testing and is simply referred to as *DgrTPS7* throughout the remainder of this chapter.

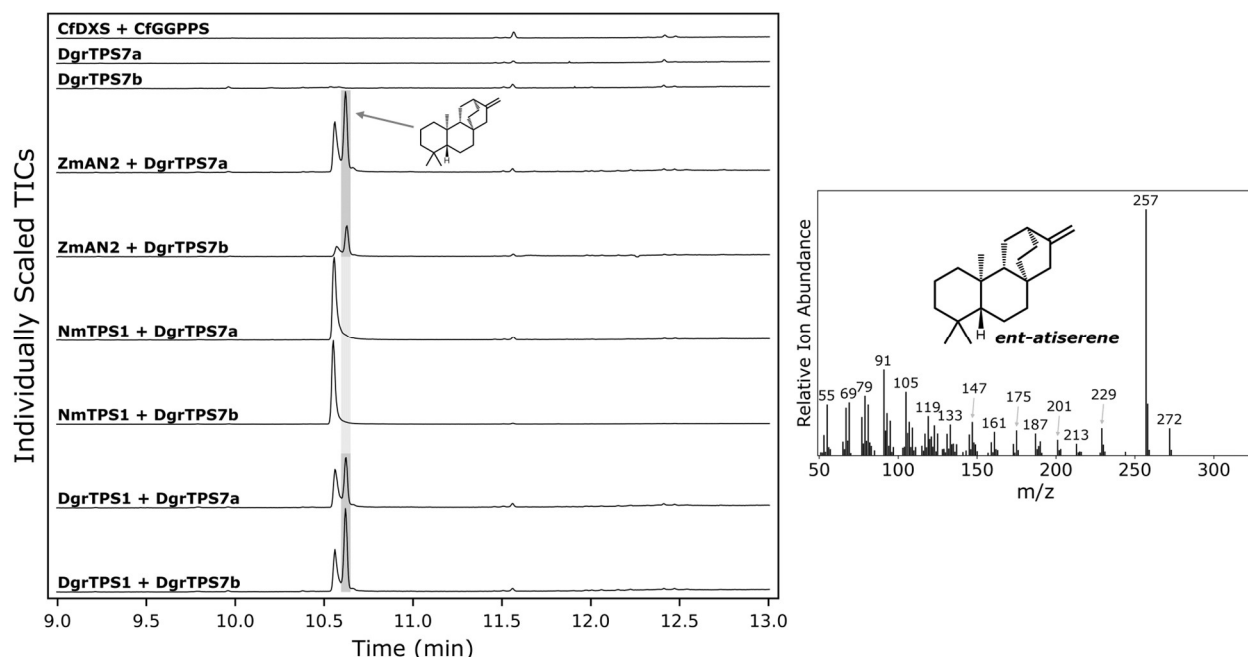


Figure 3.4: *DgrTPS7a* and *DgrTPS7b* convert *ent*-CPP to *ent*-atiserene. (Left) Transient expression of *DgrTPS7a* and *DgrTPS7b* yield *ent*-atiserene when coexpressed with an *ent*-CPP synthase (*ZmAN2* or *DgrTPS1*). *DgrTPS7a* and *DgrTPS7b* are also enantioselective and do not convert (+)-CPP (from *NmTPS1*) to a new product. Each assay has *CfDXS* and *CfGGPPS* coexpressed in addition to those listed. (Right) Mass spectrum (70 eV EI) of *ent*-atiserene.

Two Pairs of Cytochrome P450s With Overlapping Functions Oxidize ent-atiserene

Following the confirmation that a pair of terpene synthases make *ent*-atiserene, we continued with our proposed biosynthetic pathway to search for cytochrome P450s which can carry out sequential oxidations of methyl groups 19 and 20 to aldehydes. In contrast to the TPS family, the identification of P450s presents a challenge due to the number of genes that may be present in any given plant³⁹. In our transcriptome assemblies for *D. grandiflorum* and the three *Aconitum* species, a BLAST search against a reference set of P450 sequences yielded 2,061 predicted P450 transcripts. For *D. grandiflorum* alone, there were 297 after clustering shorter transcripts with greater than 95% sequence identity.

To narrow this down to a manageable number to test, we used a similar strategy to our previous work in identifying the P450 involved in the leubethanol pathway (Chapter 2)²⁸ by taking

advantage of the assumed conservation of this pathway between neighboring genera and tissue-specific accumulation of metabolites. The total transcripts from each assembly were first assigned to individual clans based on homology to the closest reference sequence, and individual phylogenies were made for distinct clans (Figures 3.S2-5). We then filtered these transcripts to include only those in *D. grandiflorum* with high root expression and with a root-expressed ortholog in each *Aconitum* assembly. This narrowed down a list of 297 possible P450s to just 7 to test.

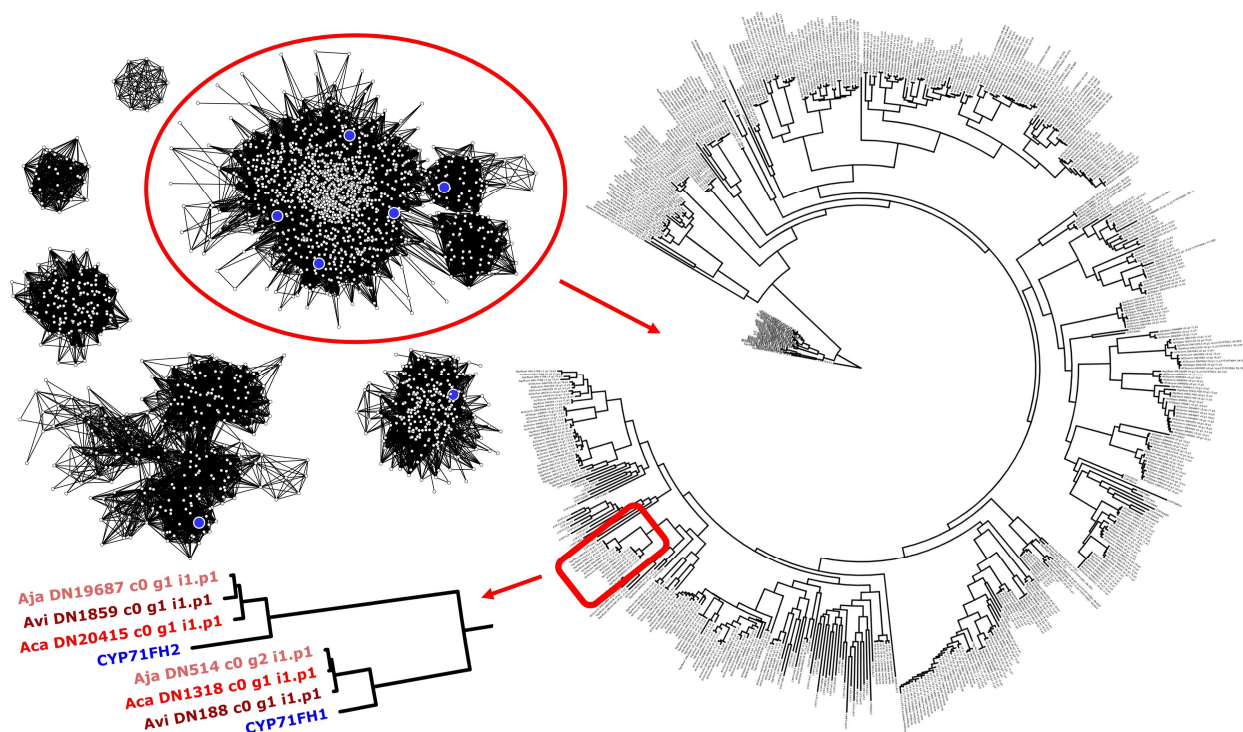


Figure 3.5: Process of filtering Cytochrome P450 transcripts for candidate selection. (Top Left) Sequence similarity network of P450 transcripts from each assembled transcriptome. Nodes (2,061) represent individual sequences and edges (211,127) represent a sequence identity of at least 32%. Large blue nodes represent candidates which were selected for testing. Individual clusters represent separate P450 clans. The largest clan (CYP71) is circled in red. Individual phylogenetic trees were generated for the four largest clusters. (Right) Example phylogenetic tree of the CYP71 clan. Highlighted in red is one section of the tree with two candidate P450s from *D. grandiflorum*. (Bottom Left) Highlighted portion of the CYP71 phylogenetic tree showing two *D. grandiflorum* P450 candidates (CYP71FH1 and CYP71FH2), which both met our filtering criteria by having high expression in root tissue and respective orthologs in each *Aconitum* assembly (Aja = *A. japonicum*; Avi = *A. vilmorinianum*; Aca = *A. carmichaelii*).

These seven P450s were cloned from *D. grandiflorum* root cDNA and tested through transient expression in *N. benthamiana*. Each candidate was coexpressed with *DgrTPS1* and *DgrTPS7*, and products were analyzed via GC-MS following ethyl acetate extraction. CYP701A127 and CYP71FH1 both showed activity in oxidizing the *ent*-atiserene backbone (Figure 3.6). Coexpression with either of these P450s showed a depletion in *ent*-atiserene and the production of respective metabolites with a molecular ion at m/z 286 and retention of m/z 257 as the highest abundance fragment ion (Figure 3.6), consistent with sequential oxidations of a methyl group to an aldehyde. Both enzymes also made a product with a molecular ion at m/z 302 (compounds **A** and **B**; Figure 3.S6), consistent with either a third oxidation of this carbon to an

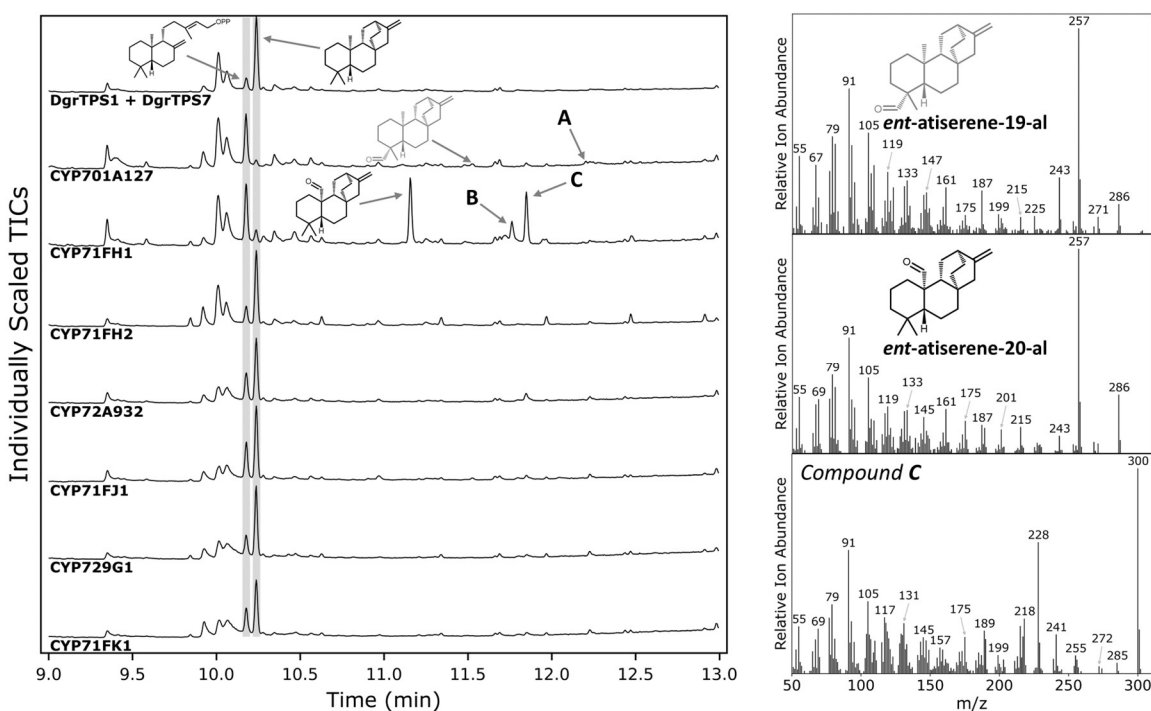


Figure 3.6: CYP701A127 and CYP71FH1 convert *ent*-atiserene to oxidized products. (Left) Coexpression of either CYP701A127 and CYP71FH1 with *DgrTPS1* and *DgrTPS7* show depletion of *ent*-atiserene and production of oxidized products, while the remainder of candidates show a similar accumulation of *ent*-atiserene to *DgrTPS1* and *DgrTPS7* alone. CYP701A127 likely makes *ent*-atiserene-19-al (as such is drawn in gray; described in the text). CYP71FH1 makes *ent*-atiserene-20-al (confirmed by NMR) and another major product (C) with an unsolved structure. Each assay has *CjDXS*, *CjGGPPS*, *DgrTPS1*, and *DgrTPS7* coexpressed in addition to those listed. (Right) Mass spectra of both aldehyde products and unknown C. Mass spectra for minor products **A** and **B** have a molecular ion of 302 m/z and are given in Figure 3.S6.

acid or addition of another hydroxyl group elsewhere. CYP71FH1 also produces a major product with a molecular ion at m/z 300 (compound **C**), which would suggest a net addition of two oxygen atoms and four oxidations from *ent*-atiserene.

For the products of CYP71FH1, we scaled up production in *N. benthamiana* to purify compounds and attempt to solve structures by NMR. While sufficient quantities were simple to produce through expression and extraction from roughly 30 g of fresh weight, purification of the two major products from each other proved challenging. One fraction purified through a silica column was sufficiently enriched for the m/z 286 product that we could confirm its identity as *ent*-atiserene-20-al through NMR (Figure 3.S7-8). The structure of the m/z 300 product was not determined. The products of CYP701A127 gave weak signals by GC and may have been shuttled away to other products through conversion by endogenous *N. benthamiana* enzymes. We tentatively assigned CYP701A127's product as *ent*-atiserene-19-al based on the mass spectrum both in terms of its own fragmentation pattern and in comparison to similar structures in the NIST database (Figure 3.S6), close retention time to *ent*-atiserene-20-al, and phylogenetic evidence that CYP701A127 is a recent duplication of its putative central metabolism paralog (likely an *ent*-kaurene oxidase that oxidizes this same carbon).

In our proposed biosynthetic pathway, we figured that a pair of P450s could work together to oxidize both methyl groups at carbons 19 and 20 to aldehydes, and so we tested whether coexpression of both of these enzymes would further the pathway. Ethyl acetate extraction and GC-MS analysis on both TPSs and P450s coexpressed revealed a depletion of both *ent*-atiserene and of both P450's respective products (Figure 3.7). These assays were also analyzed by LC-MS on 80% methanol extracts, which revealed two products from CYP701A127 (compounds **D** and **E**), four from CYP71FH1 (compounds **F-I**), and a total of five products with coexpression of both

enzymes (Figure 3.7 and Figure 3.S9). Four of the products present with both P450s coexpressed are an accumulation of CYP71FH1's products (compounds **F-I**, including its major product **G**), suggesting that these are products different than those detected by GC-MS for CYP71FH1 alone, and that CYP701A127 may share a partial functional redundancy with CYP71FH1. One additional minor product is present (compound **J**) when both are coexpressed.

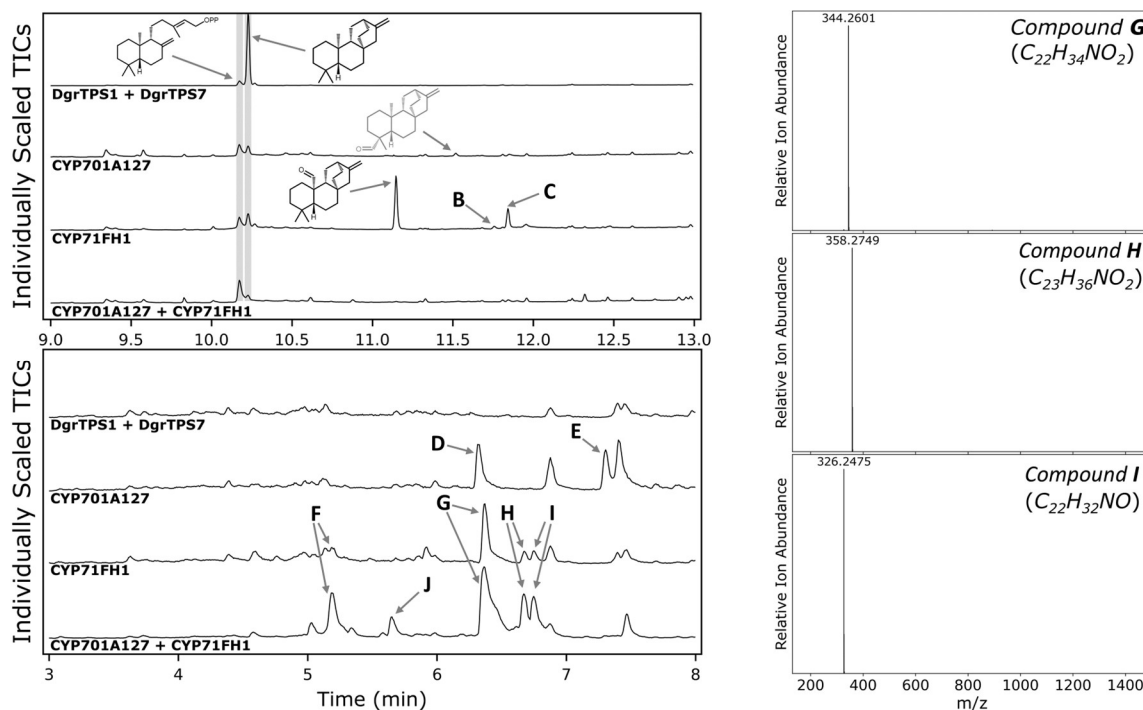


Figure 3.7: Coexpression of CYP701A127 and CYP71FH1 lead to an accumulation of the same products. (Left) GC-MS (top panel) and LC-MS (bottom panel) analysis of CYP701A127 and CYP71FH1 coexpression. Individual products of either enzyme detectable by GC-MS are no longer present when both are coexpressed. Products detectable by LC-MS for CYP701A127 are depleted upon coexpression of both P450s, however those for CYP71FH1 accumulate. One additional peak is seen upon coexpression of both enzymes (compound **J**). Each assay has *Cf*DXS, *Cf*GGPPS, *Dgr*TPS1, and *Dgr*TPS7 coexpressed in addition to those listed. (Right) Mass spectra and predicted chemical formulas for three products. Mass spectra for products not shown here are given in Figure 3.S9.

We further characterized this pair of P450s against the remaining five candidates. Coexpression of both TPSs, both P450s, and each remaining P450 candidate revealed that both CYP729G1 and CYP71FK1 can act on these products (Figures 3.8 and 3.S10). The molecular ions for each product suggest that they are each a single hydroxylation difference (additional ~16 m/z)

from major products for CYP701A127 and CYP71FH1 alone. Interestingly, despite these enzymes being evolutionarily distant (belonging to entirely different clans), both give the same product profile, with the exception of one additional product present with coexpression of CYP729G1 (compound **L**) which is not present with CYP71FK1.

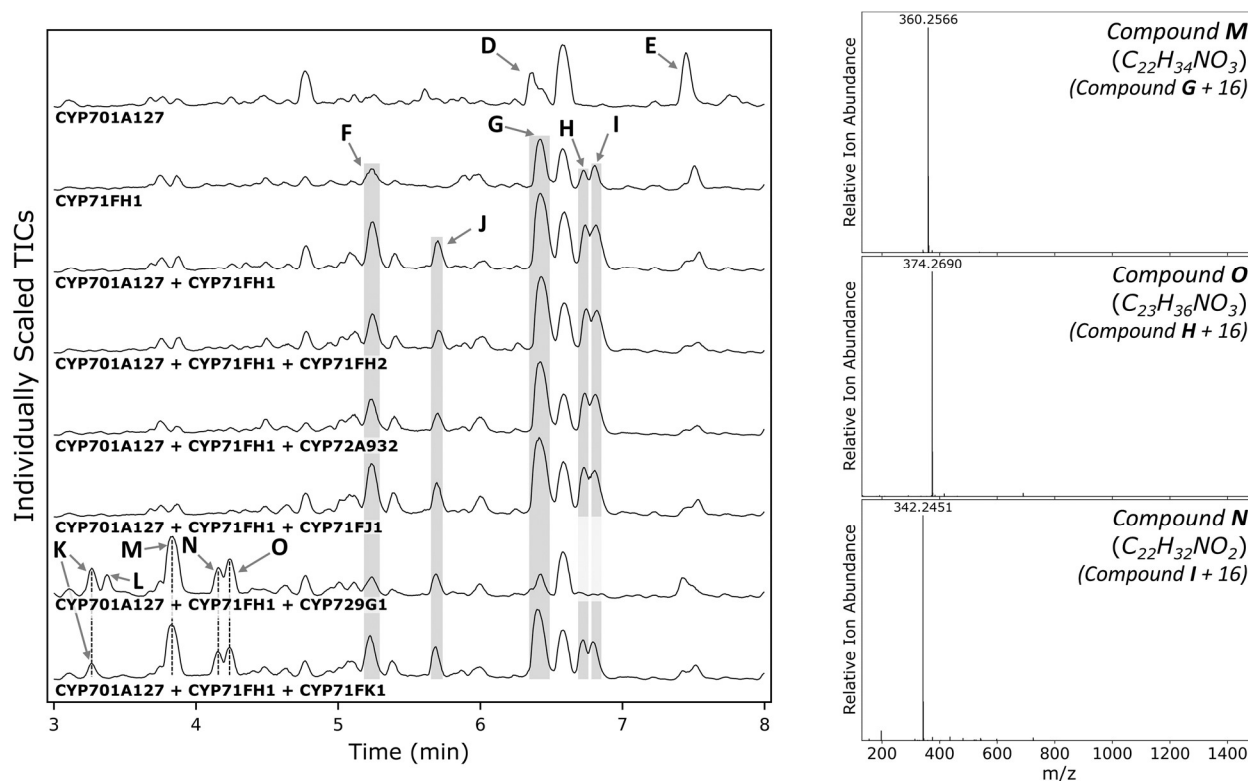


Figure 3.8: *CYP729G1* and *CYP71FK1* have redundant functions. (Left) GC-MS (top panel) and LC-MS (bottom panel) analysis of CYP701A127 and CYP71FH1 coexpression. Individual products of either enzyme detectable by GC-MS are no longer present when both are coexpressed. Products detectable by LC-MS for CYP701A127 are depleted upon coexpression, however those for CYP71FH1 accumulate. One additional peak is seen upon coexpression of both enzymes (compound **J**). Each assay has *Cf*DXS, *Cf*GGPPS, *Dgr*TPS1, and *Dgr*TPS7 coexpressed in addition to those listed. (Right) Mass spectra and predicted chemical formulas for three products. Mass spectra for products not shown here are given in Figure 3.S10.

Continuation of the Previously Proposed Biosynthetic Pathway

Accumulation of intermediates and side products is likely to occur when pathways are incompletely reconstructed or artificially altered^{3,40}, and we considered that the abundance of products from these four P450s may be due to an accumulation of intermediates which would not

occur with the coexpression of subsequent steps in the pathway. We therefore chose to continue with the pathway through screening additional candidates, as it is unclear whether the presence of multiple products is a facet of the actual pathway or simply a result of an incomplete reconstruction of the pathway and/or endogenous activity of enzymes present in *N. benthamiana*.

Considering that CYP701A127 and CYP71FH1 carry out the oxidations proposed in our initial biosynthetic pathway required for nitrogen incorporation, we proposed that this incorporation likely follows these two steps. In many alkaloid biosynthetic pathways, the formation of an alkaloid scaffold typically involves the accumulation of both an amine and aldehyde precursor⁹. The nitrogen present in the majority of diterpenoid alkaloids in *Aconitum* and *Delphinium* appears to be derived from ethylamine due to the attached -CH₂CH₃ group (Figure 3.9), while some metabolites presumably incorporate methylamine (-CH₃) or ethanolamine (-CH₂CH₂OH)^{13,14}—the origin of which could come from decarboxylation of alanine, glycine, or serine, respectively. Serine decarboxylases are present in central metabolism, and a duplication of this enzyme in *Camellia sinensis* has been shown to decarboxylate alanine into ethylamine (AlaDC) in theanine biosynthesis⁴¹. Additionally, *Spirea japonica*—an evolutionarily distinct plant which makes similar compounds—has been shown to produce isotopically labeled diterpenoid alkaloids through addition of labeled serine⁴².

The mechanism of nitrogen incorporation is also an important consideration, as the iminium cation formed through condensation of an amine and aldehyde is inherently unstable. Quenching of this cation through either a substitution or reduction⁹ can avoid spontaneous hydrolysis separating them back into their constituent parts, and in the case of diterpenoid alkaloids, it likely follows both mechanisms based on the number of bonds present on both oxidized methyl groups (Figure 3.9). Carbon 20 almost always contains an extra carbon-carbon

bond relative to *ent*-atiserene and the intermediate *ent*-atiserene-20-al, while carbon 19 does not, similar to both *ent*-atiserene and the intermediate *ent*-atiserene-19-al. This suggests that incorporation at carbon 19 requires a reductase, and at carbon 20 may involve a spontaneous intra-molecular condensation.

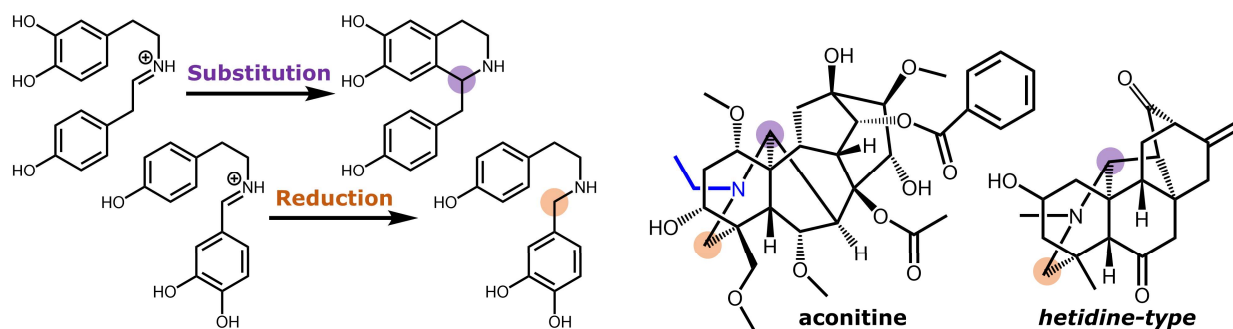


Figure 3.9: Nitrogen incorporation into diterpenoid alkaloids likely involves iminium cation resolution through reduction and substitution. (Left) Example highlighted by Lichman 2021⁹ showing how the iminium cation in norcoclaurine biosynthesis is resolved through substitution (top reaction - purple), while similar compounds from the Amaryllidaceae family involve a reduction (bottom reaction - orange). (Right) representative compounds from *Delphinium* and *Aconitum* with carbons highlighted corresponding to the proposed reaction mechanism shown on examples on the left. Highlighted in blue is the portion of aconitine proposed here to originate from ethylamine—present in the majority of diterpenoid alkaloids.

In contrast to the steps elucidated thus far, involving carbocation-mediated cyclizations (TPSs) and site-specific oxidations (P450s), the reaction of an amine and aldehyde to form an alkaloid scaffold could occur either spontaneously or through enzyme catalysis given the inherent reactivity between aldehydes and primary amines. The putative involvement of a reductase is also not straightforward in terms of how many different enzyme families this function could evolve from. To search for the next step(s), we decided to carry out coexpression analysis and see which genes were coexpressed with the first four enzymes already found in the pathway (*Dgr*TPS1, *Dgr*TPS7, CYP701A127, and CYP71FH1). Given that our own data for *D. grandiflorum* only contained single replicates of each tissue type, and that these metabolites accumulate in only one of them, we carried out this analysis on public data instead. The data collected for *A. vilmorinianum*

involved sequencing three replicates of root tissue at three different stages of development²², and so we carried out coexpression analysis on this dataset and simply BLAST searched the top hits back against our set of four transcriptomes (Figure 3.10). We found three putative reductases which were highly coexpressed with the *A. vilmorinianum* orthologs of our four initial pathway genes, and one putative cupin (named here simply as VGCRed, OxoRed, SangRed, and Cupin, respectively).

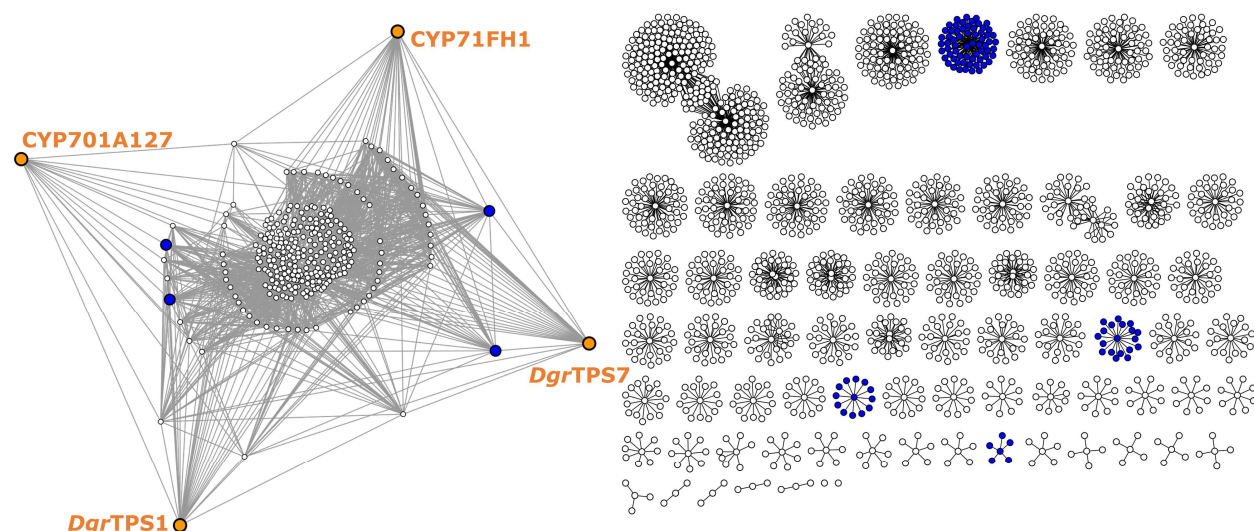


Figure 3.10: Coexpression analysis on *Aconitum vilmorinianum* and BLAST search back against the four *Delphinium/Aconitum* transcriptome assemblies. (Left) Coexpression network showing all *A. vilmorinianum* genes coexpressed with the respective orthologs of the first four steps characterized in the pathway (“anchor sequences” in orange). Nodes represent assembled transcripts and edges represent coexpression between genes determined by mutual rank (MR; cutoff: $e^{-(MR-1)/5} > 0.01$)⁴³. Genes included in this network either meet this threshold with one of the anchor sequences or with another gene that does (i.e. two degrees of separation). Nodes further from the center represent genes that meet this coexpression threshold with a greater number of anchor sequences; nodes in the center do not meet the cutoff threshold directly with any anchor sequence. Highlighted in blue are the four candidates selected for characterization. (Right) Visual representation of a BLAST search of each node in coexpression network against all four transcriptome assemblies to find orthologous genes. Nodes represent individual sequences and edges represent a sequence identity of at least 70%. Highlighted in blue are the clusters of transcripts for each candidate shown in blue in the coexpression network.

Coexpression Analysis Reveals that a Predicted Reductase is Active in the Pathway

Each of these four genes were cloned from *D. grandiflorum* root cDNA and tested for activity through transient expression in *N. benthamiana*. The alanine decarboxylase (AlaDC) from

*C. sinensis*⁴¹ was also included to supply ethylamine to the pathway, both to see if new metabolites spontaneously form with our aldehyde intermediates and to ensure that our coexpression candidates, if required, have access to ethylamine. Testing of each candidate was carried out along with either the first four enzymes (*DgrTPS1*, *DgrTPS7*, CYP701A127, and CYP71FH1) or these four plus CYP729G1.

Two major results came from coexpression of these candidates with the first four enzymes (Figure 3.11). First, coexpression of AlaDC resulted in a minor product with a proposed neutral formula of $C_{22}H_{33}NO$ (exact mass 328.2647 in ESI+; calculated m/z 328.2635). Second, coexpression of SangRed led to nearly a complete depletion in precursors and the formation of a new peak with an exact mass identical to the minor product from AlaDC. Coexpression of SangRed along with the first four steps and CYP729G1 did not deplete all of CYP729G1's products, however did lead to the formation of a new peak with a proposed neutral formula of $C_{22}H_{33}NO_2$ (exact mass 344.2611 in ESI+; calculated m/z 344.2584), suggesting that both of these enzymes compete for the products of the first four enzymes, while CYP729G1 can still hydroxylate the product of SangRed (or conversely that SangRed can convert the product of CYP729G1). Similar to the previous results with just the first four enzymes, coexpression with AlaDC led to the formation of a minor product with an identical exact mass (344.2611). Coexpression of both AlaDC and SangRed together along with the first four enzymes (or also CYP729G1) did not lead to an obvious increase in SangRed products, suggesting that ethylamine is not a substrate. Further testing revealed that SangRed produces its major product without the need for CYP701A127 and that CYP71FK1 retains its functional redundancy with CYP729G1, even in combination with SangRed (Figure 3.S11).

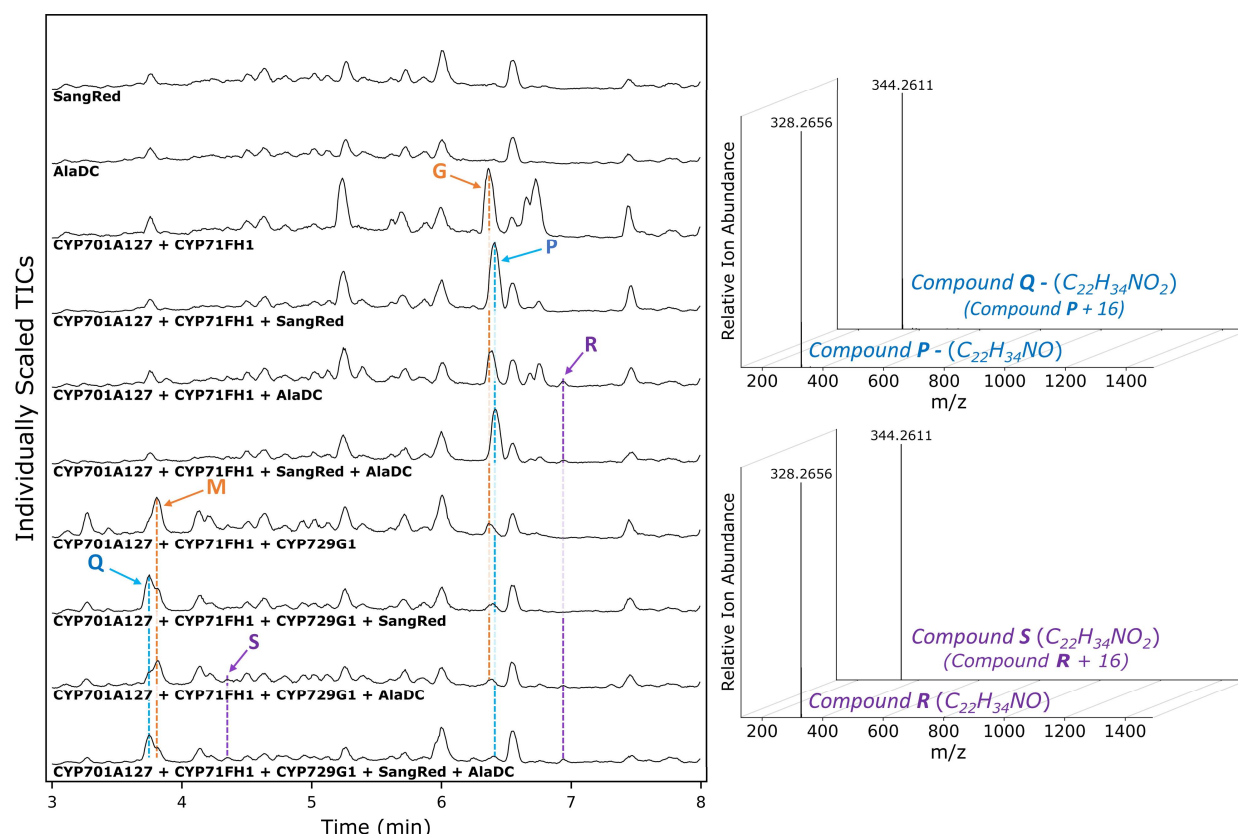


Figure 3.11: Coexpression with SangRed produces an isomer of what is produced upon supplementation with ethylamine. (Left) LC-MS analysis of SangRed and AlaDC coexpression with previous steps of the pathway. Products **G**, **H**, and **I** from the first four enzymes are depleted upon coexpression with SangRed, and a new product **P** is made. Compound **P** has an identical exact mass to a minor product **R**, which is made through coexpression of AlaDC. A new compound **Q** is made through coexpression of SangRed with the first four enzymes and CYP729G1. Compound **Q** similarly has an identical exact mass to a minor product **S**, which is made through coexpression of AlaDC. Each assay has *CfDXS*, *CfGGPPS*, *DgrTPS1*, and *DgrTPS7* coexpressed in addition to those listed. (Right) Mass spectra highlighting compounds made through coexpression with either SangRed or AlaDC and a putative difference of one hydroxylation upon addition of CYP729G1.

Discussion

Through a combination of transcriptomics comparing tissue types and genera and coexpression analysis, we have identified seven enzymes active in the biosynthetic pathway towards diterpenoid alkaloids in the Ranunculaceae family. There are hundreds of diterpenoid alkaloids in this family, and the identification of these enzymes will serve as the basis for further pathway discovery towards specific metabolites. This work highlights the usefulness of utilizing

public data as an orthogonal filter for selection of candidate enzymes beyond the analysis of a single species, as it likely would not have been possible to identify all of these enzymes otherwise given the inherent complexity of these pathways.

Immediately following our characterization of the TPS pair which makes *ent*-atiserene, Mao et al. 2021²⁷ published a characterization of the entire TPS family from *A. carmichaelii* and identified enzymes orthologous to ours. Given how straightforward the identification of these two TPSs was in comparison to the following five steps, it is perhaps surprising that so many preceding studies attempted to identify genes involved in this pathway but did not^{22–26}. Mao et al.²⁷ commented that recent attempts may have been limited by incomplete transcript assembly of genes putatively annotated as being involved in the pathway, and as such combined two different methods of RNA sequencing (Illumina and PacBio) to assemble their transcriptome. In fact, in our initial assembly, transcripts for *Dgr*TPS1 and *Dgr*TPS7 were truncated. While we didn't troubleshoot why our assembly pipeline was unable to assemble them correctly the first time, we were confident that these candidates were correct based on their expression pattern and phylogenetic origin, and we simply reassembled the transcriptome with a limited dataset (see Methods) and both genes were assembled properly.

One possible explanation for these assembly artifacts is that the genetics of members of the *Delphinium* and *Aconitum* genera are inherently complicated. *Delphinium montanum*, for example, is an autotetraploid with a predicted genome size of roughly 40 Gb³³ ($2n = 32$ ⁴⁴). The four species studied here have a range of predicted ploidy levels (*D. grandiflorum*: $2n = 16$; *A. carmichaelii*: $2n = 32/64$ – depending on cultivar; *A. japonicum*: $2n = 32$; *A. vilmorinianum*: $2n = 16$)⁴⁴, and it has been suggested that, at least in the *Aconitum* genus, there may have been multiple recent events of polyploidization and diploidization³². This fits with the model of our initial biosynthetic

pathway—and the phylogenetic relationships of these genes—in which we predicted that the first three steps may be recent duplications of central metabolism enzymes given the similarity of these predicted intermediates to those in gibberellin biosynthesis³⁰. While we didn't characterize the putative central metabolism copies of these genes, Mao et al.²⁷ demonstrated a pair of recently-duplicated *ent*-CPP synthases and *ent*-kaurene/atiserene synthases in their analysis. CYP701A127, which we assigned as an *ent*-atiserene oxidase (making *ent*-atiserene-19-al) also belongs to the same family as CYP701A3, the *ent*-kaurene oxidase involved in central metabolism in *Arabidopsis*⁴⁵.

It should be noted that *DgrTPS1*—being an *ent*-CPP synthase—is technically not an enzyme which makes a specialized metabolite. Given its relative expression (~75x higher in roots) over its putative central metabolism paralog (*DgrTPS2*), however, it is clearly dedicated to specialized metabolism. A similar phenomenon is seen in both *Oryza sativa*⁴⁶ and *Zea mays*⁴⁷, where two copies of an *ent*-CPP synthase are present; one which is involved in gibberellin biosynthesis and another which is inducible by pathogens for the production of defensive *ent*-CPP-derived specialized metabolites. Given the presence of duplicate *ent*-CPP synthases in each of these independent lineages of plants, there is likely a strong evolutionary pressure for the ability to tightly regulate these competing pathways.

Throughout the process, we varied the approach to identify each class of enzyme based on what information was necessary. For the terpene synthases, for example, few enough transcripts were present in our assembly that we relied solely on data from *D. grandiflorum*, as the choice of candidates to test was obvious given just this single dataset. For the P450s, the *Aconitum* datasets were essential given the presence of nearly 300 unique transcripts in our *D. grandiflorum* assembly. Had we not chosen to work with a neighboring genus, we may not have been able to

filter candidates down to just seven that we tested, as the only orthologous genes present across each species in our analysis have persisted throughout roughly 27 million years since the speciation of the two genera⁴⁸. Notably, three of the P450s shown to be active are founding members of new subfamilies (denoted by the ending of “1”). Finally, even with tissue and species-specific transcriptomic data, the following steps were not obvious, and so coexpression analysis allowed us to search for new candidates without prior knowledge of which enzyme families to search.

Throughout the process of characterizing various steps in the pathway, not every intermediate product was identified. Often it can be difficult to differentiate “actual” intermediates in terms of whether the observed products are relevant to the pathway or simply a result of an incomplete reconstruction or a heterologous host’s interference of the native pathway. In the process of discovering the forskolin pathway, for example, coexpression of an incomplete set of genes in *N. benthamiana* led to an accumulation of many side products that did not occur once the entire pathway was reconstructed (five P450s acting on a single diterpene scaffold and at least sixteen total products)⁴⁰. A similar example can be seen with accumulation of precursors and side products for the scopolamine pathway in *A. belladonna* following virus-induced gene silencing of various pathway steps³. We identified the activity of the two TPSs and confirmed our predicted activity of two P450s, but following this confirmation, we decided to test enzymes in different combinations to identify new steps in case the side products seen were similar artifacts. Given the emergence of a single product upon coexpression with SangRed, this is the most obvious target for structural elucidation to further the pathway.

The presence of a minor product forming upon coexpression with AlaDC was expected based on the presence of aldehydes in our intermediates, however the amount of product that would form was uncertain. We proposed that ethylamine was the source of nitrogen in this pathway,

however if that is the case, it is likely enzyme-catalyzed based on the poor conversion resulting from spontaneous condensation. It is more likely, however, that it follows a different mechanism than is proposed, as the product of SangRed converts nearly all of the products of CYP701A127 and CYP71FH1 to a single product which is likely an isomer of this spontaneous condensation based on an identical exact mass but differing retention time. The substrates and mechanism of SangRed is still unknown, and difficult to predict given its low degree of homology to other characterized enzymes.

Beyond the immediate questions emerging from this progress, more challenges remain in the discovery of diterpenoid alkaloid pathways. Perhaps most important is the differentiation between the C20 and C19 metabolites, and when this occurs. The greatest challenge will likely be the reconstruction of an entire pathway to a specific product, rather than the initial scaffold-forming steps investigated here which are presumably common to each metabolite. Based on the structure of aconitine (see Figure 3.1), there are potentially twenty or more steps involved in its biosynthesis. Further pathway discovery of such downstream steps will likely require different methodology employed here given the species-specificity of some of these products. This may benefit from accurate cross-species metabolomic data to differentiate chemical conversions present in distinct lineages in conjunction with the cross-species transcriptomics employed here.

Materials and Methods

Plant material, RNA isolation, and cDNA synthesis

D. grandiflorum plants were grown in a greenhouse under ambient photoperiod and 24°C day/17°C night temperatures. RNA isolation from flowers, leaves, and roots, quality assessment,

RNA sequencing, and cDNA synthesis was carried out as described in Miller et al. 2020²⁸ (in parallel with samples prepped for *L. frutescens*; see Chapter 2).

D. grandiflorum and Aconitum genera de novo transcriptome assembly and analysis

RNA-seq data were obtained through RNA sequencing on an Illumina HiSeq 4000 for *D. grandiflorum* and the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) for *A. carmichaelii* (PRJNA415989)²⁴, *A. japonicum* (PRJDB4889), and *A. vilmorinianum* (PRJNA667080)²². Transcriptome assembly and analysis was carried out exactly as described in Miller et al. 2020²⁸ (see Chapter 2), with the exception of adaptor trimming, which was done with TrimGalore (v0.6.5; <https://github.com/FelixKrueger/TrimGalore>). CD-HIT (v4.8.1)^{50,51} was used for clustering of *D. grandiflorum* P450 sequences. Sequence similarity networks were made with BLAST (v2.7.1+) and visualized with Cytoscape⁵².

Initial assembly of the *D. grandiflorum* transcriptome resulted in incomplete transcripts for *DgrTPS1* and *DgrTPS7* (only ~75% coverage of reference sequences), and although this was prior to our characterization of these enzymes, we noted that these transcripts were most likely misassembled given their high expression and likelihood of being involved in the pathway. Reassembly of the *D. grandiflorum* transcriptome was therefore done with only data acquired from root tissue, with reads from each tissue type mapped to this assembly. Transcripts for both of these genes in the new assembly aligned to the entire length of reference sequences, and so this assembly was used for further analysis.

Coexpression analysis

Our assembly for *A. vilmorinianum* was used for coexpression analysis. To minimize the computational burden, we reduced the analysis through clustering by 99% identity with CD-HIT (v4.8.1)^{50,51}, calculated expression levels through mapping reads to this clustered transcriptome, and eliminated any transcript with no samples that had at least 20% the expression level (in TPM) as any sample for either TPS. Coexpression analysis was carried out as described by Wisecaver et al. 2017⁴³ (pipeline at: <https://github.itap.purdue.edu/jwisecav/mr2mods>). The resulting coexpression network shown in Figure 3.10 shows only genes with one or two degrees of separation from any of the first four genes in the pathway (*DgrTPS1*, *DgrTPS7*, *CYP701A127*, and *CYP71FH1*) based on a mutual rank (MR) cutoff of $e^{-(MR-1)/5} > 0.01$. Orthologs from each transcriptome were found with BLAST (v2.7.1+) and visualized with Cytoscape⁵².

Cloning

PCR amplification from cDNA, cloning, and constructs used for transient expression in *N. benthamiana* were carried out as described in Miller et al. 2020²⁸ for plastidial tests with GGPP (see Chapter 2). Constructs for *ZmAN2*, *NmTPS1*, and *NmTPS2* in pEAQ (used as positive controls for *ent*-CPP, (+)-CPP, and *ent*-kaurene biosynthesis, respectively) were made by Johnson et al. 2019⁵³.

Transient expression in N. benthamiana, product scale-up, and NMR analysis

Transient expression in *N. benthamiana* for screening assays was carried out exactly as described in Miller et al. 2020²⁸ (see Chapter 2), with the exception of solvents used to extract

each set of assays as described in the main text. For *ent*-atiserene and *ent*-atiserene-20-al scaleup, three whole plants were infiltrated with a syringe, and approximately 15/30 g of fresh weight were extracted with hexane/ethyl acetate (respectively). Products were purified through silica chromatography with 10% ethyl acetate : 90% hexane as the mobile phase. Initial purification was carried out with approximately 100 mL of oven-dried silica, and fractions were collected in approximately 3 mL increments and assessed for purity by GC-FID. Fractions containing desired products were further purified with approximately 1.5 mL oven-dried silica in a Pasteur pipette, with fractions collected in 1 mL increments and purity assessed by GC-FID. NMR analysis was carried out on a Bruker 800 MHz spectrometer equipped with a TCl cryoprobe using CDCl₃ as the solvent. CDCl₃ peaks were referenced to 7.26 and 77.00 ppm for ¹H and ¹³C spectra, respectively.

GC-MS analysis

All GC-MS analyses were performed on hexane or ethyl acetate extracts (described for each case in the text) with an Agilent 7890A GC with an Agilent VF-5ms column (30 m x 250 µm x 0.25 µm, with 10m EZ-Guard) and an Agilent 5975C mass spectrometer. The inlet was set to 250°C splitless injection of 1 µL, He carrier gas (1 ml/min), and the detector was activated following a 3 min solvent delay. Mass spectra were generated using 70 eV electron ionization with a scan range of m/z 50 to 350. The following method was used for analysis of each sample presented in the text: temperature ramp start 40°C, hold 1 min, 40°C/min to 200°C, hold 2 min, 20°C/min to 280°C, 40°C/min to 320°C; hold 5 min. Figures for chromatograms and mass spectra were generated with Pyplot.

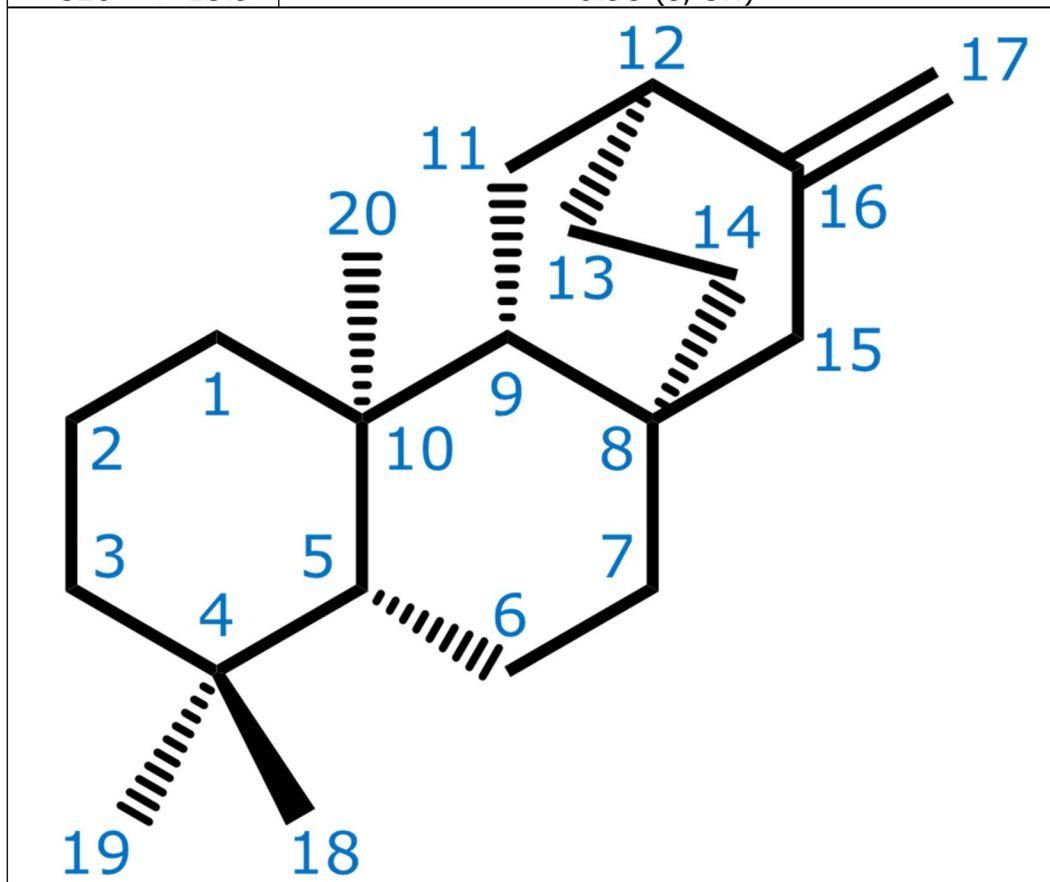
LC-MS analysis

All LC-MS analyses were performed on 80% methanol : 20% H₂O *N. benthamiana* extracts with a Waters Xevo G2-XS quadrupole ToF mass spectrometer with a Waters ACQUITY column manager and Waters ACQUITY BEH C18 column (2.1 x 100 mm; 1.7 µm). Injection volume for each sample was 10 µL, and flow rate was set to 0.3 mL/min with a column temperature of 40°C. The mobile phase consisted of 10 mM ammonium formate (pH 2.8) (Solvent A) and acetonitrile (Solvent B) with the following method: initial 99% A : 1 % B , continuous gradient to 2% A : 98% B over 12 min, hold for 1.5 min, continuous gradient to 99% A : 1% B over 0.1 min, hold 1.5 min. Mass spectra were generated through electrospray ionization in positive-ion mode with leucine enkephalin as a lockmass, and continuum peak acquisition were collected with a mass range of m/z 50-1500 and a scan duration of 0.2 s. Capillary and cone voltage were 3.0 kV and 40 V, respectively, cone and desolvation gas flow rates were 40 and 600 L/h, respectively, and source and desolvation temperatures were 100°C and 350°C, respectively. High-energy spectra were generated with argon as the collision gas and a voltage ramp from 20 to 80 V. Figures for chromatograms and mass spectra were generated with Pyplot.

APPENDIX

Table 3.S1: ^1H and ^{13}C chemical shifts for ent-atiserene. CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively.

^{13}C NMR	δ	^1H NMR
C1	39.4	0.79 (dd, $J = 3.7, 13.1$ Hz, 1H); 1.53 (m, 1H)
C2	18.2	1.38 (m, 1H); 1.59 (m, 1H)
C3	42.2	1.15 (m, 1H); 1.38
C4	33.1	-
C5	56.3	0.82 (dd, $J = 1.7, 12.1$ Hz, 1H)
C6	18.8	1.34 (m, 1H); 1.48 (ddt, $J = 13.3, 4.6, 2.4$ Hz, 1H)
C7	39.5	1.14 (m, 1H); 1.18 (m, 1H)
C8	33.5	-
C9	52.8	1.16 (m, 1H)
C10	37.7	-
C11	28.6	1.42 (m, 1H); 1.59 (m, 1H)
C12	36.6	2.24 (p, $J = 3.0$ Hz, 1H)
C13	28.7	0.99 (m, 1H); 1.94 (m, 1H)
C14	27.4	1.59 (m, 1H); 1.62 (m, 1H)
C15	48.3	1.91 (m, 1H); 2.05 (m, 1H)
C16	153.2	-
C17	104.4	4.58 (q, $J = 2.1$ Hz, 1H); 4.74 (q, $J = 2.3$ Hz, 1H)
C18	33.5	0.87 (s, 3H)
C19	21.7	0.84 (s, 3H)
C20	13.9	0.98 (s, 3H)



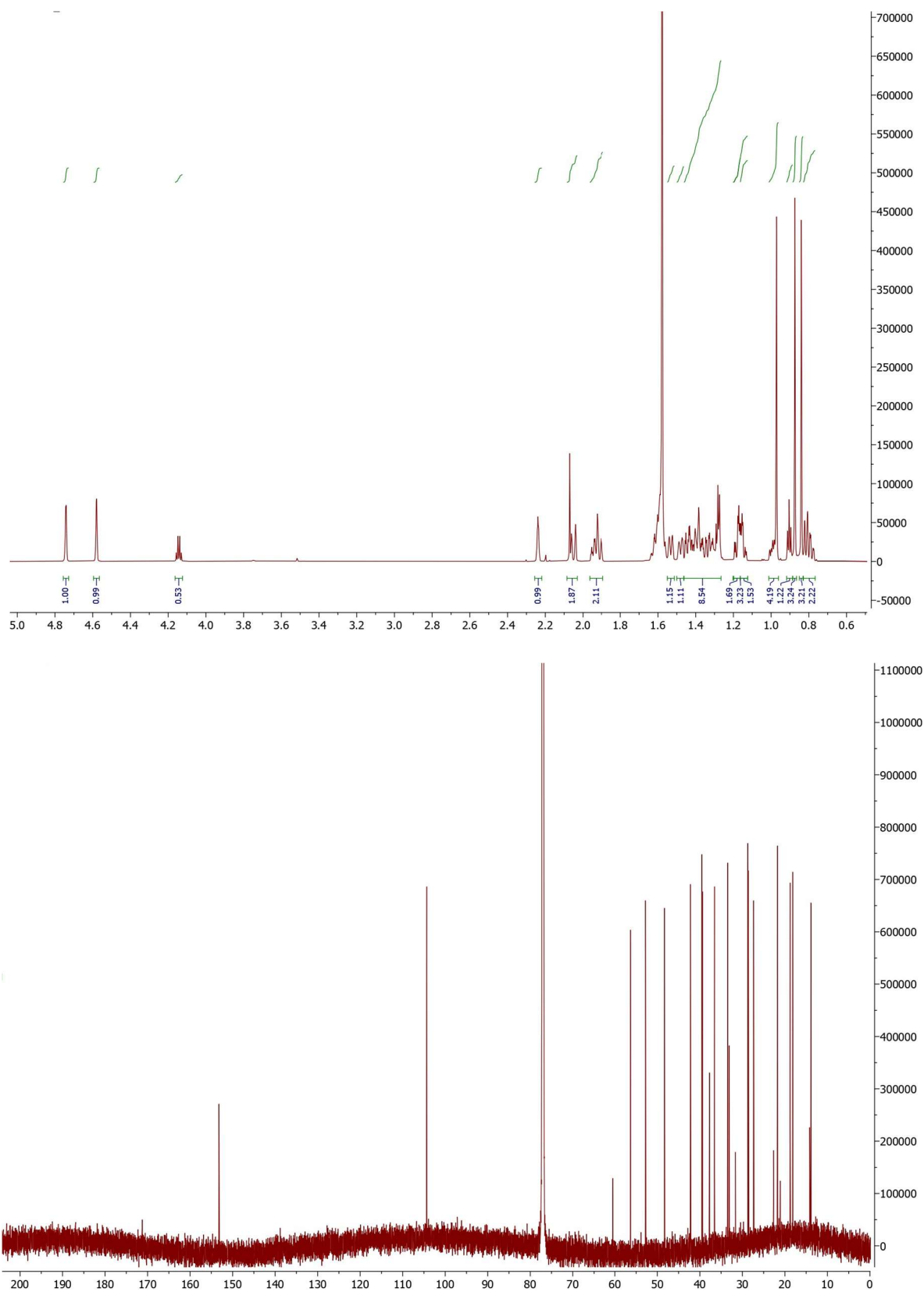


Figure 3.S1: ^1H , ^{13}C , HSQC, H2BC, HMBC, COSY, and NOESY NMR spectra for *ent-atiserene*.

Figure 3.S1 (cont'd)

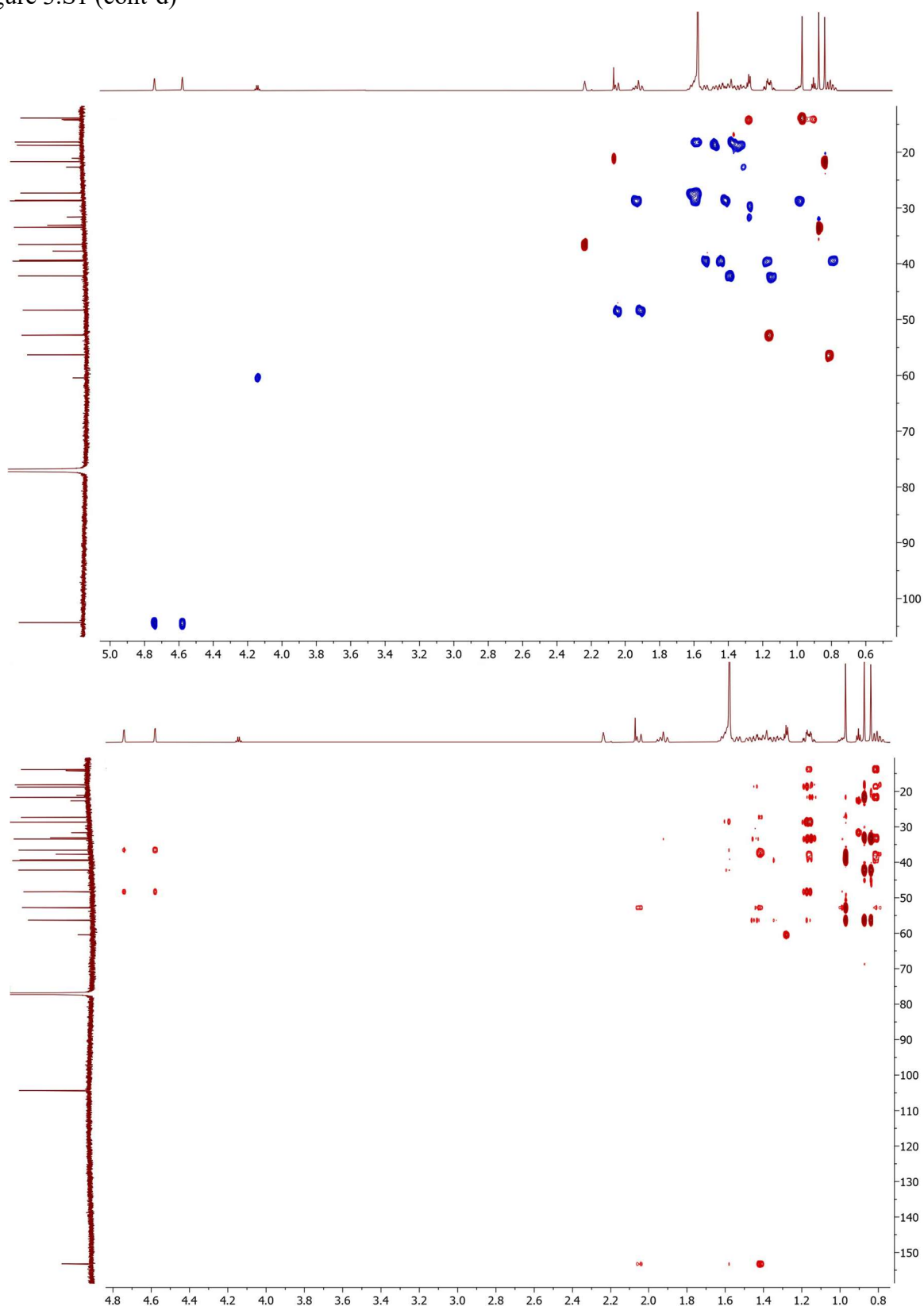


Figure 3.S1 (cont'd)

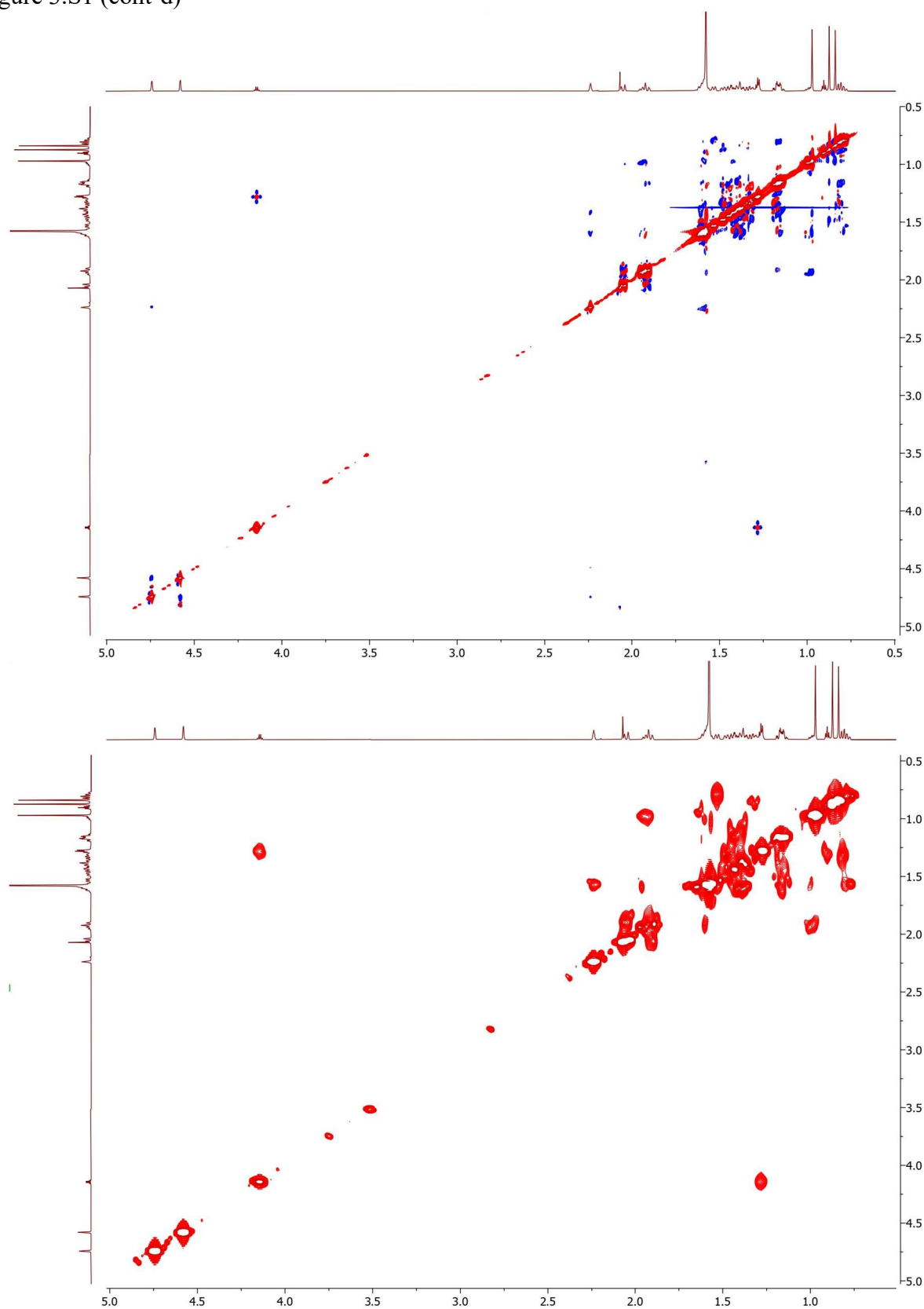




Figure 3.S2: Maximum likelihood phylogenetic tree of candidate P450s from the CYP71 clan. Branch lengths indicate substitutions per site and numbers at nodes represent percent support from 1,000 bootstrap replicates. Candidates in each tree are not clustered by sequence identity and so each assembled transcript is present.

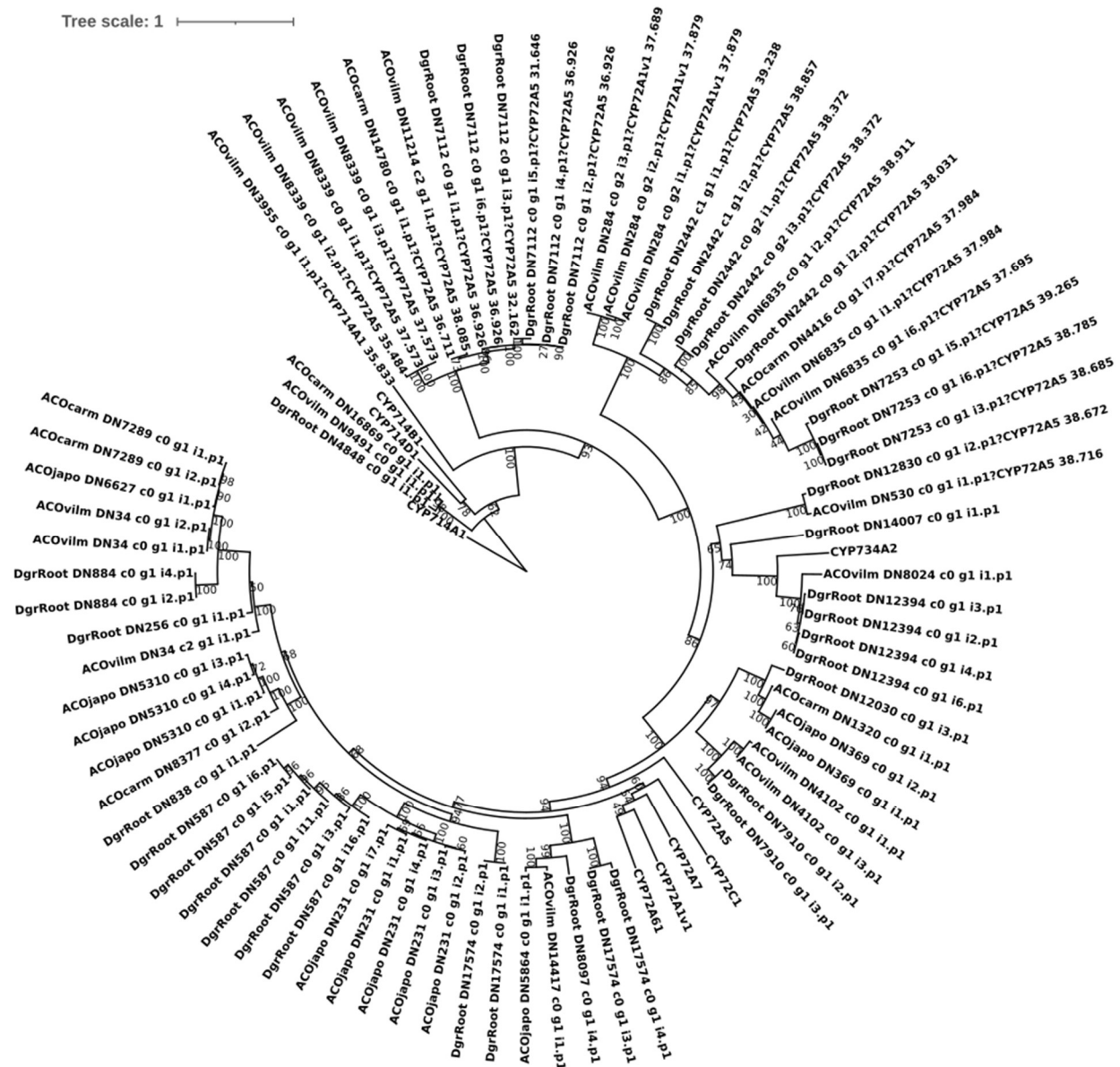
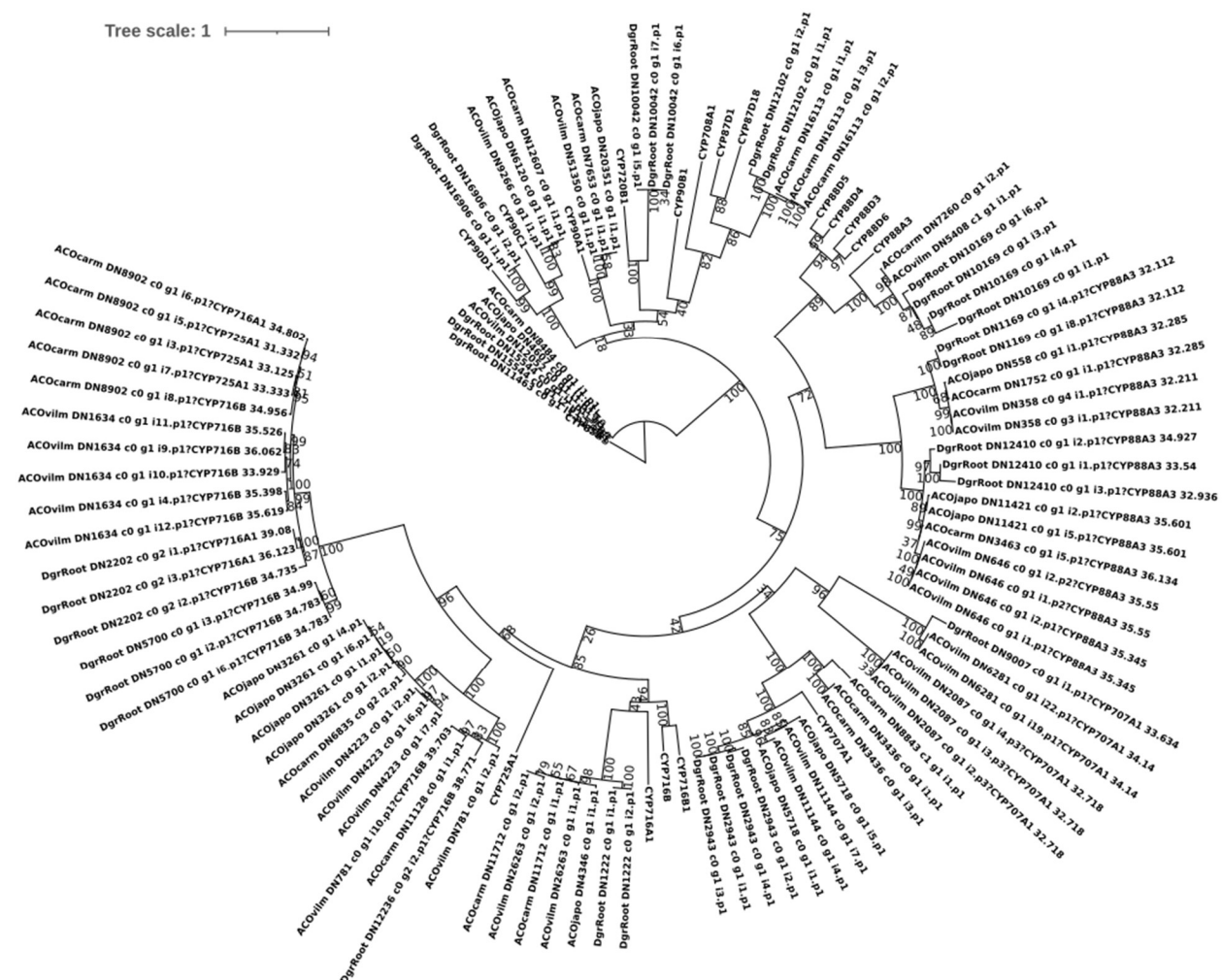


Figure 3.S3: Maximum likelihood phylogenetic tree of candidate P450s from the CYP72 clan. Branch lengths indicate substitutions per site and numbers at nodes represent percent support from 1,000 bootstrap replicates. Candidates in each tree are not clustered by sequence identity and so each assembled transcript is present.



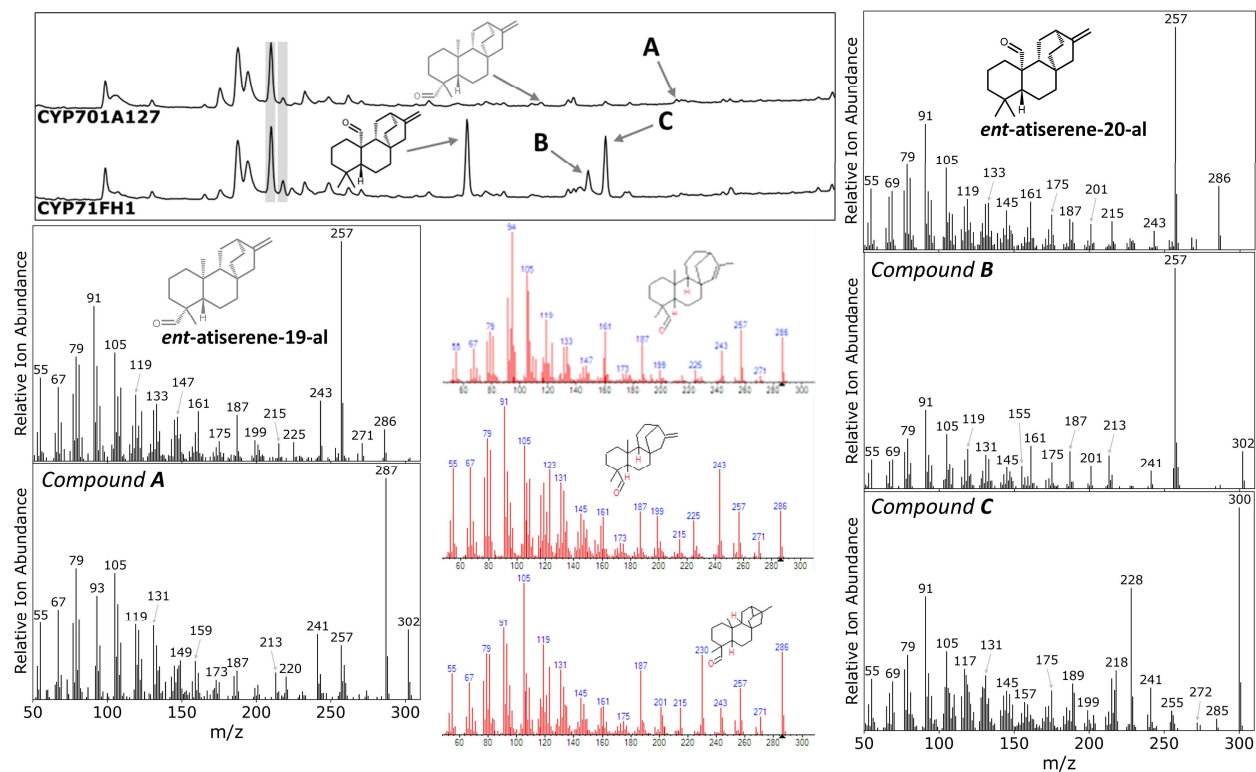


Figure 3.S6: Mass spectra for all compounds shown in Figure 3.6 in the main text. (Top Left) Section of Figure 3.6 relevant to these compounds. (Bottom Left) Mass spectra for compounds made through coexpression of CYP701A127 with previous pathway steps. (Middle) Close matches for *ent-atiserene-19-al* from the NIST database with mass spectra and structures. (Right) Mass spectra for compounds made through coexpression of CYP71FH1 with previous pathway steps.

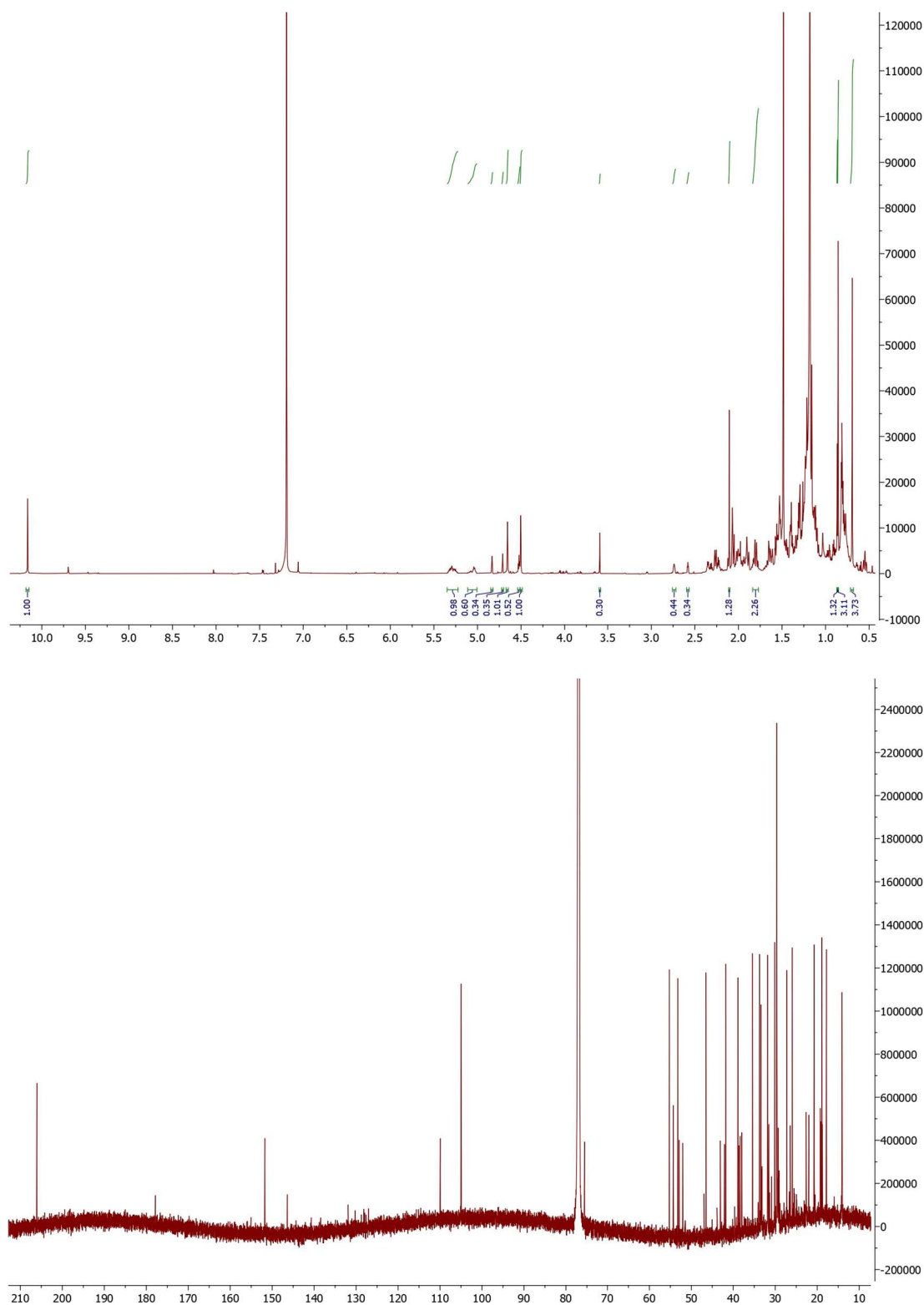


Figure 3.S7: ^1H , ^{13}C , HSQC, H2BC, HMBC, COSY, and NOESY NMR spectra of ent-atiserene-20-al. Aldehyde peak is present in ^1H spectrum at 10.16 ppm, which has the same integration value as terminal alkene protons (4.50 and 4.66 ppm). This product was not completely purified from Compound C (peaks at 4.71 and 4.83 ppm are likely terminal alkene protons for Compound C).

Figure 3.S7 (cont'd)

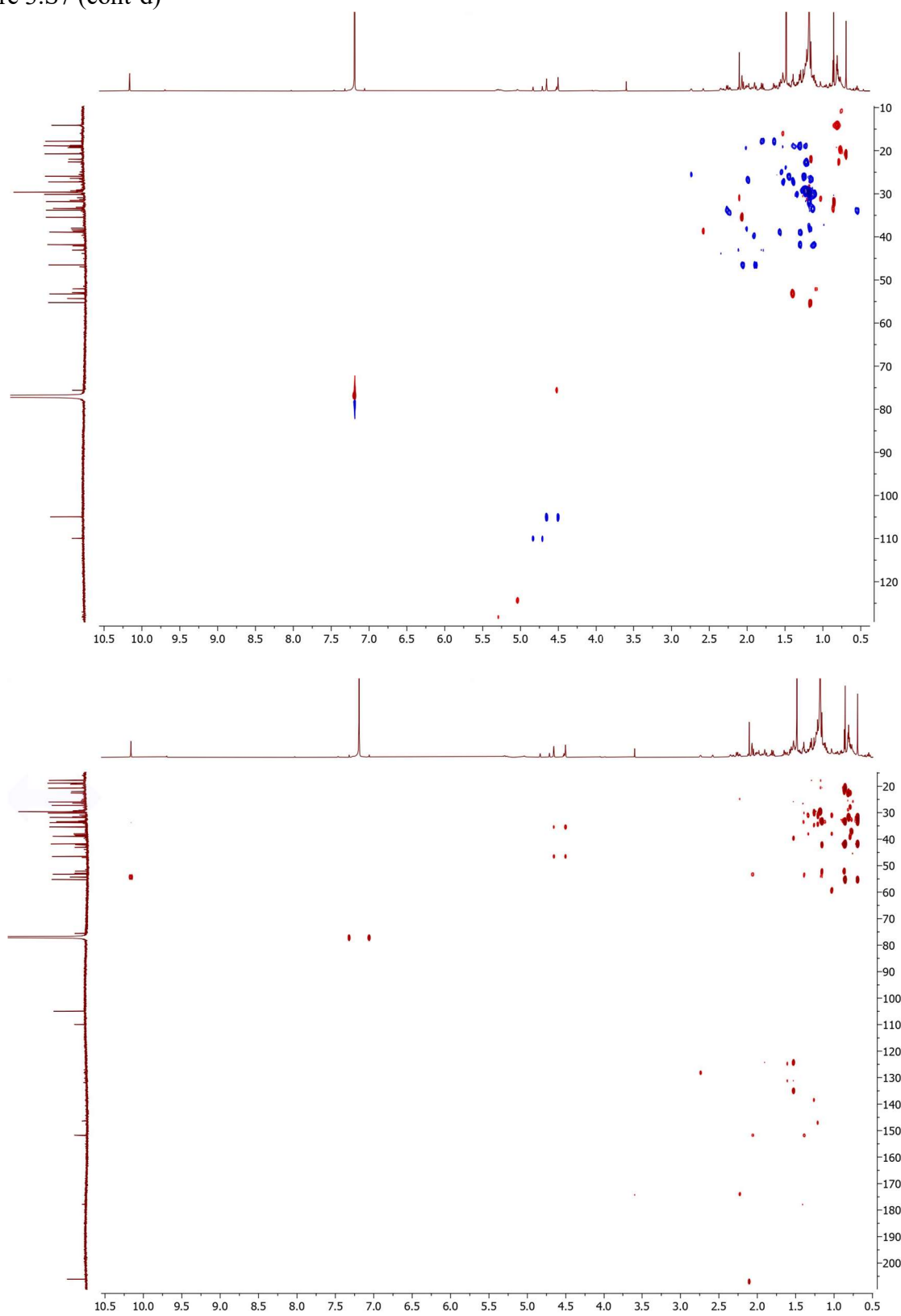
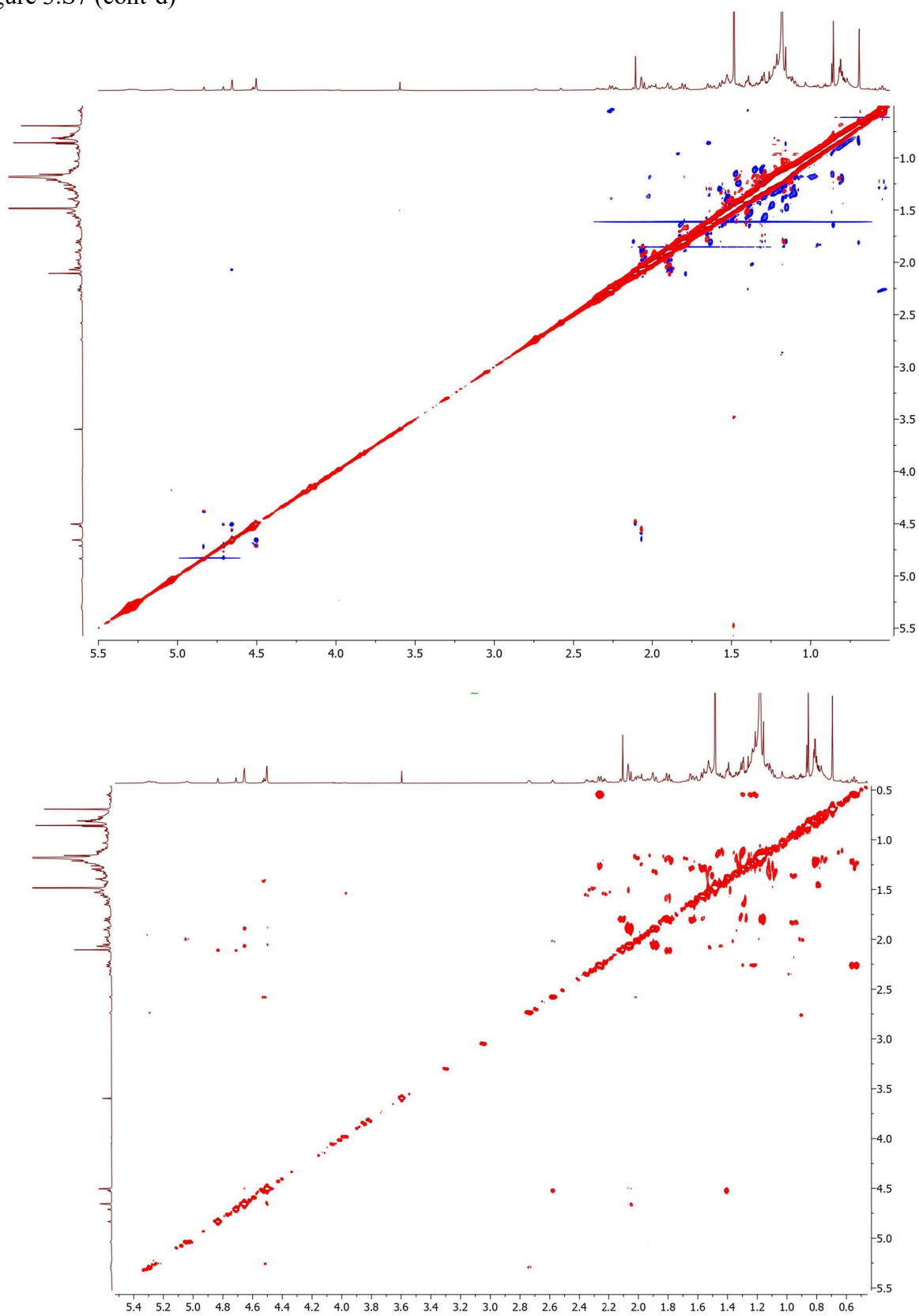


Figure 3.S7 (cont'd)



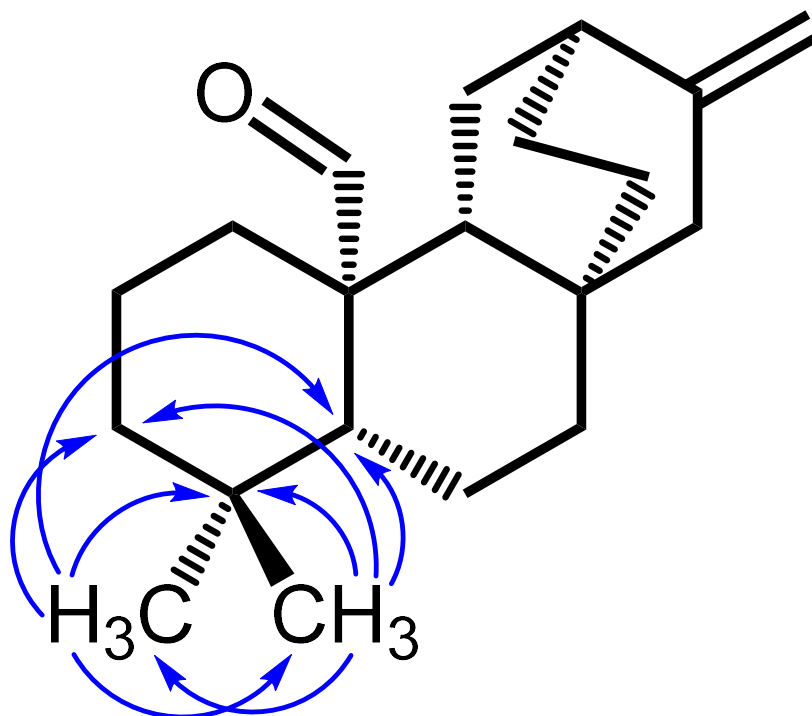


Figure 3.S8: Select HMBC correlations for *ent-atiserene-20-al*. Correlations drawn show methyl groups for carbons 18 and 19 are retained following conversion of *ent-atiserene* by CYP71FH1.

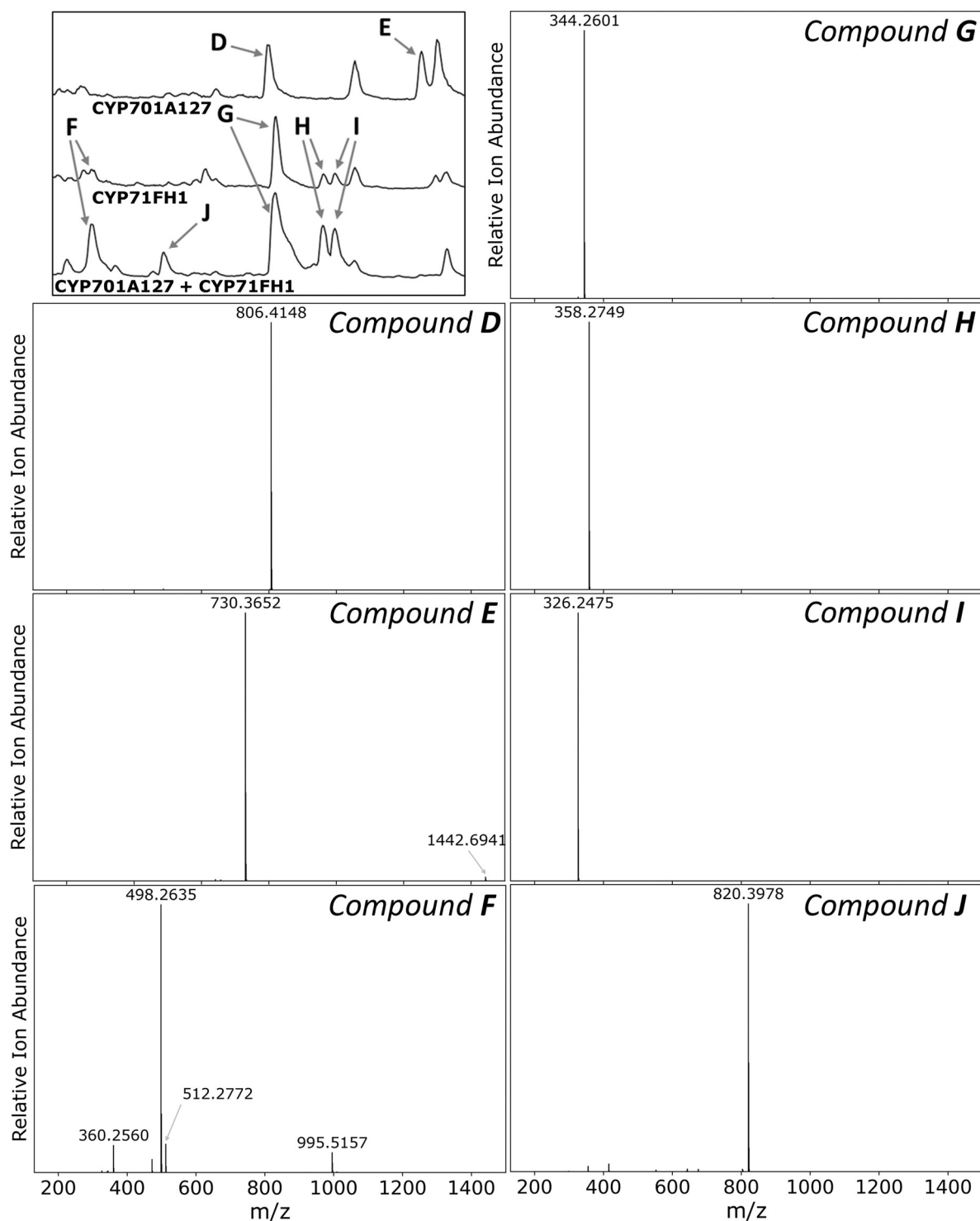


Figure 3.S9: Mass spectra for all compounds shown in Figure 3.7 in the main text. Relevant portion of Figure 3.7 is shown in the top left.

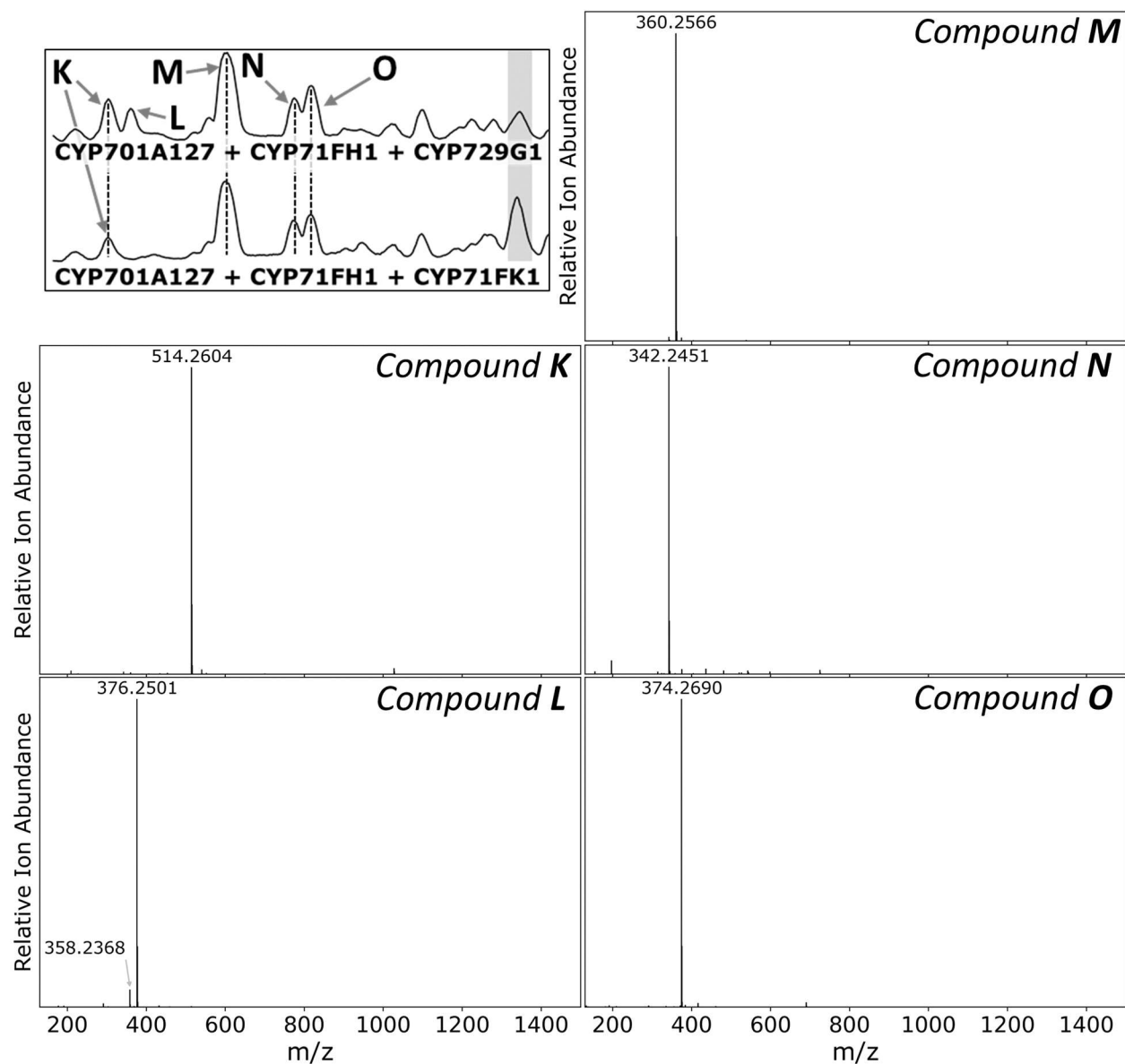


Figure 3.S10: Mass spectra for all compounds shown in Figure 3.8 in the main text. Relevant portion of Figure 3.8 is shown in the top left.

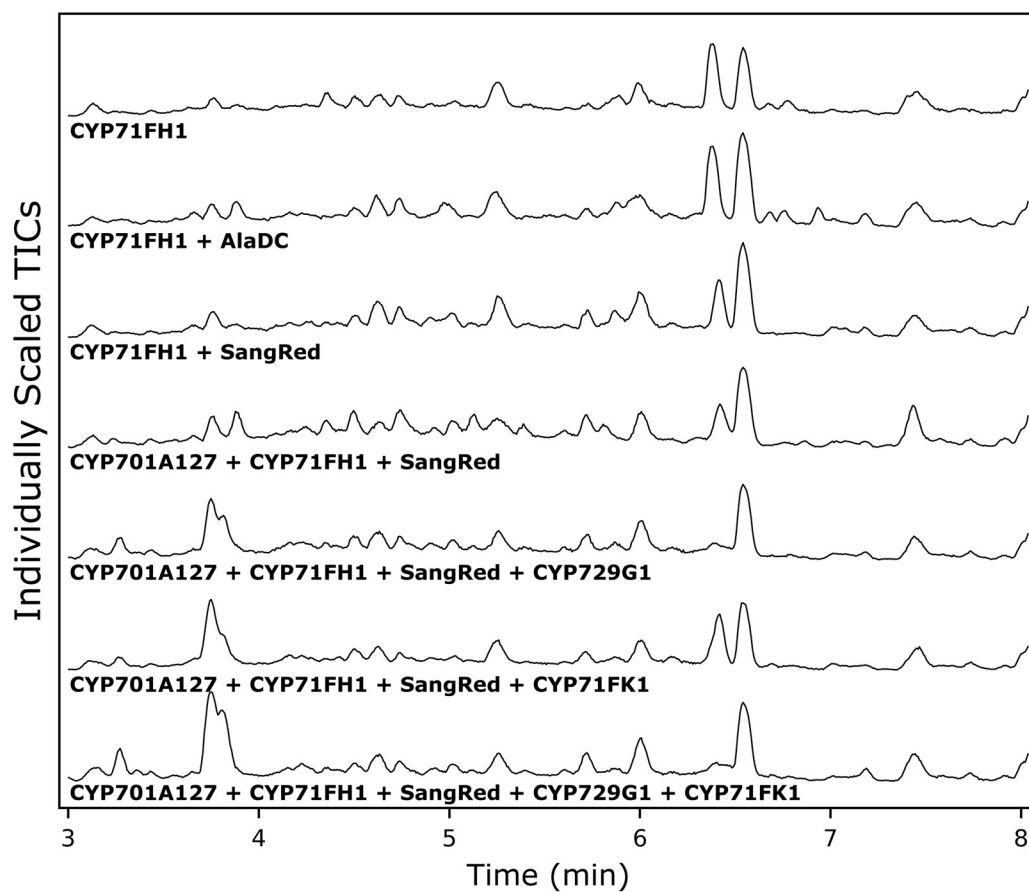


Figure 3.S11: *CYP729G1* and *CYP71FK1* still have similar activity when coexpressed with *SangRed*. Data shown are LC-MS total ion chromatograms. Each assay has *CfDXS*, *CfGGPPS*, *DgrTPS1*, and *DgrTPS7* coexpressed in addition to those listed.

REFERENCES

REFERENCES

- (1) Galanie, S.; Thodey, K.; Trenchard, I. J.; Filsinger Interrante, M.; Smolke, C. D. Complete Biosynthesis of Opioids in Yeast. *Science* **2015**, *349* (6252), 1095–1100. <https://doi.org/10.1126/science.aac9373>.
- (2) Nett, R. S.; Lau, W.; Sattely, E. S. Discovery and Engineering of Colchicine Alkaloid Biosynthesis. *Nature* **2020**, *584* (7819), 148–153. <https://doi.org/10.1038/s41586-020-2546-8>.
- (3) Bedewitz, M. A.; Jones, A. D.; D'Auria, J. C.; Barry, C. S. Tropinone Synthesis via an Atypical Polyketide Synthase and P450-Mediated Cyclization. *Nat Commun* **2018**, *9*, 5281. <https://doi.org/10.1038/s41467-018-07671-3>.
- (4) Wrenbeck, E. E.; Bedewitz, M. A.; Klesmith, J. R.; Noshin, S.; Barry, C. S.; Whitehead, T. A. An Automated Data-Driven Pipeline for Improving Heterologous Enzyme Expression. *ACS Synth. Biol.* **2019**, *8* (3), 474–481. <https://doi.org/10.1021/acssynbio.8b00486>.
- (5) *Biosynthesis of medicinal tropane alkaloids in yeast* | *Nature*. <https://www.nature.com/articles/s41586-020-2650-9> (accessed 2021-04-15).
- (6) Pan, Q.; Mustafa, N. R.; Tang, K.; Choi, Y. H.; Verpoorte, R. Monoterpenoid Indole Alkaloids Biosynthesis and Its Regulation in *Catharanthus Roseus*: A Literature Review from Genes to Metabolites. *Phytochem Rev* **2016**, *15* (2), 221–250. <https://doi.org/10.1007/s11101-015-9406-4>.
- (7) Caputi, L.; Franke, J.; Farrow, S. C.; Chung, K.; Payne, R. M. E.; Nguyen, T.-D.; Dang, T.-T. T.; Soares Teto Carqueijeiro, I.; Koudounas, K.; Dugé de Bernonville, T.; Ameyaw, B.; Jones, D. M.; Vieira, I. J. C.; Courdavault, V.; O'Connor, S. E. Missing Enzymes in the Biosynthesis of the Anticancer Drug Vinblastine in Madagascar Periwinkle. *Science* **2018**, *360* (6394), 1235–1239. <https://doi.org/10.1126/science.aat4100>.
- (8) Qu, Y.; Safonova, O.; De Luca, V. Completion of the Canonical Pathway for Assembly of Anticancer Drugs Vincristine/Vinblastine in *Catharanthus Roseus*. *The Plant Journal* **2019**, *97* (2), 257–266. <https://doi.org/10.1111/tj.14111>.
- (9) Lichman, B. R. The Scaffold-Forming Steps of Plant Alkaloid Biosynthesis. *Nat. Prod. Rep.* **2021**, *38* (1), 103–129. <https://doi.org/10.1039/D0NP00031K>.
- (10) Oneto, J. F. The Alkaloids of Species of *Garrya*. I. Isolation of Alkaloids**University of California, College of Pharmacy, San Francisco. *Journal of the American Pharmaceutical Association (Scientific ed.)* **1946**, *35* (7), 204–207. <https://doi.org/10.1002/jps.3030350703>.
- (11) Ma, Y.; Mao, X.-Y.; Huang, L.-J.; Fan, Y.-M.; Gu, W.; Yan, C.; Huang, T.; Zhang, J.-X.; Yuan, C.-M.; Hao, X.-J. Diterpene Alkaloids and Diterpenes from *Spiraea Japonica* and Their

Anti-Tobacco Mosaic Virus Activity. *Fitoterapia* **2016**, *109*, 8–13. <https://doi.org/10.1016/j.fitote.2015.11.019>.

(12) Hart, N.; Johns, S.; Lamberton, J.; Soares, H.; Willing, R. New Alkaloids of the Ent-Kaurene Type From Anopterus Species (Escalloniaceae). I. The Structure and Reactions of Anopterine. *Aust. J. Chem.* **1976**, *29* (6), 1295–1318. <https://doi.org/10.1071/ch9761295>.

(13) Yin, T.; Cai, L.; Ding, Z. An Overview of the Chemical Constituents from the Genus Delphinium Reported in the Last Four Decades. *RSC Advances* **2020**, *10* (23), 13669–13686. <https://doi.org/10.1039/D0RA00813C>.

(14) Nyirimigabo, E.; Xu, Y.; Li, Y.; Wang, Y.; Agyemang, K.; Zhang, Y. A Review on Phytochemistry, Pharmacology and Toxicology Studies of Aconitum. *J Pharm Pharmacol* **2015**, *67* (1), 1–19. <https://doi.org/10.1111/jphp.12310>.

(15) Csupor, D.; Wenzig, E. M.; Zupkó, I.; Wölkart, K.; Hohmann, J.; Bauer, R. Qualitative and Quantitative Analysis of Aconitine-Type and Lipo-Alkaloids of Aconitum Carmichaelii Roots. *Journal of Chromatography A* **2009**, *1216* (11), 2079–2086. <https://doi.org/10.1016/j.chroma.2008.10.082>.

(16) Zhou, G.; Tang, L.; Zhou, X.; Wang, T.; Kou, Z.; Wang, Z. A Review on Phytochemistry and Pharmacological Activities of the Processed Lateral Root of Aconitum Carmichaelii Debeaux. *J Ethnopharmacol* **2015**, *160*, 173–193. <https://doi.org/10.1016/j.jep.2014.11.043>.

(17) Liu, X.-Y.; Wang, F.-P.; Qin, Y. Synthesis of Three-Dimensionally Fascinating Diterpenoid Alkaloids and Related Diterpenes. *Acc. Chem. Res.* **2021**, *54* (1), 22–34. <https://doi.org/10.1021/acs.accounts.0c00720>.

(18) Gong, J.; Chen, H.; Liu, X.-Y.; Wang, Z.-X.; Nie, W.; Qin, Y. Total Synthesis of Atropurpuran. *Nat Commun* **2016**, *7* (1), 12183. <https://doi.org/10.1038/ncomms12183>.

(19) Owens, K. R.; McCowen, S. V.; Blackford, K. A.; Ueno, S.; Hirooka, Y.; Weber, M.; Sarpong, R. Total Synthesis of the Diterpenoid Alkaloid Arcutinidine Using a Strategy Inspired by Chemical Network Analysis. *J. Am. Chem. Soc.* **2019**, *141* (35), 13713–13717. <https://doi.org/10.1021/jacs.9b05815>.

(20) Pang, L.; Liu, C.-Y.; Gong, G.-H.; Quan, Z.-S. Synthesis, in Vitro and in Vivo Biological Evaluation of Novel Lappaconitine Derivatives as Potential Anti-Inflammatory Agents. *Acta Pharm Sin B* **2020**, *10* (4), 628–645. <https://doi.org/10.1016/j.apsb.2019.09.002>.

(21) Cherney, E. C.; Baran, P. S. Terpenoid-Alkaloids: Their Biosynthetic Twist of Fate and Total Synthesis. *Isr J Chem* **2011**, *51* (3–4), 391–405. <https://doi.org/10.1002/ijch.201100005>.

(22) Li, Y.-G.; Mou, F.-J.; Li, K.-Z. De Novo RNA Sequencing and Analysis Reveal the Putative Genes Involved in Diterpenoid Biosynthesis in Aconitum Vilmorinianum Roots. *3 Biotech* **2021**, *11* (2), 96. <https://doi.org/10.1007/s13205-021-02646-6>.

- (23) Pal, T.; Malhotra, N.; Chanumolu, S. K.; Chauhan, R. S. Next-Generation Sequencing (NGS) Transcriptomes Reveal Association of Multiple Genes and Pathways Contributing to Secondary Metabolites Accumulation in Tuberous Roots of *Aconitum Heterophyllum* Wall. *Planta* **2015**, *242* (1), 239–258. <https://doi.org/10.1007/s00425-015-2304-6>.
- (24) Rai, M.; Rai, A.; Kawano, N.; Yoshimatsu, K.; Takahashi, H.; Suzuki, H.; Kawahara, N.; Saito, K.; Yamazaki, M. De Novo RNA Sequencing and Expression Analysis of *Aconitum Carmichaelii* to Analyze Key Genes Involved in the Biosynthesis of Diterpene Alkaloids. *Molecules* **2017**, *22* (12). <https://doi.org/10.3390/molecules22122155>.
- (25) Yang, Y.; Hu, P.; Zhou, X.; Wu, P.; Si, X.; Lu, B.; Zhu, Y.; Xia, Y. Transcriptome Analysis of *Aconitum Carmichaelii* and Exploration of the Salsolinol Biosynthetic Pathway. *Fitoterapia* **2020**, *140*, 104412. <https://doi.org/10.1016/j.fitote.2019.104412>.
- (26) Zhao, D.; Shen, Y.; Shi, Y.; Shi, X.; Qiao, Q.; Zi, S.; Zhao, E.; Yu, D.; Kennelly, E. J. Probing the Transcriptome of *Aconitum Carmichaelii* Reveals the Candidate Genes Associated with the Biosynthesis of the Toxic Aconitine-Type C19-Diterpenoid Alkaloids. *Phytochemistry* **2018**, *152*, 113–124. <https://doi.org/10.1016/j.phytochem.2018.04.022>.
- (27) Mao, L.; Jin, B.; Chen, L.; Tian, M.; Ma, R.; Yin, B.; Zhang, H.; Guo, J.; Tang, J.; Chen, T.; Lai, C.; Cui, G.; Huang, L. Functional Identification of the Terpene Synthase Family Involved in Diterpenoid Alkaloids Biosynthesis in *Aconitum Carmichaelii*. *Acta Pharmaceutica Sinica B* **2021**. <https://doi.org/10.1016/j.apsb.2021.04.008>.
- (28) Miller, G. P.; Bhat, W. W.; Lanier, E. R.; Johnson, S. R.; Mathieu, D. T.; Hamberger, B. The Biosynthesis of the Anti-Microbial Diterpenoid Leubethanol in *Leucophyllum Frutescens* Proceeds via an All-Cis Prenyl Intermediate. *The Plant Journal* **2020**, *104* (3), 693–705. <https://doi.org/10.1111/tpj.14957>.
- (29) Jin, B.; Cui, G.; Guo, J.; Tang, J.; Duan, L.; Lin, H.; Shen, Y.; Chen, T.; Zhang, H.; Huang, L. Functional Diversification of Kaurene Synthase-Like Genes in *Isodon Rubescens*. *Plant Physiology* **2017**, *174* (2), 943–955. <https://doi.org/10.1104/pp.17.00202>.
- (30) Grennan, A. K. Gibberellin Metabolism Enzymes in Rice. *Plant Physiology* **2006**, *141* (2), 524–526. <https://doi.org/10.1104/pp.104.900192>.
- (31) Kong, H.; Zhang, Y.; Hong, Y.; Barker, M. S. Multilocus Phylogenetic Reconstruction Informing Polyploid Relationships of *Aconitum* Subgenus *Lycotemon* (Ranunculaceae) in China. *Plant Syst Evol* **2017**, *303* (6), 727–744. <https://doi.org/10.1007/s00606-017-1406-y>.
- (32) Park, S.; An, B.; Park, S. Recurrent Gene Duplication in the Angiosperm Tribe Delphinieae (Ranunculaceae) Inferred from Intracellular Gene Transfer Events and Heteroplasmic Mutations in the Plastid MatK Gene. *Sci Rep* **2020**, *10* (1), 2720. <https://doi.org/10.1038/s41598-020-59547-6>.

- (33) Salvado, P.; Aymerich Boixader, P.; Parera, J.; Vila Bonfill, A.; Martin, M.; Quélenec, C.; Lewin, J.-M.; Delorme-Hinoux, V.; Bertrand, J. A. M. Little Hope for the Polyploid Endemic Pyrenean Larkspur (*Delphinium Montanum*): Evidences from Population Genomics and Ecological Niche Modeling. *Ecology and Evolution* **2022**, *12* (3), e8711. <https://doi.org/10.1002/ece3.8711>.
- (34) Xu, J.-B.; Li, Y.-Z.; Huang, S.; Chen, L.; Luo, Y.-Y.; Gao, F.; Zhou, X.-L. Diterpenoid Alkaloids from the Whole Herb of *Delphinium Grandiflorum* L. *Phytochemistry* **2021**, *190*, 112866. <https://doi.org/10.1016/j.phytochem.2021.112866>.
- (35) Li, Y.; Gao, F.; Zhang, J.-F.; Zhou, X.-L. Four New Diterpenoid Alkaloids from the Roots of *Aconitum Carmichaelii*. *Chem. Biodivers.* **2018**, *15* (7), e1800147. <https://doi.org/10.1002/cbdv.201800147>.
- (36) Yamashita, H.; Takeda, K.; Haraguchi, M.; Abe, Y.; Kuwahara, N.; Suzuki, S.; Terui, A.; Masaka, T.; Munakata, N.; Uchida, M.; Nunokawa, M.; Kaneda, K.; Goto, M.; Lee, K.-H.; Wada, K. Four New Diterpenoid Alkaloids from *Aconitum Japonicum* Subsp. *Subcuneatum*. *J Nat Med* **2018**, *72* (1), 230–237. <https://doi.org/10.1007/s11418-017-1139-9>.
- (37) Yin, T.-P.; Cai, L.; Fang, H.-X.; Fang, Y.-S.; Li, Z.-J.; Ding, Z.-T. Diterpenoid Alkaloids from *Aconitum Vilmorinianum*. *Phytochemistry* **2015**, *116*, 314–319. <https://doi.org/10.1016/j.phytochem.2015.05.002>.
- (38) Andersen-Ranberg, J.; Kongstad, K. T.; Nielsen, M. T.; Jensen, N. B.; Pateraki, I.; Bach, S. S.; Hamberger, B.; Zerbe, P.; Staerk, D.; Bohlmann, J.; Møller, B. L.; Hamberger, B. Expanding the Landscape of Diterpene Structural Diversity through Stereochemically Controlled Combinatorial Biosynthesis. *Angewandte Chemie International Edition* **2016**, *55* (6), 2142–2146. <https://doi.org/10.1002/anie.201510650>.
- (39) Nelson, D.; Werck-Reichhart, D. A P450-Centric View of Plant Evolution. *The Plant Journal* **2011**, *66* (1), 194–211. <https://doi.org/10.1111/j.1365-313X.2011.04529.x>.
- (40) Pateraki, I.; Andersen-Ranberg, J.; Jensen, N. B.; Wubshet, S. G.; Heskes, A. M.; Forman, V.; Hallström, B.; Hamberger, B.; Motawia, M. S.; Olsen, C. E.; Staerk, D.; Hansen, J.; Møller, B. L.; Hamberger, B. Total Biosynthesis of the Cyclic AMP Booster Forskolin from *Coleus Forskohlii*. *eLife* **2017**, *6*, e23001. <https://doi.org/10.7554/eLife.23001>.
- (41) Bai, P.; Wang, L.; Wei, K.; Ruan, L.; Wu, L.; He, M.; Ni, D.; Cheng, H. Biochemical Characterization of Specific Alanine Decarboxylase (AlaDC) and Its Ancestral Enzyme Serine Decarboxylase (SDC) in Tea Plants (*Camellia Sinensis*). *BMC Biotechnology* **2021**, *21* (1), 17. <https://doi.org/10.1186/s12896-021-00674-x>.
- (42) Zhao, P.-J.; Gao, S.; Fan, L.-M.; Nie, J.-L.; He, H.-P.; Zeng, Y.; Shen, Y.-M.; Hao, X.-J. Approach to the Biosynthesis of Atisine-Type Diterpenoid Alkaloids. *J. Nat. Prod.* **2009**, *72* (4), 645–649. <https://doi.org/10.1021/np800657j>.

- (43) Wisecaver, J. H.; Borowsky, A. T.; Tzin, V.; Jander, G.; Kliebenstein, D. J.; Rokas, A. A Global Coexpression Network Approach for Connecting Genes to Specialized Metabolic Pathways in Plants. *Plant Cell* **2017**, *29* (5), 944–959. <https://doi.org/10.1105/tpc.17.00009>.
- (44) Bosch i Daniel, M.; Simon Pallisé, J.; López i Pujol, J.; Blanché i Vergés, C. DCDB: An Updated on-Line Database of Chromosome Numbers of Tribe Delphinieae (Ranunculaceae). **2016**.
- (45) Morrone, D.; Chen, X.; Coates, R. M.; Peters, R. J. Characterization of the Kaurene Oxidase CYP701A3, a Multifunctional Cytochrome P450 from Gibberellin Biosynthesis. *Biochemical Journal* **2010**, *431* (3), 337–347. <https://doi.org/10.1042/BJ20100597>.
- (46) Prisic, S.; Xu, M.; Wilderman, P. R.; Peters, R. J. Rice Contains Two Disparate Ent-Copalyl Diphosphate Synthases with Distinct Metabolic Functions. *Plant Physiol* **2004**, *136* (4), 4228–4236. <https://doi.org/10.1104/pp.104.050567>.
- (47) Harris, L. J.; Saparno, A.; Johnston, A.; Prisic, S.; Xu, M.; Allard, S.; Kathiresan, A.; Ouellet, T.; Peters, R. J. The Maize An2 Gene Is Induced by Fusarium Attack and Encodes an Ent-Copalyl Diphosphate Synthase. *Plant Mol Biol* **2005**, *59* (6), 881–894. <https://doi.org/10.1007/s11103-005-1674-8>.
- (48) Kumar, S.; Stecher, G.; Suleski, M.; Hedges, S. B. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* **2017**, *34* (7), 1812–1819. <https://doi.org/10.1093/molbev/msx116>.
- (49) Minami, H.; Dubouzet, E.; Iwasa, K.; Sato, F. Functional Analysis of Norcoclaurine Synthase in *Coptis Japonica*. *J Biol Chem* **2007**, *282* (9), 6274–6282. <https://doi.org/10.1074/jbc.M608933200>.
- (50) Li, W.; Godzik, A. Cd-Hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* **2006**, *22* (13), 1658–1659. <https://doi.org/10.1093/bioinformatics/btl158>.
- (51) Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data. *Bioinformatics* **2012**, *28* (23), 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>.
- (52) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13* (11), 2498–2504. <https://doi.org/10.1101/gr.1239303>.
- (53) Johnson, S. R.; Bhat, W. W.; Bibik, J.; Turmo, A.; Hamberger, B.; Hamberger, B. A Database-Driven Approach Identifies Additional Diterpene Synthase Activities in the Mint Family (Lamiaceae). *J Biol Chem* **2019**, *294* (4), 1349–1362. <https://doi.org/10.1074/jbc.RA118.006025>.

CHAPTER 4

Repurposing Terpene Synthases for the Conversion of Synthetic Geranylgeranyl Diphosphate Derivatives

Abstract

Terpenoids are an incredibly diverse class of specialized metabolites with tens of thousands of compounds identified in plants, and their biosynthesis stems from only a small handful of starting substrates. Nearly all diterpenoids are derived from a single substrate—geranylgeranyl diphosphate (GGPP)—which is cyclized to hundreds of diterpene backbones through either single-step or multi-step conversion by terpene synthases (TPSs). Many of these enzymes have been demonstrated to exhibit a high degree of substrate promiscuity in the conversion of substrates which are found throughout the plant kingdom but not necessarily in the TPS's species of origin. Here we extend this to demonstrate that diterpene synthases (diTPSs) can convert synthetic GGPP derivatives which are not found in nature into new terpene backbones. Screening of twenty synthetic substrates against twenty-four enzymes in 438 unique combinations resulted in fifty-four successful combinations and the identification of fifty-six new compounds. Structural characterization by NMR for six select compounds was carried out, showing that many of these products are a direct translation of the synthetic substrate into a respective derivative of the enzymes' native products. This screening of substrate tolerance revealed trends to inform future design principles for TPS biocatalysis, and we demonstrate a high degree of promiscuity in several enzymes with native functions involved in the biosynthesis of medicinal natural products. This work demonstrates the remarkable potential of TPSs as biocatalysts in the semi-biosynthesis of non-natural compounds and methods to access derivatives of natural terpene products which would be difficult to achieve strictly through chemical synthesis.

Introduction

Terpenoids are the largest group of specialized metabolites in plants with tens of thousands of known compounds¹ and a vast array of practical uses such as flavors, fragrances, pesticides, and pharmaceuticals. Despite the enormous number and the diversity of structures present throughout this class of metabolites, their biosynthesis stems from only a small handful of starting substrates. This begins with the conversion of prenyl diphosphate substrates to terpene scaffolds through carbocation-mediated cyclization reactions by terpene synthases (TPSs)². Diterpenoids—twenty carbon terpenoids typically derived from geranylgeranyl diphosphate (GGPP)—exhibit an especially broad range of terpene scaffold structures due to the modular nature of their biosynthesis³. While monoterpenes (C₁₀) and sesquiterpenes (C₁₅) are typically limited to single step conversion by a TPS, diterpene scaffolds are typically synthesized in two steps: first by a class II diterpene synthase (diTPS) which cyclizes GGPP to a common decaline (“labdane”) core, followed by a class I diTPS which further modifies this labdane intermediate to a finalized diterpene scaffold⁴. Diterpenes can also be synthesized in a single step by cyclization of GGPP directly by a class I diTPS², similar to the origin of most mono- and sesquiterpenes, allowing for an even greater diversity of structures.

The combinatorial nature of diterpene biosynthesis has led to a sizable amount of research into the catalytic promiscuity of class I diTPSs which convert labdane-type intermediates^{3,5–7} (termed here as “second step” class I diTPSs to distinguish them from the “single-step” diTPSs which directly convert GGPP). Some previous studies have tested the promiscuity of class II diTPSs⁸ and single-step class I diTPSs⁹, however research in this area lags behind by comparison. While all of the above examples include testing TPSs with substrates which are not necessarily found in the enzyme’s species of origin, each was carried out with substrates found in nature.

Previous research has also been carried out in the conversion of non-natural terpene precursors, although it is limited to single-step class I and mostly sesquiterpene biosynthesis, with examples such as the conversion of hydroxylated^{10–12} and methylated^{12–14} substrates, those with heteroatom insertions¹⁵, and close structural isomers^{16,17}.

Here we take advantage of the modular nature of combinatorial diterpene biosynthesis—and include single-step class I diTPSs as well—to generate a range of new products by substituting the starting substrate (GGPP) with twenty synthetic GGPP derivatives. These substrates were intentionally designed to probe the chemical space that these TPSs are capable of converting—not solely to guarantee a maximal number of successful combinations. Modifications to the native GGPP substrate ranged from small changes such as the addition of a methyl (**1**) or fluoro (**3**) group to those which introduced a larger structural change such as the addition of a phenyl ring (**12–14**) or ethoxy group (**20**). Modifications were also made along the length of the substrate in the example of substrates **8–10**, which respectively introduce an ether between each isoprenyl subunit. In total, 436 enzyme(s) and synthetic substrate combinations resulted in 56 products detectable by GC-MS. This work demonstrates the potential of TPSs as biocatalysts in the conversion of non-native substrates and highlights trends seen both at the substrate and enzyme level for combinations which worked most effectively. We show that many combinations lead to a direct translation of an enzyme's native mechanism with the modification of the substrate carried through to a final product, suggesting that this can be an effective method for the semi-biosynthetic derivatization of natural products which may be difficult to access through chemical synthesis alone.

Results and Discussion

TPSs Utilized in this Study and Screening Process

TPSs chosen for screening are summarized in Table 4.1. Screening-scale assays were carried out *in vitro* with purified enzymes and a two-phase system with a hexane overlay to continuously extract dephosphorylated products. Screening of each combination of class II and single-step class I TPS with substrates **1-20** was first analyzed through GC-FID (Figures 4.S11 and 4.S14), with potential hits rescreened for validation by GC-MS (Figures 4.S8-10). Class II/class I combinations screens were carried out based on select class II/substrate combinations and were also screened initially by GC-FID (Figures 4.S12 and 4.S13) followed by GC-MS (Figures 4.S7 and 4.S10). A summary of all active combinations is given in Figure 4.1, and mass spectra for each product (**21-74**) is given in Figures 4.S10.

Screening of Class II diTPSs

Eleven class II diTPSs which make nine different products with GGPP were screened against all twenty synthetic substrates, demonstrating that class II TPSs can, in fact, convert synthetic GGPP derivatives. In particular, substrates **3** and **8** were converted by the greatest number of enzymes (seven and six, respectively). Conversion of **3** is perhaps unsurprising since a simple hydrogen to fluorine substitution does little to change the structure of the substrate, and the location of this substitution does not interfere with the cyclization mechanism of a class II TPS. While the latter holds true for **8**, the former does not, as the head-to-tail length of the substrate is increased by the addition of an ether.

Table 4.1: List of enzymes used in this study. References written in substrates column show studies which demonstrated the listed substrate promiscuity. Native products are not written for second-step class I TPSs as they vary based on substrate.

	Enzyme	Substrate(s)	Native Product	Species of Origin	Characterized By
Class II	ZmAN2	GGPP	ent-copalyl diphosphate	<i>Zea mays</i>	Harris <i>et al.</i> 2005 ¹⁸
	OsSCS	GGPP	syn-copalyl diphosphate	<i>Oryza sativa</i>	Xu <i>et al.</i> 2004 ¹⁹
	TwTPS21	GGPP	ent-copal-8-ol diphosphate	<i>Trypterygium wilfordii</i>	Hansen <i>et al.</i> 2016 ²⁰
	PcTPS1	GGPP	(10 <i>R</i>)-labda-8,13E-dienyl diphosphate	<i>Pogostemon cablin</i>	Johnson & Bhat <i>et al.</i> 2019 ⁵
	ShTPS1	GGPP	kolavenyl diphosphate	<i>Salvia hispanica</i>	unpublished
	CamTPS2	GGPP	kolavenyl diphosphate	<i>Callicarpa americana</i>	Hamilton <i>et al.</i> 2020 ²¹
	ArTPS2	GGPP	neo-cleroda-4(18),13E-dienyl diphosphate	<i>Ajuga reptans</i>	Johnson & Bhat <i>et al.</i> 2019 ⁵
	CfTPS1	GGPP	(+)-copalyl diphosphate	<i>Coleus forskohlii</i>	Pateraki <i>et al.</i> 2014 ²²
	CfTPS16	GGPP	(+)-copalyl diphosphate	<i>Coleus forskohlii</i>	Johnson & Bhat <i>et al.</i> 2019 ⁵
	CfTPS2	GGPP	(+)-copal-8-ol diphosphate	<i>Coleus forskohlii</i>	Pateraki <i>et al.</i> 2014 ²²
	LlTPS1	GGPP	peregrinol diphosphate	<i>Leonotis leonurus</i>	Johnson & Bhat <i>et al.</i> 2019 ⁵
Class I	DgTPS1	GGPP	casbene	<i>Daphne genkwa</i>	unpublished
	ElCAS	GGPP	casbene	<i>Euphorbia lathyris</i>	Luo <i>et al.</i> 2016 ²³
	EpoiTPS1	GGPP	cembrenol	<i>Euphorbia poissonii</i>	unpublished
	LfTPS1	NNPP	dihydroserrulatene	<i>Leucophyllum frutescens</i>	Miller <i>et al.</i> 2020 ²⁴
	EsTPS1	NNPP	dihydroserrulatene	<i>Eremophila serrulata</i>	Miller <i>et al.</i> 2020 ²⁴
	PvTPS4	FPP	δ cadinene	<i>Prunella vulgaris</i>	Johnson <i>et al.</i> 2019 ⁹
	PvHVS	GGPP	hydroxy vulgarisane	<i>Prunella vulgaris</i>	Johnson <i>et al.</i> 2019 ⁹
	EpTPS1	ent labdanes ³	-	<i>Euphorbia peplus</i>	Zerbe & Hamberger <i>et al.</i> 2013 ²⁵
	LlTPS4a	(+) labdanes ⁵	-	<i>Leonotis leonurus</i>	Johnson & Bhat <i>et al.</i> 2019 ⁵
	CfTPS3	most labdanes ³	-	<i>Coleus forskohlii</i>	Pateraki <i>et al.</i> 2014 ²²
	OmTPS3	(+) labdanes ⁵	-	<i>Origanum majorana</i>	Johnson & Bhat <i>et al.</i> 2019 ⁵
	SsSS	all labdanes ^{3,5,6}	-	<i>Salvia sclarea</i>	Caniard <i>et al.</i> 2012 ²⁶
	KgTS	all labdanes ⁶	-	<i>Kitasatospora griseola</i>	Dairi <i>et al.</i> 2001 ²⁷

The contrast between the conversion of **8** and lack of conversion for substrates **9** and **10** highlight that the cyclization mechanism of a class II diTPS likely has little tolerance for GGPP modifications in the first three isoprenyl subunits distal to the diphosphate. Despite only differing by the location of an identical modification, the ether groups present in **9** and **10** are both within the range of carbons which form the bicyclic labdane core. In fact, no substrates which involved a modification in the first three isoprenyl subunits were shown to be converted by any class II enzyme.

Conversion of both **1** and **3** led to products which are direct translations of the substrate's modification to the diTPS's native product, based on product profile and similar mass spectra as highlighted in Figure 4.2. For example, the native product of *Cf*TPS2 is (+)-copal-8-ol diphosphate²², and two major products detectable by GC-MS following dephosphorylation are two stereoisomers of manoyl oxide. The two major products detected for conversion of **1** and **3** by *Cf*TPS2 are likely stereoisomers of each other based on their mass spectra, and have major peaks shifted relative to manoyl oxide by respective amounts corresponding to their modification (+14 for **1**, +18 for **3**). Similar examples for other class II diTPSs can be seen in Figure 4.S9. While not as obvious in terms of products made, the unusual substrates **17** and **18** were also converted to products by four and three class II enzymes, respectively, showing that these enzymes can convert a variety of synthetic GGPP derivatives that vary by more than a singular small modification.

Perhaps most interesting is the conversion of **4**, **5**, and **8** to **69** across four different enzymes (Figure 4.S9), which has a retention time and mass spectrum consistent with that of a sesquiterpene (m/z 204; Figure 4.S10). This could plausibly arise from cyclization of these substrates into a labdane-type product, followed by cleavage of the diphosphate and subsequent loss of the entire modified group, as these substrates all contain an ether in the same location and vary only by

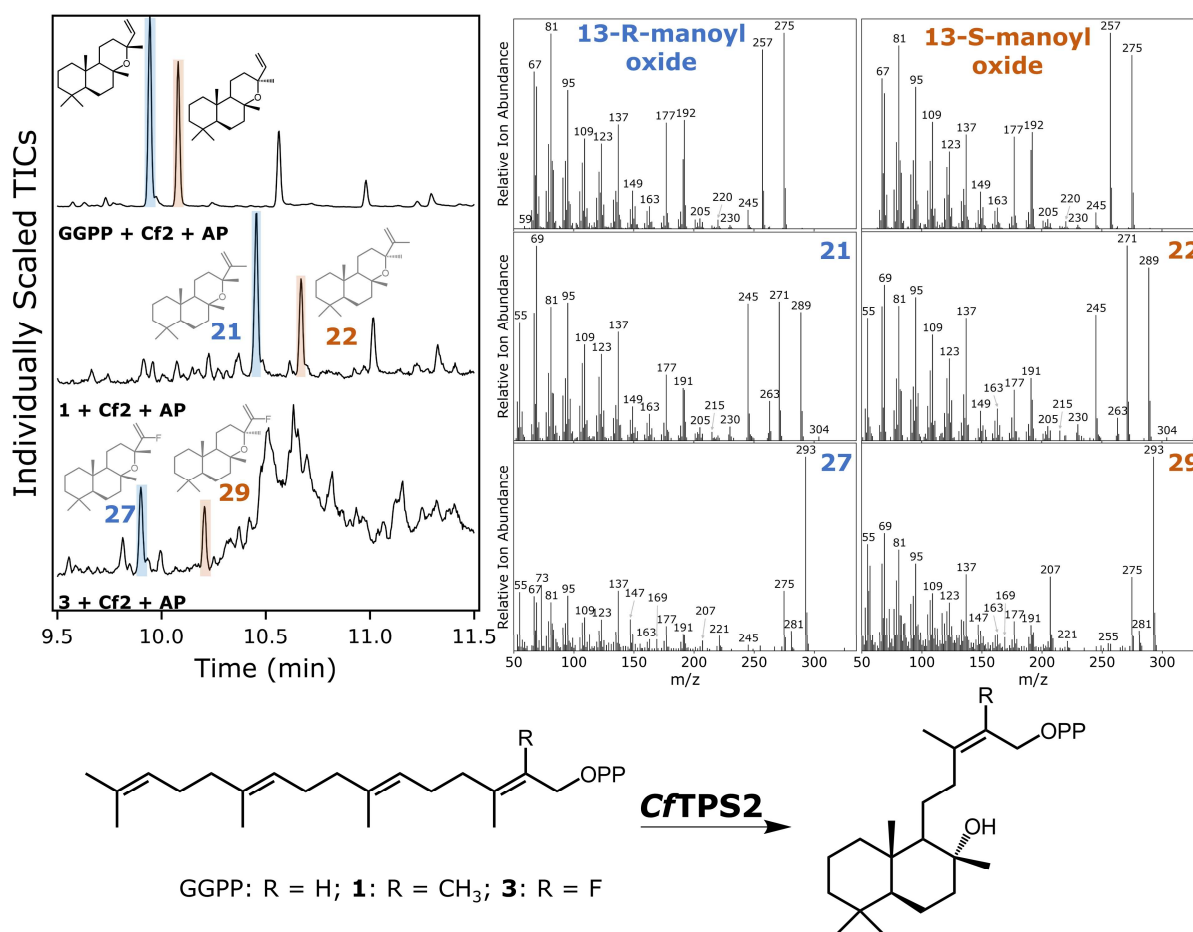


Figure 4.2: Example of class II TPS conversion of modified substrates to derivatives of native products. (Top Left) GC-MS chromatograms of CfTPS2 products with GGPP, **1**, and **3**. Highlighted in blue and orange are stereoisomers of manoyl oxide and likely derivatives of manoyl oxide from conversion of **1** and **3**, with putative structures drawn in gray. (Top Right) Mass spectra of each product, showing a similar fragmentation pattern for manoyl oxide derivatives with major peaks shifted by the substrate's modification (+14 for **1**, +18 for **3**). (Bottom) Predicted structures of the diphosphate intermediates. This structural assignment is supported by NMR on the product of CfTPS2 and SsSS with **1** (Table 4.S1 and Figure 4.S1) and **3** (Table 4.S2 and Figure 4.S2).

modifications on the side of this ether proximal to the diphosphate. This suggests that substrates similar to these may be utilized for the synthesis of labdane-type sesquiterpenes, by mimicking both GGPP in terms of the approximate length of the molecule and FPP in terms of having a potential leaving group after the third isoprenyl subunit. A similar phenomenon is highlighted below for conversion of these substrates by the sesquiterpene synthase PvTPS4 (see Figure 4.4).

Two enzymes—*Cf*TPS2 and *Sh*TPS1—were notably superior in terms of the number of synthetic substrates that each could convert. Grouping of all class II TPSs based on different aspects of their mechanism does not show any obvious trends with respect to the number of substrates that each could convert. For example, grouping those which involve final carbocation quenching through deprotonation (*Zm*AN2: 0 substrates; *Os*SCS: 0; *Pc*TPS1: 0; *Cf*TPS1: 3; *Cf*TPS16: 6), water capture (*Tw*TPS21: 1; *Cf*TPS2: 7; *Ll*TPS1: 0), or a series of methyl and hydride shifts (*Sh*TPS1: 7; *Cam*TPS2: 2; *Ar*TPS2: 2) shows variation within each group. Even pairs of TPSs which make the same native product with GGPP show differences (*Sh*TPS1 and *Cam*TPS2; *Cf*TPS1 and *Cf*TPS16). The contrast between *Cf*TPS1 and *Cf*TPS16 suggests that phylogenetic distance from other promiscuous enzymes could be a better predictor for promiscuity than the enzyme's native mechanism (identical mechanism but *Cf*TPS16 is more closely related to *Cf*TPS2, the most promiscuous enzyme tested).

Screening of Class II/Class I Combinations

Thirteen of these class II/substrate combinations were further screened for combinatorial testing with six second-step class I diTPSs. These enzymes were selected based on their range of native functions and substrate promiscuity, with two enzymes which have been demonstrated to convert every labdane intermediate that they encounter (*Ss*SS^{3,5,6} and *Kg*TS⁶). These two were shown here to be the most promiscuous as well, leading to products in six and five combinations, respectively (Figure 4.1). Only two class II/substrate combinations were shown to have a functional pairing with the other four class I enzymes (**1** with *Cf*TPS2 and **3** with *Cf*TPS2).

The absence of activity of both *Ss*SS and *Kg*TS with the other **3**/class II products is surprising given that these enzyme pairings are functional with GGPP as a starting substrate⁶.

Products detected in class II screening for *TwTPS21* with **3**, for example, are likely the *ent* enantiomers of products highlighted in Figure 4.2 for *CfTPS2* with **3** (**27b** and **29b**; same retention times and mass spectra shown in Figure 4.S9-10), and the natural intermediate *ent*-copal-8-ol has been shown to be converted by *SsSS*^{3,6}, *KgTS*⁶, and *CfTPS3*³. Given that this detection of products was reliant on a coupled assay, this is possibly due to a discrepancy in activity between *CfTPS2* and the other three class II enzymes tested in combination assays with **3**, as this does not decouple the activity of these class I enzymes from the reliance on substrate availability. Conversion of intermediates derived from substrates **4** and **5** also did not show any conversion through class I combinations, likely due to the difference in structure proximal to the phosphate which—in functional combinations—is either not present (**8**), or only a minor change from GGPP (**1** and **3**) which could still allow for the native mechanisms of these respective class I enzymes to be carried out.

Reactions for five of these combinatorial products were scaled up for structural determination by NMR. Consistent with the derivatization of natural products highlighted in Figure 4.2, four out of five combinations showed a direct translation of each enzyme's native mechanism carried through both steps with each modified substrate. *CfTPS2* and *SsSS* make sclareol with GGPP as a substrate³, and the same combination resulted in the formation of 14-methyl (14-methyl-(+)-copal-8,13-ol; **24**), 14-fluoro (14-fluoro-(+)-copal-8,13-ol; **31**), and 11-oxo (11-oxo-(+)-copal-8,13-ol; **46**) derivatives of sclareol with substrates **1**, **3**, and **8**, respectively. Similarly, *CfTPS1* and *SsSS*, which make (+)-13*R*-manool with GGPP³, made 14-fluoro-(+)-copal-13-ol (**34**). The one exception is the combination of *ShTPS1* and *SsSS* converting substrate **8**, which resulted in 11-oxo-*ent*-copal-8,13-ol (**53**), despite kolavenyl diphosphate synthases typically making kolavelool when paired with *SsSS*^{3,21}. The mechanism of a kolavenyl diphosphate

synthase involves the quenching of a carbocation through a series of methyl and hydride shifts, rather than methyl group deprotonation (copalyl diphosphate) or water capture (copal-8-ol diphosphate)^{2,6,7}. Notably, this cation is the branching point in the differentiation between these three products, and it could be that the conversion of **8** by *ShTPS1* involves water quenching rather than a series of methyl and hydride shifts due to slight positional changes in catalysis from the larger substrate.

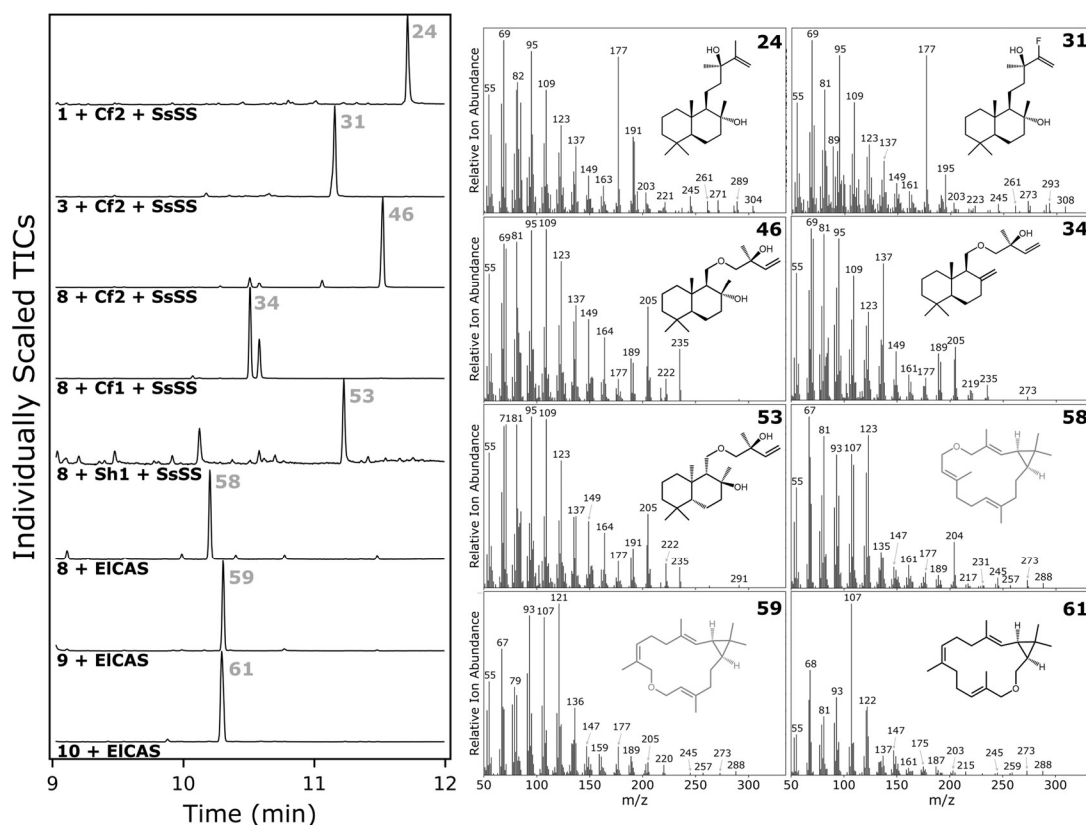


Figure 4.3: Structures of select products. (Left) GC-MS chromatograms for products which were scaled up for NMR analysis (and products of **8/9** + *EICAS*). (Right) Mass spectra for each compound labeled on chromatograms, with structures solved by NMR (Tables 4.S1-6 and Figures 4.S1-6). Drawn in gray for **58** and **59** are putative structures based on the enzyme's native mechanism and similarities to **61** in retention time and mass spectra.

Screening of Single-Step class I TPSs

In addition to combinatorial biosynthesis by pairs of class II/class I TPSs, diterpenes can be synthesized in a single step through the activity of a class I enzyme². The majority of mono-

and sesquiterpenes are made in this manner but is relatively uncommon in diterpene biosynthesis. The majority of examples of diTPSs that can carry this out are compartment-switching members of the TPS-a subfamily (typically cytosolic sesquiterpene synthases), as switching of compartments grants these enzymes access to new substrates (e.g. GGPP) *in vivo*, which is a phenomenon that has been seen in at least five independent plant families^{9,24}. Each TPS screened here is an example of a compartment-switching TPS-a, with the exception of *Pv*TPS4, which is a cytosolic member of this subfamily which natively functions as a sesquiterpene synthase, but is closely related to a set of enzymes which have switched compartments (including *Pv*HVS, also screened here)⁹.

Screening of each of these TPSs led to three results which are immediately apparent. First, three enzymes were non-functional with all twenty substrates. In the case of *Lf*TPS1 and *Es*TPS1, this likely due to the substrate specificity that these two enzymes exhibit towards their native substrate NNPP (nerylneryl diphosphate: the all-*cis* stereoisomer of GGPP) as neither have any activity with GGPP²⁴. For *Pv*HVS, the native mechanism of this enzyme in converting GGPP involves cyclization throughout the entire length of the molecule resulting in 4 rings²⁸, and a modification to any part of its substrate could result in an interruption of this mechanism. This follows the same concept as why no class II enzymes could convert substrates **9** or **10**, as these would interrupt the native cyclization mechanism.

Second, TPSs which make casbene and cembrenol were functional in converting both the methyl (**1**) and ether (**8-10**) substrates. In contrast to the lack of activity with *Pv*HVS, these mechanisms are much less complicated and only involve the first and last isoprenyl subunits in the cyclization mechanism. The methyl addition to **1** is likely a small enough change to not interfere with these mechanisms, while the ethers present in **8-10** are located away from these isoprenyl

subunits. Scaleup and NMR on the product of **10** and *E/CAS* revealed that **61** is 13-oxo-casbene (Figure 4.3), showing that this diTPS's native mechanism can progress with the addition of an ether at this position. The ability for *E/CAS* to convert **8** and **9** as well suggest that the structures of **58** and **59** could be 5- and 9-oxocasbene, respectively (Figure 4.3). Surprisingly, *DgTPS1* was able to convert a small amount of **13** (Figure 4.S8) to products **63** and **64**, both of which have molecular ions consistent with this substrate dephosphorylated (310 m/z; Figure 4.S10). This was the only example of a TPS able to convert any of the phenyl ring substrates (**12-14**).

Third, *PvTPS4* (the only sesquiterpene synthases tested here) converted **4**, **5**, **8**, **9**, **12**, **17**, and **18** to cadinene—its native sesquiterpene product (Figure 4.4). This is especially interesting in that all of these substrates share the same structure as its native substrate FPP, with the exception of the modifications between the third isoprenyl subunit and the diphosphate, which all begin with an ether. The ability for this enzyme to convert *all* of these substrates to cadinene implies that each can be initially converted to a farnesyl cation—the starting point in the mechanism of a sesquiterpene synthase following dephosphorylation of FPP²⁹. A proposed mechanism is shown for substrate **4** in Figure 4.4, which suggests that this ether allows for the entire modified portion of these substrates to act as a leaving group, resulting in a farnesyl cation. This could explain why small amounts of product **69** (a putative sesquiterpene) can be seen through the conversion of **4**, **5**, and **8** by class II TPSs, as they could be cyclized prior to dephosphorylation that results in this same type of loss.

General Trends

As highlighted throughout each set of screens, successful combinations tended to be those where the modifications present in the substrate are compatible with the native mechanism of the

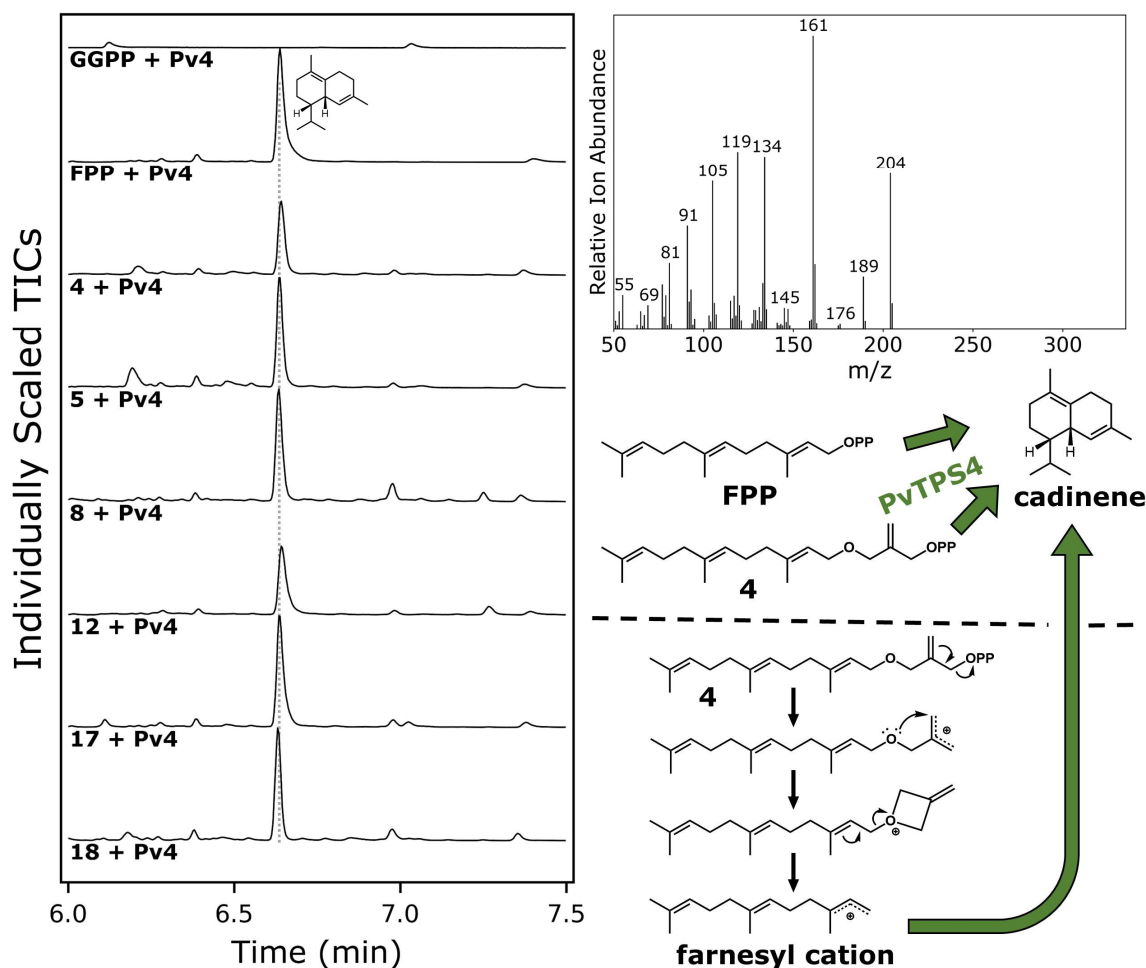


Figure 4.4: *PvTPS4* converts six modified substrates to cadinene. (Left) GC-MS chromatographs showing each substrate converted to cadinene by *PvTPS4*. (Top Right) Mass spectrum of cadinene. (Bottom Right) Proposed mechanism for the formation of a farnesyl cation from 4.

enzymes converting them. In the case of class II enzymes, no substrates with modifications to the first three isoprenyl subunits—involved in the cyclization mechanism towards a labdane core—resulted in successful conversions. Likewise, no successful conversion was seen in class II/class I combination screening for substrates 4 or 5, likely due to these modifications being too dissimilar from GGPP proximal to the diphosphate (in contrast to 1 and 3) for any of the class I enzymes to convert the modified class II intermediates. Finally, *PvHVS* could not convert any substrate, likely because its native cyclization mechanism involves modifications along the entire length of the substrate.

These similarities in substrate structure, which presumably allow them to be converted by diTPSs, generally translate to similarities in product structure as well. The product profile for class II products (see Figure 4.2) and in solved structures (see Figure 4.3) demonstrate that the products of successful combinations are largely derivatives of products which would be made with GGPP. The design of synthetic substrates for the purpose of derivatizing a given terpene scaffold may benefit from a careful inspection of the cyclization mechanism of the enzyme used to make it, as specific modifications which interfere with this mechanism may be prohibitory for enzyme function.

Some enzymes screened here had a remarkable ability to convert a wide range of substrates while others converted none at all. Consistent with prior studies^{3,5,6}, *SsSS* and *KgTS* were capable of converting a range of class II-derived intermediates. Second-step class I diTPSs have received considerable attention for their substrate promiscuity, largely because of the number of labdane intermediates which exist in nature. In contrast, fewer studies have addressed the promiscuity of class II and single-step class I diTPSs, likely owing to the more limited number of natural substrates available for testing. Some class II diTPSs have been shown to convert NNPP to irregular labdane structures⁸, and some single step class I diTPSs have been shown to convert various prenyl diphosphate substrates with varying length and stereochemistry⁹. This work demonstrates that promiscuous enzymes which can convert a range of substrates can be found in both categories: especially *CfTPS2* and *ShTPS1* for class II enzymes and *ElCAS* for class I.

Future Perspectives

The use of chemically synthesized GGPP derivatives has given us the opportunity to probe the substrate promiscuity of these enzymes with respect to substrates that do not occur in nature.

This offers a unique approach towards making diterpene derivatives that are not accessible otherwise, allowing for the testing of potential applications for these new molecules. Production of these derivatized metabolites may or may not be possible in a purely biosynthetic system. For methyl derivatives (e.g. **1**), previous work has demonstrated the incorporation of methyltransferases for methylation of GPP (in the same position as **1** with respect to the diphosphate) for C11 monoterpenes¹⁴, and an unusual precursor pathway from lepidoptera (butterflies and moths) producing modified FPP precursors for C16 sesquiterpenes¹³. For ether derivatives (e.g. **8-10**), one could imagine a prenyl-transferase which incorporates a hydroxylated dimethylallyl diphosphate into a growing prenyl diphosphate precursor, although we are unaware of an example either found in nature or engineered to do so. Fluorinated substrates (e.g. **3**) could require the discovery or engineering of a halogenase to carry out a desired fluorination, although very few fluorinated natural products and fluorinases have been discovered^{30,31} and this would likely be a significant challenge. The semi-biosynthetic methodology employed here merges the advantages of reactions which are more feasible through synthetic chemistry with those that are more feasible through biocatalysis.

The ability to derivatize terpene scaffolds through this methodology may be especially useful when considering that many terpenoids are used for medicinal purposes. In contrast to “combinatorial chemistry” methods that have historically been employed in drug discovery which sample thousands of combinatorial libraries^{32,33}, derivatization of a particular natural product of interest allows for the sampling of chemical space around what has already been demonstrated to be effective. Forskolin is a bioactive compound which acts as a cyclic-AMP booster³⁴, and is derived from the terpene scaffold 13*R*-(+)-manoyl oxide^{22,35}, which we have derivatized here (Figure 4.2). Likewise, prostratin has been used in the treatment of HIV^{36,37}, and ingenol mebutate

in the treatment of actinic keratosis (a skin disorder preceding skin cancer)^{38,39}: both derived from casbene²³ which was also derivatized here (Figure 4.3). The production of derivatized variants of any of these final metabolites would require the conversion of these modified substrates and intermediates through further steps in these biosynthetic pathways, highlighting the importance of expanding this research into downstream classes of enzymes (cytochrome P450s in particular^{40,41}). In contrast to the examples above, sclareol has been demonstrated to have anti-inflammatory properties^{42,43} and we have directly derivatized this compound with no need for further biosynthetic steps.

Screening with these twenty substrates has demonstrated the range of substrate promiscuity for these TPSs beyond what would be possible solely with substrates found in nature. A particularly interesting avenue of research would be to figure out how to achieve the same level of substrate promiscuity in TPSs which carry out different native reactions. One could envision multiple approaches to this end: the first being sampling of more enzymes for activity against these substrates, as our analysis only included one or two enzymes for each natural product represented. The higher promiscuity of *Cf*TPS16 relative to *Cf*TPS1 indicates that sampling enzymes phylogenetically closer to those with known promiscuity (i.e. *Cf*TPS2) could be an effective strategy. Another approach would be to make specific mutations in a promiscuous enzyme to alter product profile and see if this results in the retention of substrate promiscuity. *SsSS* for example, has been engineered through single-residue substitutions to switch the stereochemistry of its hydroxylation⁴⁴. Simple active site mutations have been made to alter the product profile of class II TPSs⁴⁵⁻⁴⁷, and it would be interesting to see whether mutations could be made, for example, to *Cf*TPS2 to engineer it into a (+)-kolavenyl diphosphate synthase with retention of its substrate promiscuity.

Materials and Methods

Cloning and Sources of Genes Used

Original characterization of each enzyme tested in this study is listed in Table 4.S1. Coding sequences for each gene were PCR amplified with overhangs for In-Fusion cloning into the bacterial expression vector pET28-b(+). Enzymes which are natively targeted to the plastid had their N-terminal transit peptides removed—predicted through a combination of TargetP (v 1.1⁴⁸) and sequence alignments with Clustal Omega (v1.2.4⁴⁹), and C-terminal stop codons were removed for the addition of a C-terminal 6x His tag from the pET28-b(+) vector.

Phylogenetic Tree

Amino acid sequences for each enzyme (except the bacterial KgTS) were truncated according to their predicted N-terminal transit peptide length. A bifunctional *ent*-CPP/*ent*-kaurene synthase from *Physcomitrium patens* (*PpCPS/KS*)⁵⁰ was also included as an outgroup (not shown in Figure 4.S1). Sequences were aligned with Clustal Omega (v1.2.4⁴⁹) and a maximum-likelihood phylogenetic tree with 1,000 bootstraps was generated with RAxML (v8.0.0⁵¹) and visualized with iTOL (v5⁵²). Three different trees were generated with different random starting seeds (-p option in RAxML) and each tree displayed the same topology as that shown in Figure 4.S1.

Enzyme Expression and Purification

Expression and purification of TPSs was carried out as described in Johnson *et al.* 2019⁹ and Miller *et al.* 2020²⁴ (See Chapter 2) with 50 mL of expression culture for enzyme production in sufficient quantities for GC-FID/MS screening-scale assays. For scaled-up expression/purification, all methods following the inoculation and growth of an overnight culture

were directly scaled up by a factor equal to the change in expression culture volume (i.e. 5x scaleup for 250 mL of expression culture). Ni-NTA columns were set up with 100 μ L bed volume (His60 Ni Superflow Resin; Clontech Laboratories) per 50 mL of expression culture and lysate supernatants were loaded twice prior to washing and elution with the same buffers as described for screening-scale purification (with appropriate volume scaling as detailed above). Eluted enzymes were desalted following the same method for expression/purification at all scales.

In vitro assays

Modified GGPP substrates were synthesized by Matthew Giletto and Edmund Ellsworth (Michigan State University Medicinal Chemistry Facility) and obtained at approximately 50% purity by weight (phosphate impurities) and were resuspended in 70% methanol : 30% water to 2 mg/mL (or to saturation). Screening-scale assays (for GC-FID/GC-MS analysis) were carried out with 50 μ g enzyme ($\sim$ 0.75 μ M for class II enzymes and *EpTPS1*, $\sim$ 1 μ M for remaining class I enzymes) and 10 μ g substrate ($\sim$ 30 μ M depending on substrate) in 750 μ L Reaction Buffer (50 mM HEPES, pH 7.2, 7.5 mM MgCl₂, and 5% (v/v) glycerol). For single-step class I assays, 500 μ L of hexane overlay was added immediately and reactions were carried out at 30°C for 16 hours. For class II assays, reactions were carried out at 30°C for 3 hours prior to the addition of alkaline phosphatase (1 U; Promega) and 500 μ L hexane overlay, and reactions proceeded at 30°C for 16 hours. For class II/class I combination assays, the same method as listed above for class II assays was carried out except for the addition of the respective class I enzyme instead of a phosphatase. All reactions were carried out in 1.5 mL vials. Reaction mixtures were vortexed and centrifuged to re-separate aqueous and organic layers. The entire hexane layer was removed, transferred to a

new vial, and dried down to ~50 μ L at 40°C and constant air flow prior to analysis by GC-FID or GC-MS.

GC-FID/MS analysis

All GC-MS analyses were performed on an Agilent 7890A GC with an Agilent VF-5ms column (30 m x 250 μ m x 0.25 μ m, with 10m EZ-Guard) with the inlet set to 250°C splitless injection of 1 μ L and He carrier gas (1 ml/min). Initial screening was carried out with a flame ionization detector, and rescreening by mass spectrometry was carried out with an Agilent 5975C mass spectrometer using 70 eV electron ionization. All analyses used the same GC method as listed in Chapter 3: temperature ramp start 40°C, hold 1 min, 40°C/min to 200°C, hold 2 min, 20°C/min to 280°C, 40°C/min to 320°C; hold 5 min. All figures for chromatograms and mass spectra were generated with Pyplot.

Scaleup and NMR

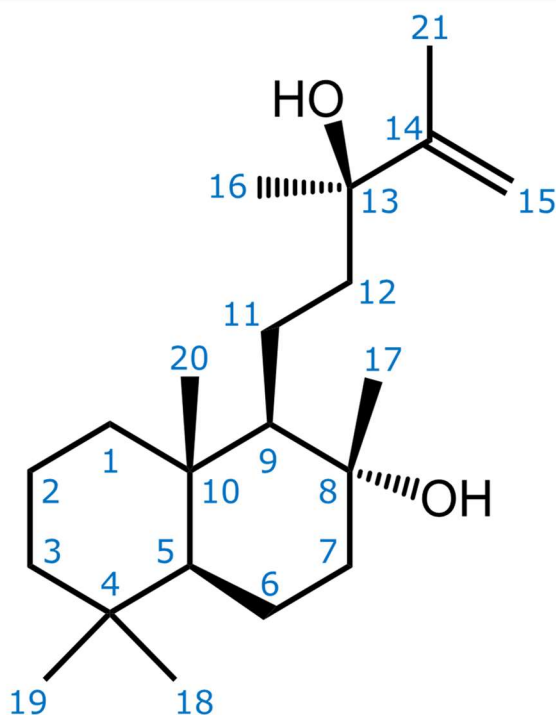
Scaleup production of compounds for NMR analysis was carried out with a direct 50x scaleup of screening-scale assays: 2.5 mg enzyme and 500 μ g substrate in 37.5 mL Reaction Buffer, split evenly between two 40 mL vials, with 12.5 mL hexane overlay (each). Reactions were carried out as described above for screening scale assays. Products were dried down at 40°C and constant air flow and resuspended in 250 μ L hexane. Resuspended products were purified through silica chromatography and loaded into ~1.5 mL oven-dried silica gel preequilibrated with hexane in a Pasteur pipette, and eluted with the following series of solvents: 3 mL 100% hexane, 3 mL 5% ethyl acetate in hexane, 3 mL 10 % ethyl acetate in hexane, and 100% ethyl acetate. Fractions were collected in 1 mL each and assessed for purity by GC-FID. Pure fractions were dried, washed and

dried twice with CDCl_3 , and resuspended in ~ 600 μL CDCl_3 . All NMR analysis was carried out on a Bruker 800 MHz spectrometer equipped with a TCI cryoprobe. CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively.

APPENDIX

Table 4.S1: ^1H and ^{13}C chemical shifts for (+)-14-methylcopal-8,13-ol (**24**). CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Stereochemistry as drawn is based on the native mechanisms of *CfTPS2* and *SsSS*. Shifts highlighted in gray are tentative assignments.

^{13}C NMR	δ	^1H NMR
C1	39.7	1.12 (m, 1H); 1.67 (m, 1H)
C2	18.5	1.45 (m, 1H); 1.59 (m, 1H)
C3	42.0	1.16 (m, 1H); 1.38 (m, 1H)
C4	33.2	-
C5	56.1	0.93 (m, 1H)
C6	20.6	1.27 (m, 1H); 1.66 (m, 1H)
C7	44.6	1.43 (m, 1H); 1.84 (m, 1H)
C8	74.7	-
C9	61.6	1.12 (m, 1H)
C10	39.2	-
C11	19.1	1.31 (m, 2H)
C12	43.3	1.67 (m, 1H); 1.77 (m, 1H)
C13	76.0	-
C14	151.1	-
C15	109.2	4.84 (s, 1H); 5.02 (s, 1H)
C16	27.6	1.33 (s, 3H)
C17	24.3	1.17 (s, 3H)
C18	33.4	0.88 (s, 3H)
C19	21.5	0.8 (s, 3H)
C20	15.5	0.8 (s, 3H)
C21	19.6	1.79 (s, 3H)



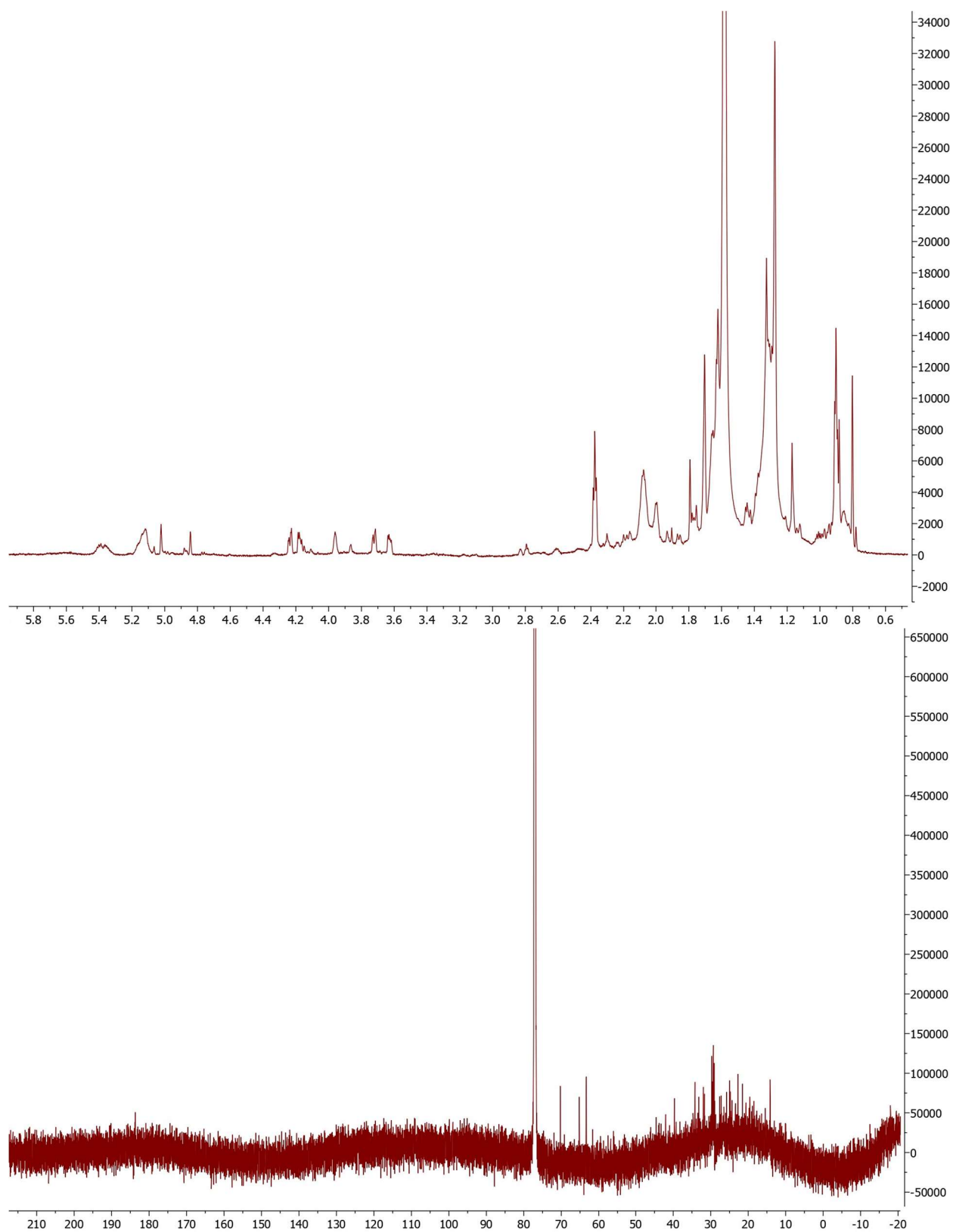


Figure 4.S1: NMR spectra (^1H , ^{13}C , HSQC, H2BC, HMBC, NOESY, COSY) for (+)-14-methylcopal-8,13-ol (24).

Figure 4.S1 (cont'd)

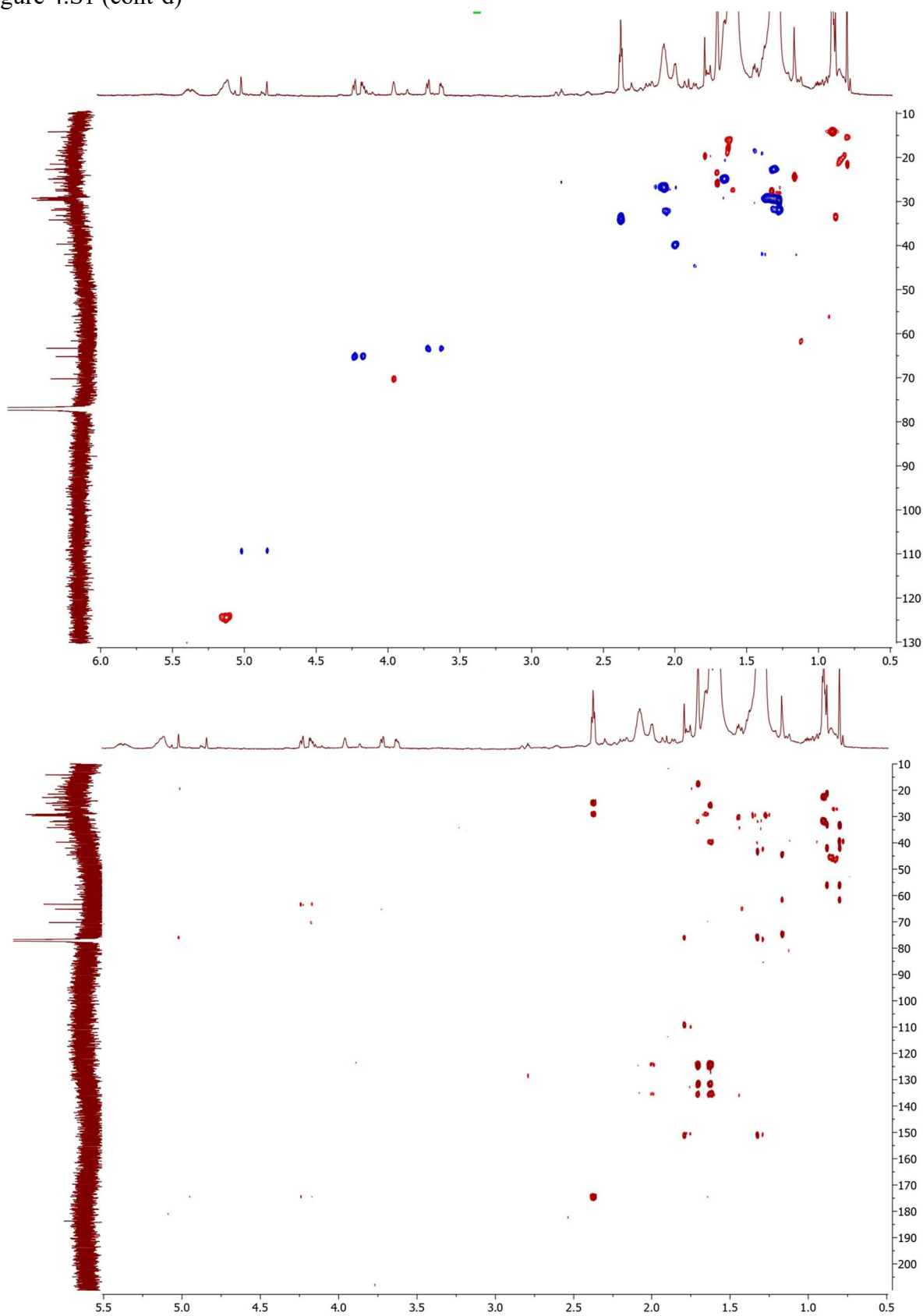


Figure 4.S1 (cont'd)

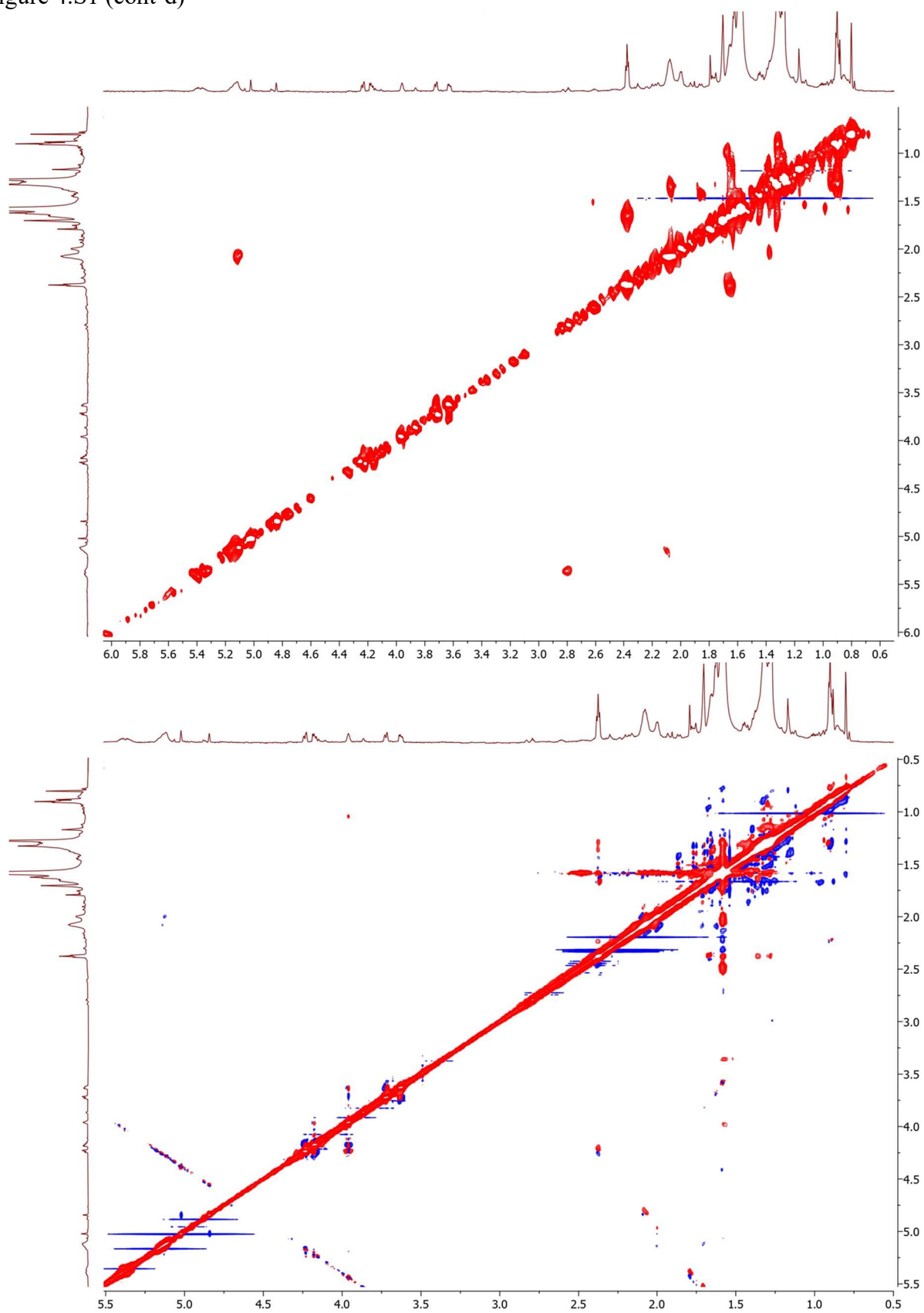
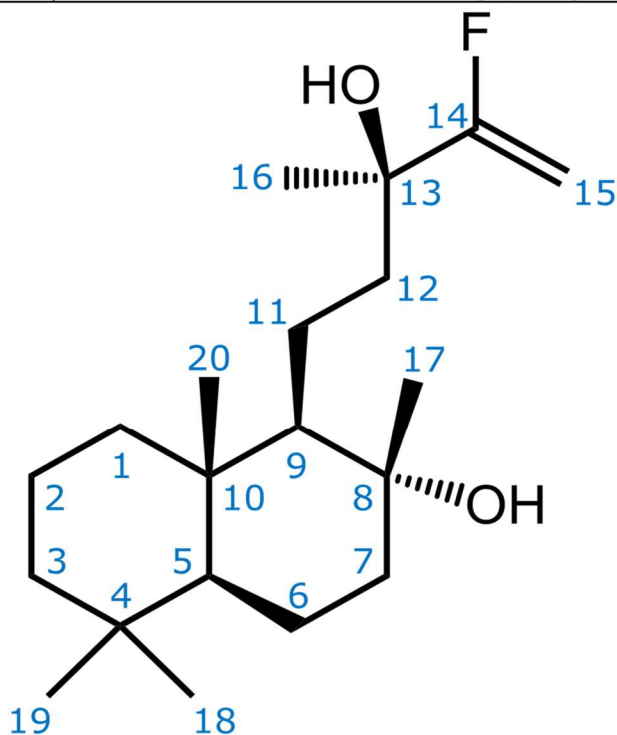


Table 4.S2: ^1H and ^{13}C chemical shifts for (+)-14-fluorocopal-8,13-ol (**31**). CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Stereochemistry as drawn is based on the native mechanisms of *Cf*TPS2 and *Ss*SS.

^{13}C NMR	δ	^1H NMR
C1	39.6	0.99 (m, 1H); 1.65 (m, 1H)
C2	18.4	1.46 (m, 1H); 1.60 (m, 1H)
C3	42.0	1.17 (m, 1H); 1.39 (m, 1H)
C4	33.3	-
C5	56.0	0.95 (m, 1H)
C6	20.5	1.28 (m, 1H); 1.67 (m, 1H)
C7	44.6	1.45 (d, $J = 13.6$ Hz, 1H); 1.86 (t, $J = 11.2$ Hz, 1H)
C8	75.0	-
C9	60.8	1.22 (s, 1H)
C10	39.2	-
C11	19.0	1.40 (m, 1H); 1.54 (m, 1H)
C12	41.7	1.78 (m, 1H); 1.85 (m, 1H)
C13	73.0	-
C14	170.2	-
C15	88.3	4.63 (m, 1H); 4.69 (m, 1H)
C16	25.8	1.39 (, 1H)
C17	24.4	1.19 (s, 1H)
C18	33.4	0.89 (s, 1H)
C19	21.5	0.80 (s, 1H)
C20	15.3	0.81 (s, 1H)



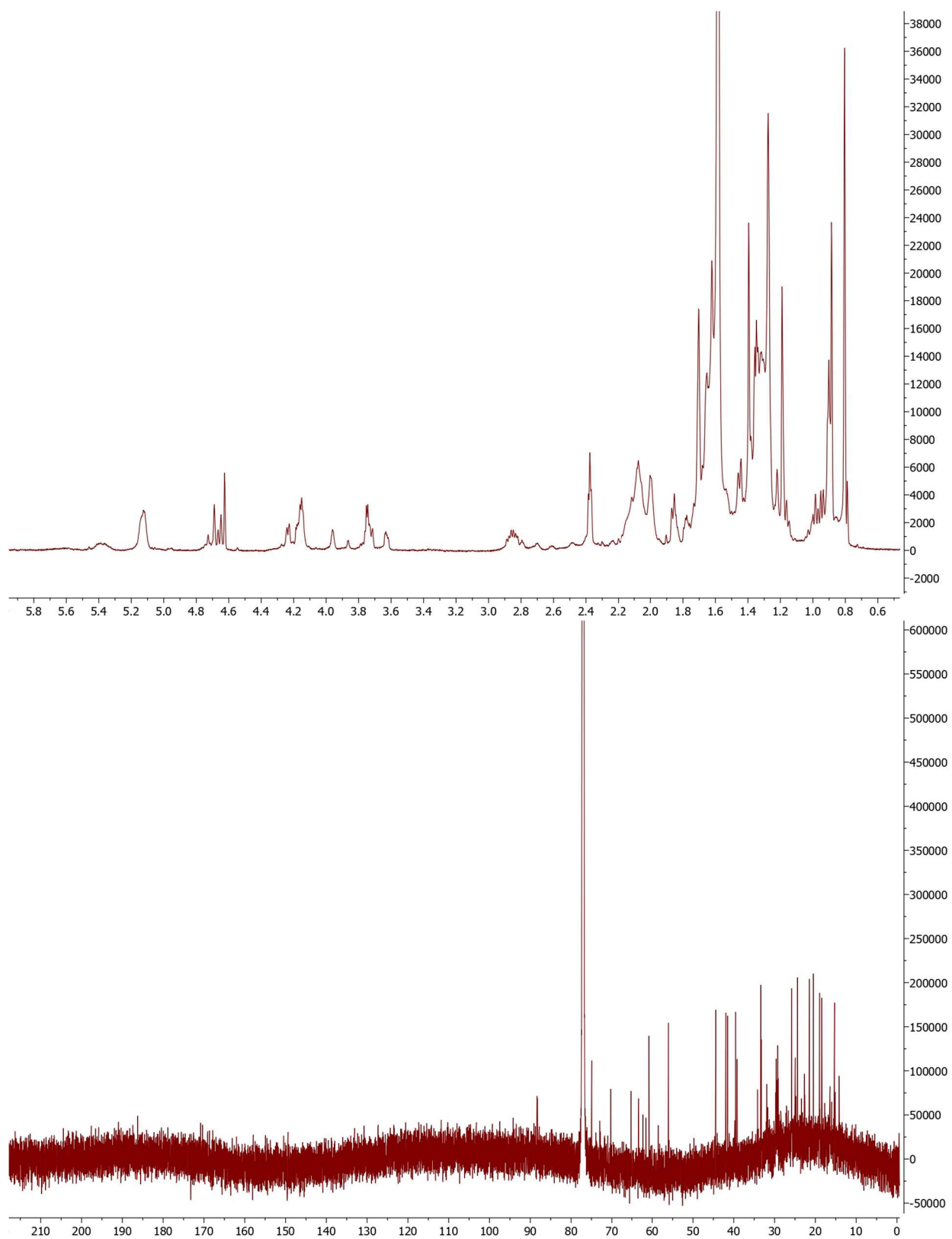


Figure 4.S2: NMR spectra (^1H , ^{13}C , HSQC, H2BC, HMBC, NOESY, COSY) for (+)-14-fluorocopal-8,13-ol (**31**).

Figure 4.S2 (cont'd)

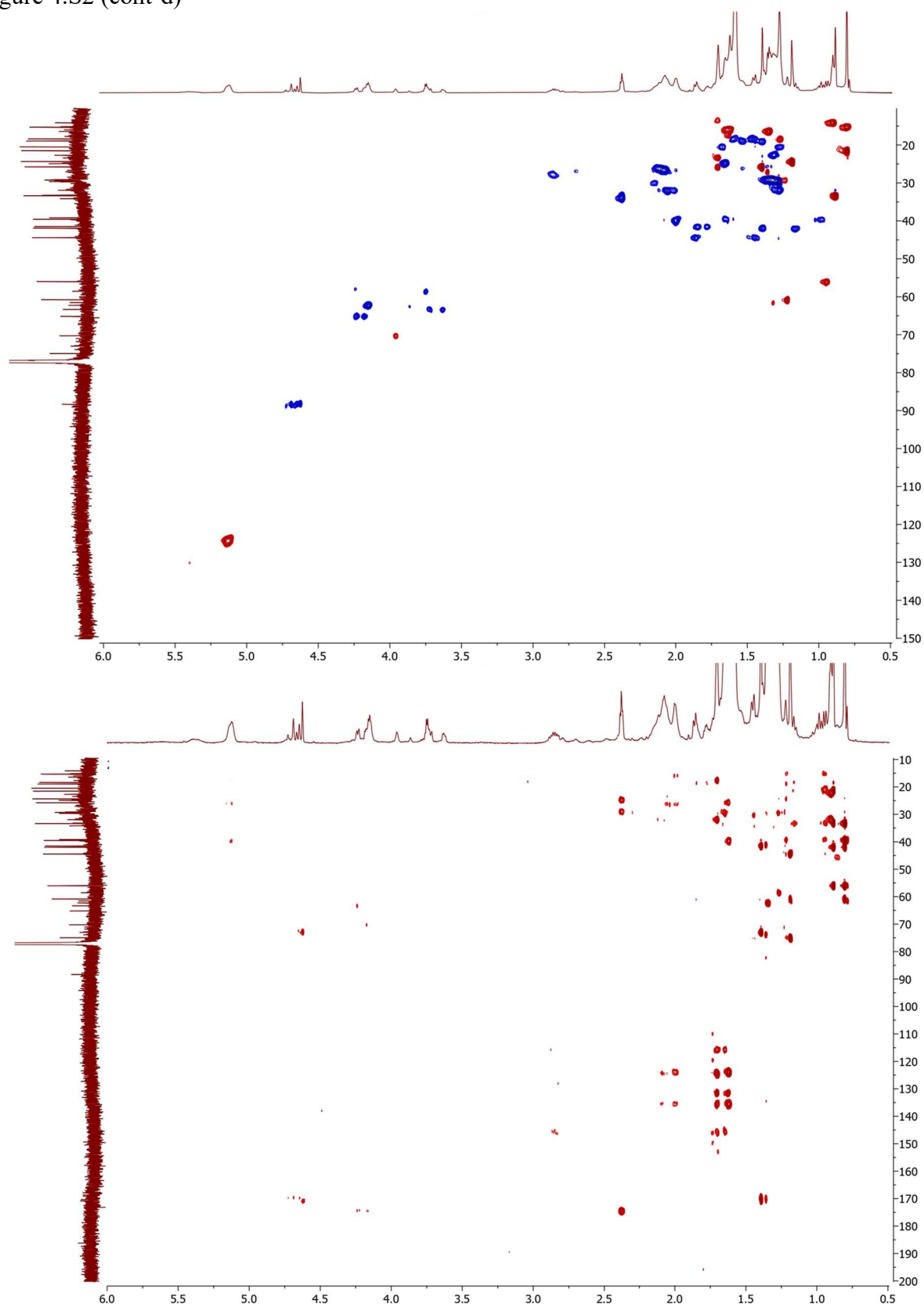


Figure 4.S2 (cont'd)

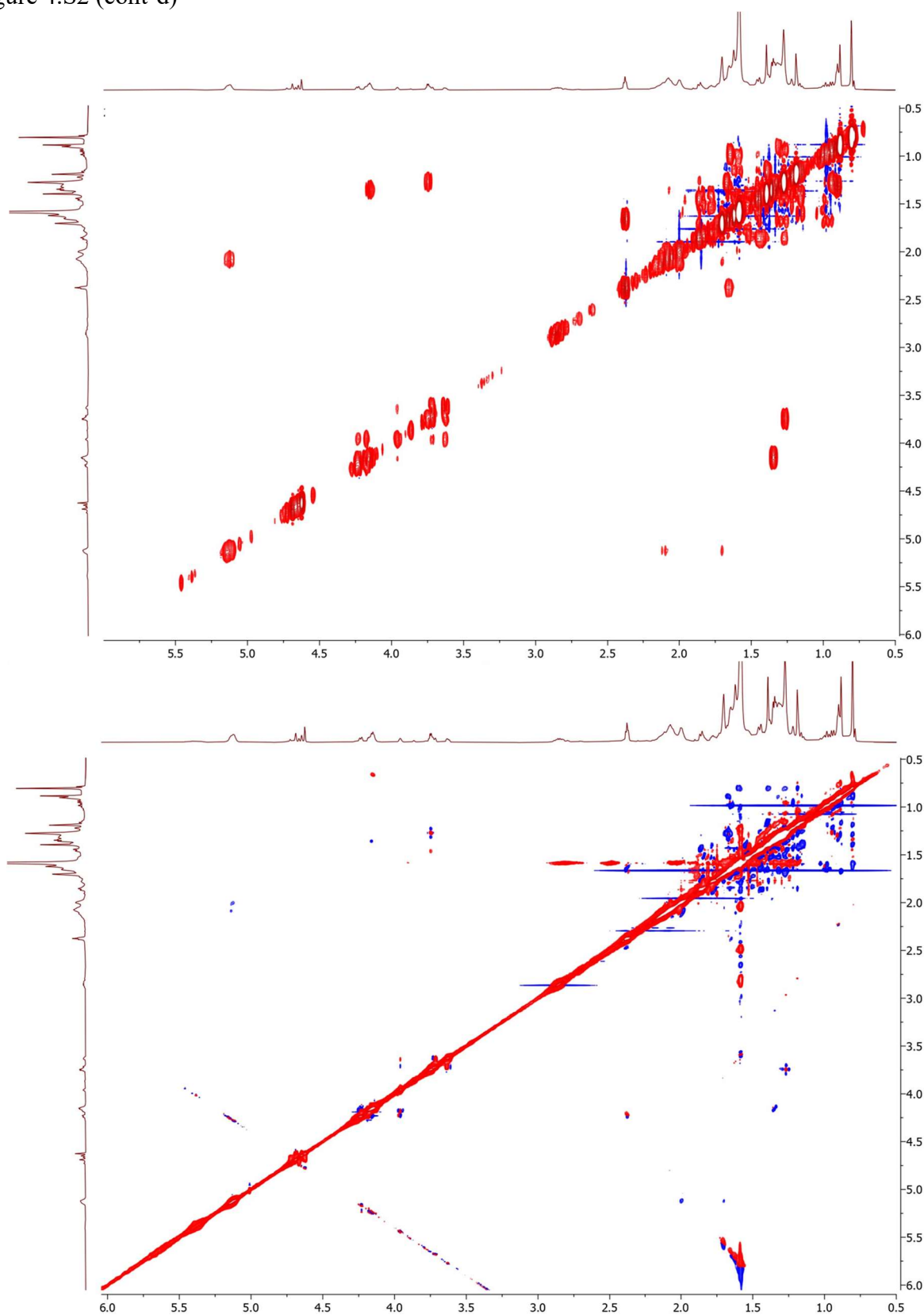
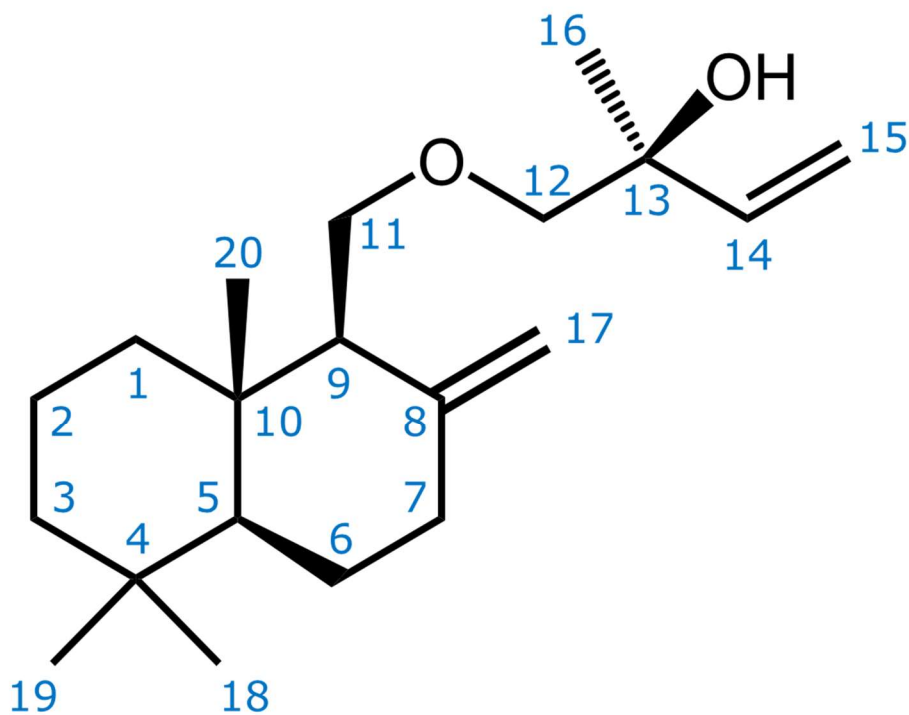


Table 4.S3: ^1H and ^{13}C chemical shifts for (+)-11-oxo-copal-13-ol (**34**). CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Stereochemistry as drawn is based on the native mechanisms of *Cf*TPS16 and *Ss*SS.

^{13}C NMR	δ	^1H NMR
C1	39.3	1.12 (m, 1H); 1.61 (m, 1H)
C2	19.0	1.41 (m, 1H); 1.48 (m, 1H)
C3	42.1	1.12 (m, 1H); 1.33 (m, 1H)
C4	33.6	-
C5	55.0	1.06 (m, 1H)
C6	23.9	1.24 (m, 1H); 1.64 (m, 1H)
C7	37.6	2.31 (d, $J = 13.4$ Hz, 1H)
C8	171.3	-
C9	55.4	1.93 (m, 1H)
C10	38.8	-
C11	68.2	3.57 (m, 2H)
C12	77.9	3.2 (m, 2H)
C13	72.6	-
C14	142.6	5.83 (d, $J = 6.2$ Hz, 1H)
C15	113.4	5.02 (d, $J = 10.7$ Hz, 1H); 5.21 (d, $J = 17.2$ Hz, 1H)
C16	24.2	1.16 (s, 3H)
C17	107.5	4.49 (s, 1H); 4.77 (s, 1H)
C18	21.8	0.74 (s, 3H)
C19	33.7	0.8 (s, 3H)
C20	15.2	0.66 (s, 3H)



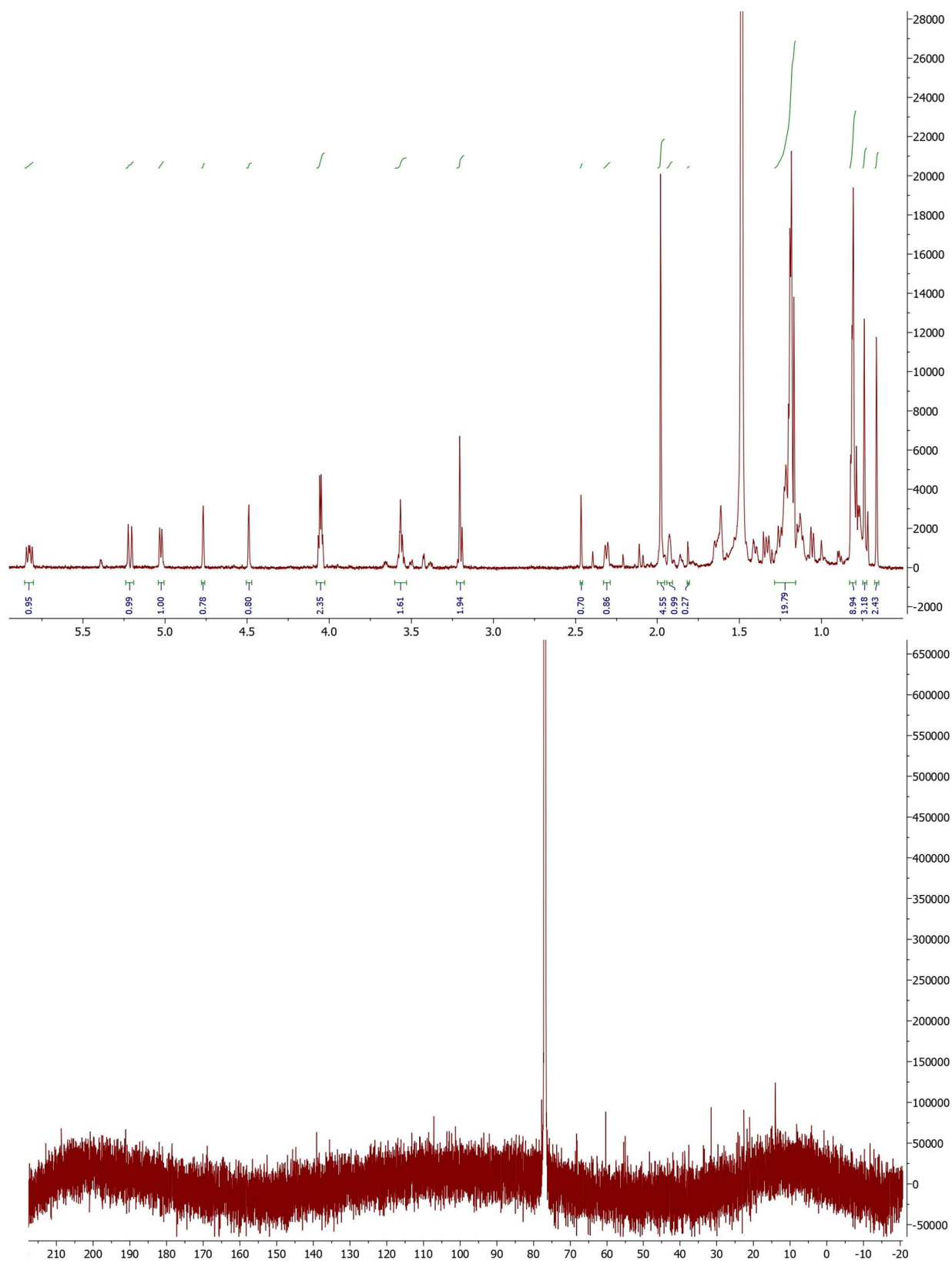


Figure 4.S3: NMR spectra (^1H , ^{13}C , HSQC, H2BC, HMBC, NOESY, COSY) for (+)-11-oxo-copal-8-ol (34).

Figure 4.S3 (cont'd)

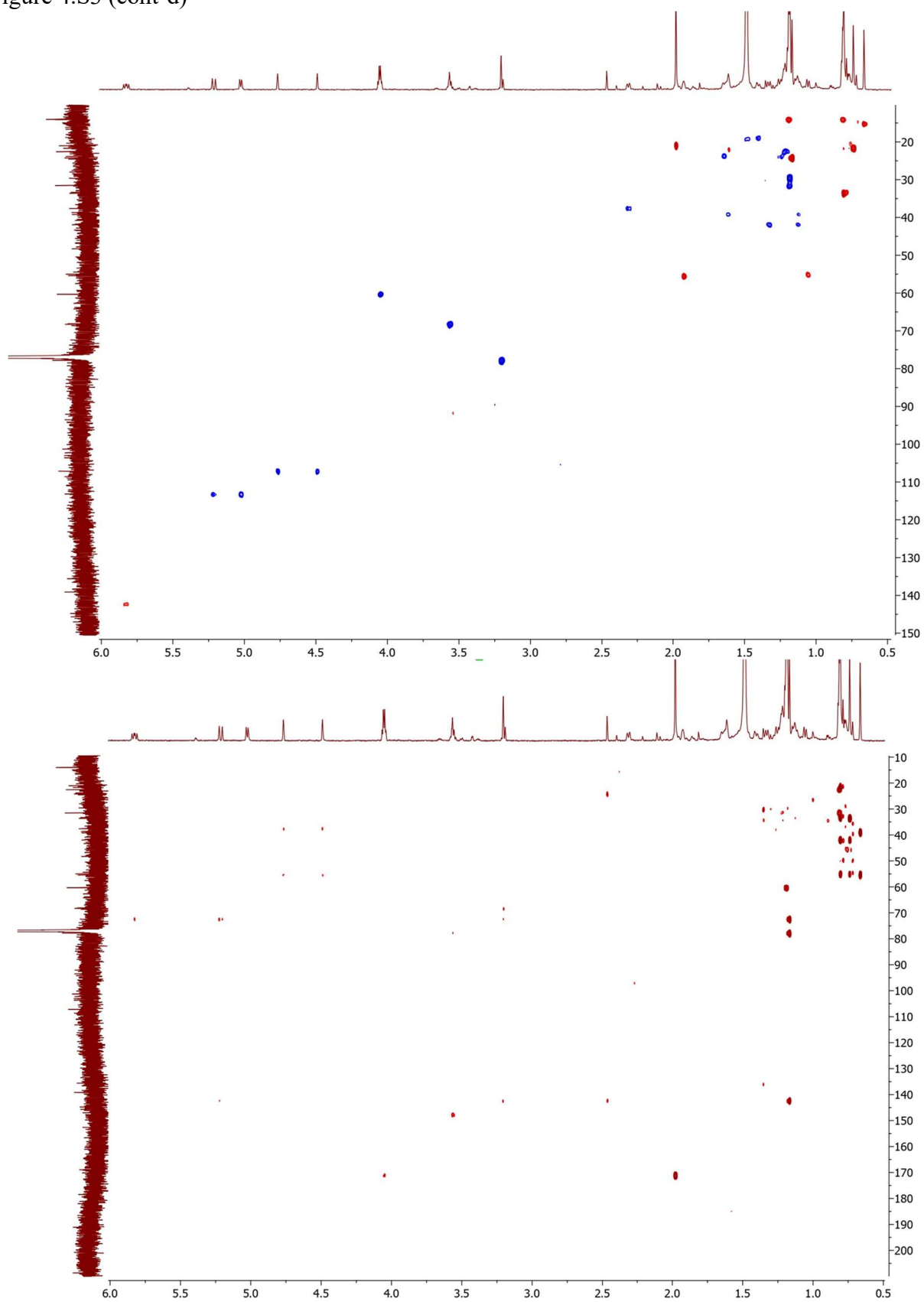


Figure 4.S3 (cont'd)

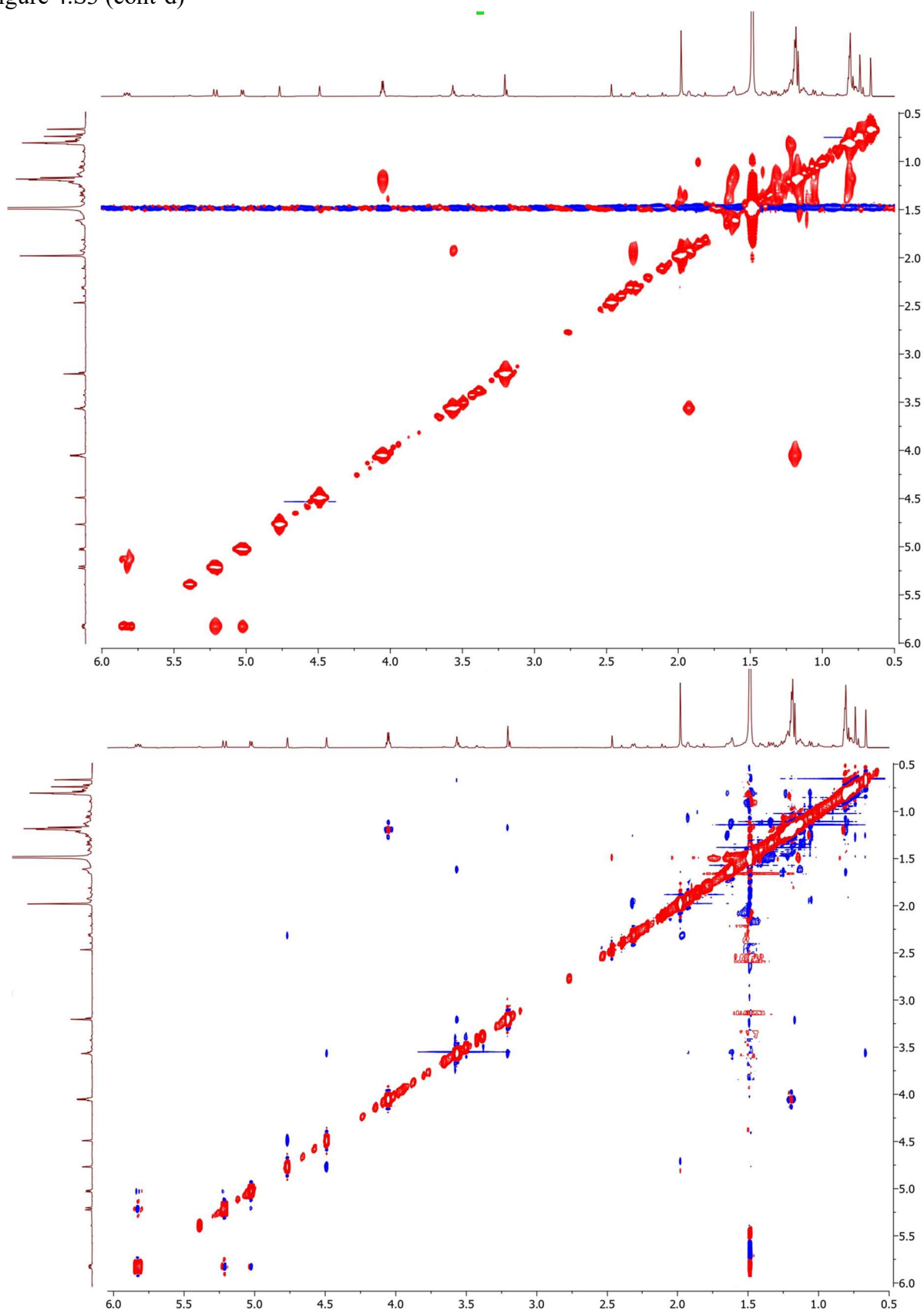
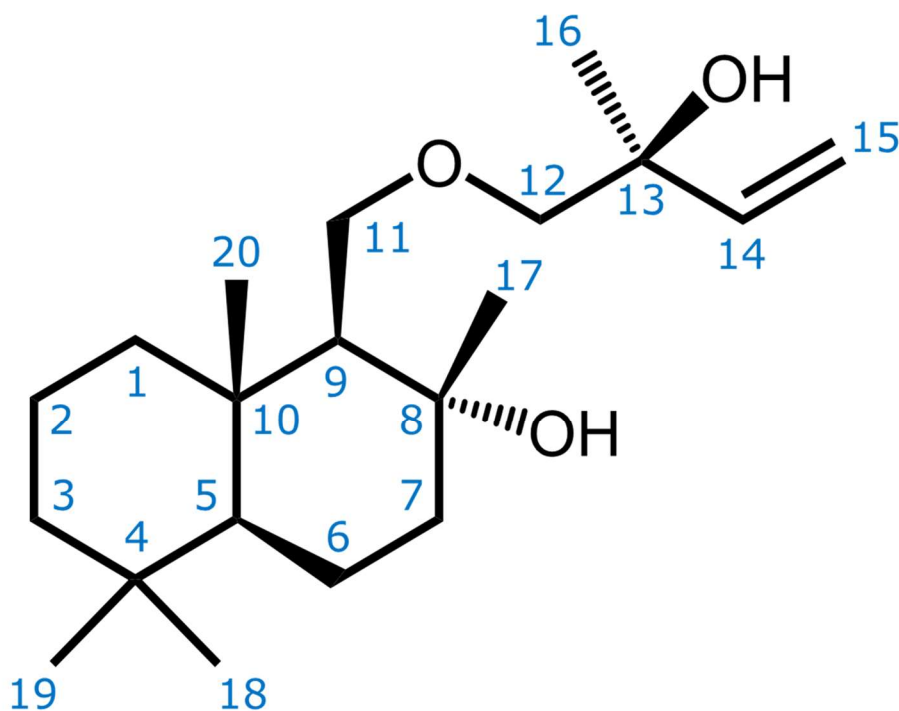


Table 4.S4: ^1H and ^{13}C chemical shifts for (+)-11-oxo-copal-8,13-ol (**46**). CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Stereochemistry as drawn is based on the native mechanisms of *Cf*TPS2 and *Ss*SS.

^{13}C NMR	δ	^1H NMR
C1	40.2	1.09 (m, 1H); 1.70 (m, 1H)
C2	18.6	1.46 (m, 1H); 1.59 (m, 1H)
C3	41.7	1.19 (m, 1H); 1.40 (d, $J = 13.3$ Hz, 1H)
C4	33.3	-
C5	55.8	0.99 (m, 1H)
C6	19.8	1.24 (m, 1H); 1.65 (m, 1H)
C7	42.8	1.52 (m, 1H); 1.88 (d, $J = 12.5$ Hz, 1H)
C8	72.9	-
C9	58.1	1.68 (m, 1H)
C10	37.3	-
C11	70.7	3.74 (t, $J = 9.6$ Hz, 1H); 3.81 (m, 1H)
C12	78.8	3.35 (d, $J = 8.5$ Hz, 1H); 3.45 (d, $J = 8.5$ Hz, 1H)
C13	72.8	-
C14	142.4	5.91 (m, 1H)
C15	113.9	5.19 (d, $J = 11.7$ Hz, 1H); 5.35 (d, $J = 17.3$ Hz, 1H)
C16	24.3	1.26 (s, 1H)
C17	25.1	1.27 (s, 1H)
C18	33.6	0.90 (s, 1H)
C19	21.7	0.81 (s, 1H)
C20	16.0	0.82 (s, 1H)



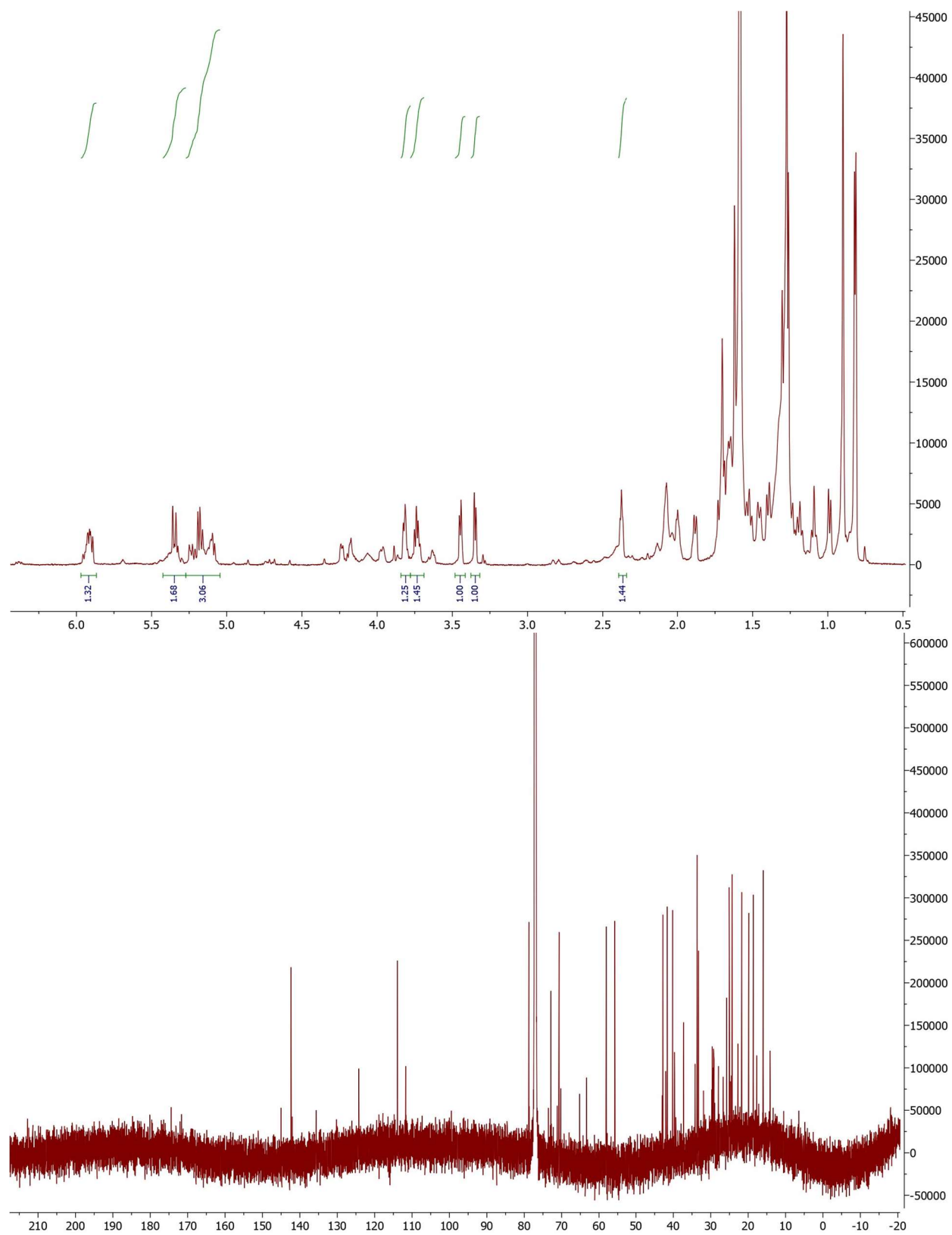


Figure 4.S4: NMR spectra (^1H , ^{13}C , HSQC, H2BC, HMBC, NOESY, COSY) for (+)-11-oxo-copal-8,13-ol (46).

Figure 4.S4 (cont'd)

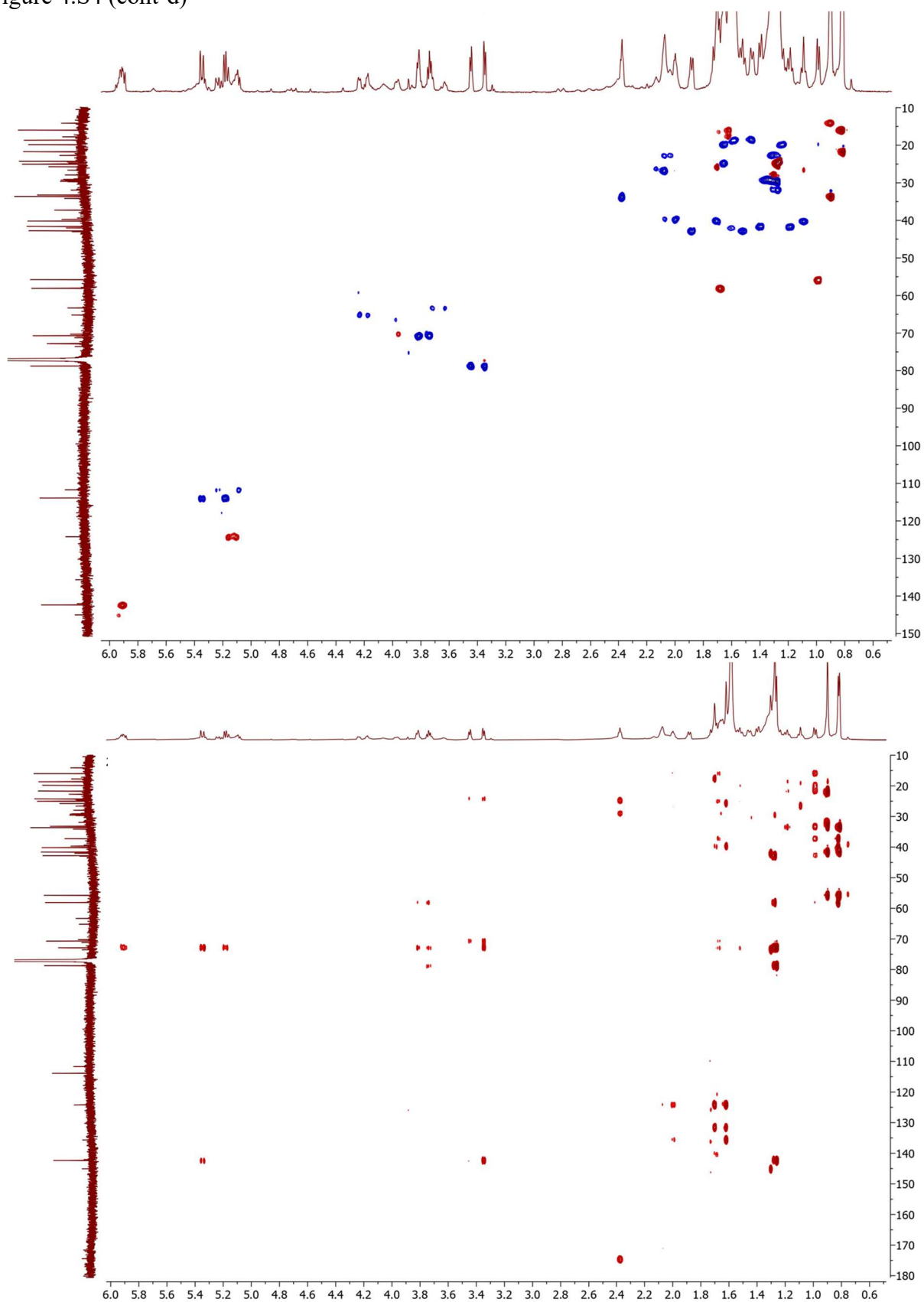


Figure 4.S4 (cont'd)

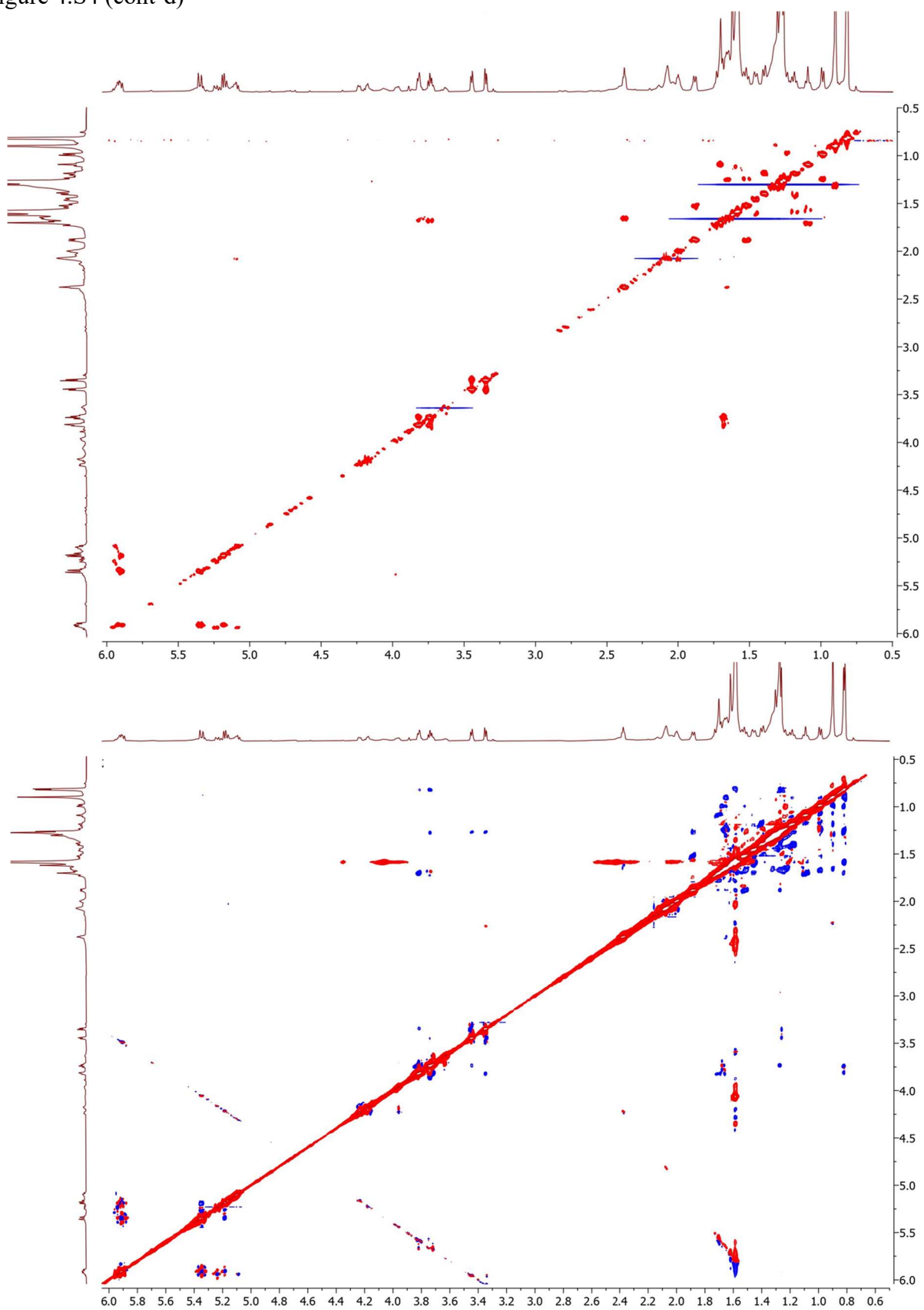
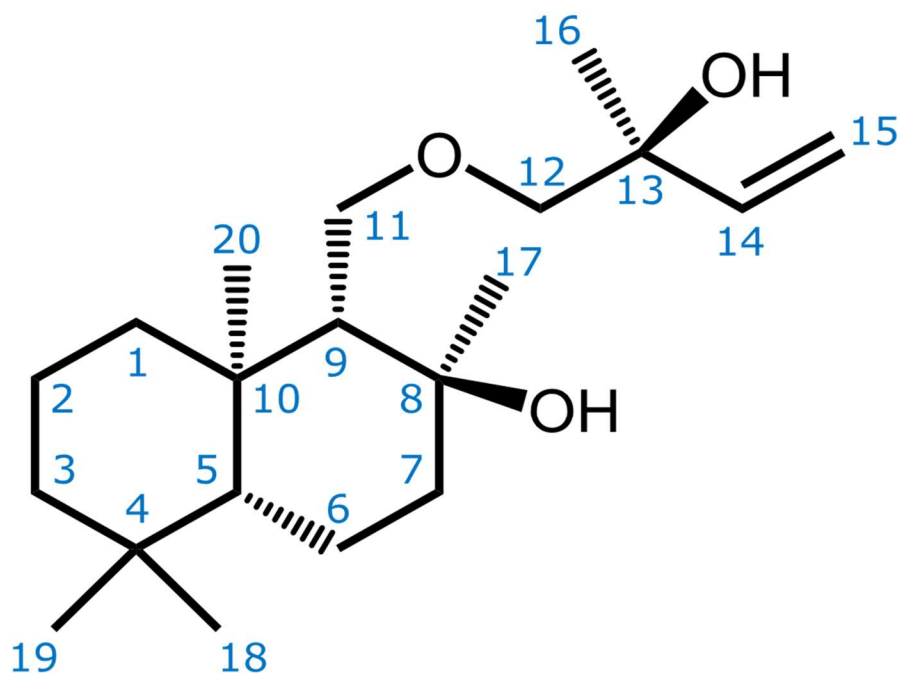


Table 4.S5: ^1H and ^{13}C chemical shifts for ent-11-oxo-copal-8,13-ol (**53**). CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Stereochemistry as drawn is based on the native mechanisms of *Sh*TPS1 and *Ss*SS.

^{13}C NMR	δ	^1H NMR
C1	40.2	1.09 (t, $J = 11.4$ Hz, 1H); 1.71 (m, 1H)
C2	18.6	1.46 (m, 1H); 1.58 (m, 1H)
C3	41.6	1.19 (m, 1H); 1.40 (m, 1H)
C4	33.3	-
C5	55.8	0.99 (d, $J = 12.0$ Hz, 2H)
C6	19.9	1.24 (m, 1H); 1.65 (m, 1H)
C7	42.8	1.52 (m, 1H); 1.88 (m, 1H)
C8	72.8	-
C9	58.1	1.68 (m, 1H)
C10	37.2	-
C11	70.7	3.73 (t, $J = 10.0$ Hz, 1H); 3.81 (d, $J = 11.9$ Hz, 1H)
C12	78.8	3.35 (d, $J = 9.1$ Hz, 1H); 3.45 (d, $J = 9.2$ Hz, 1H)
C13	72.8	-
C14	142.4	5.91 (dd, $J = 17.3, 11.0$ Hz, 1H)
C15	114.0	5.19 (d, $J = 11.5$ Hz, 1H); 5.35 (d, $J = 17.4$ Hz, 1H)
C16	24.3	1.26 (s, 3H)
C17	25.1	1.27 (s, 3H)
C18	33.6	0.90 (s, 3H)
C19	21.7	0.81 (s, 3H)
C20	16.0	0.82 (s, 3H)



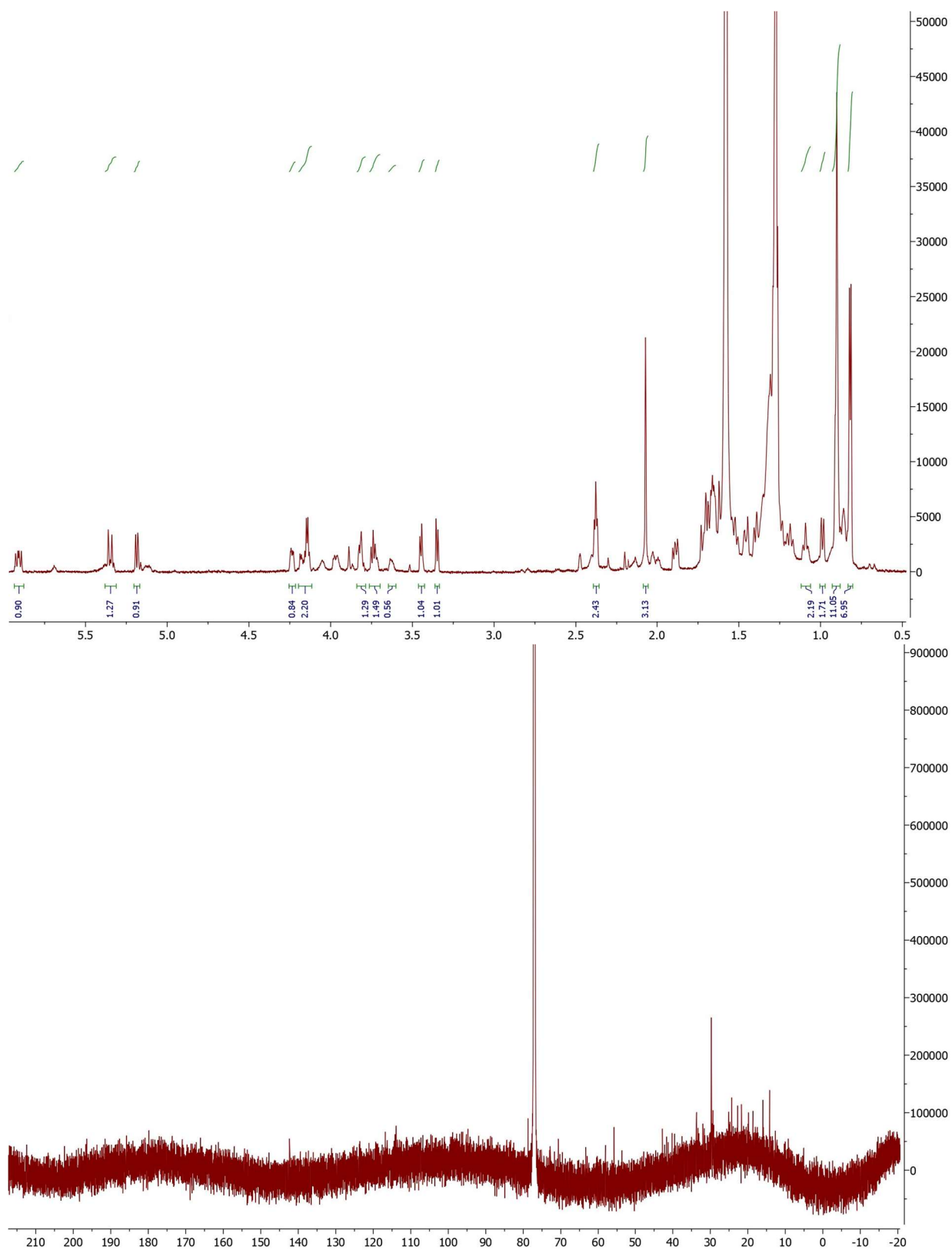


Figure 4.S5: NMR spectra (^1H , ^{13}C , HSQC, H2BC, HMBC, NOESY, COSY) for ent-11-oxo-copal-8,13-ol (53).

Figure 4.S5 (cont'd)

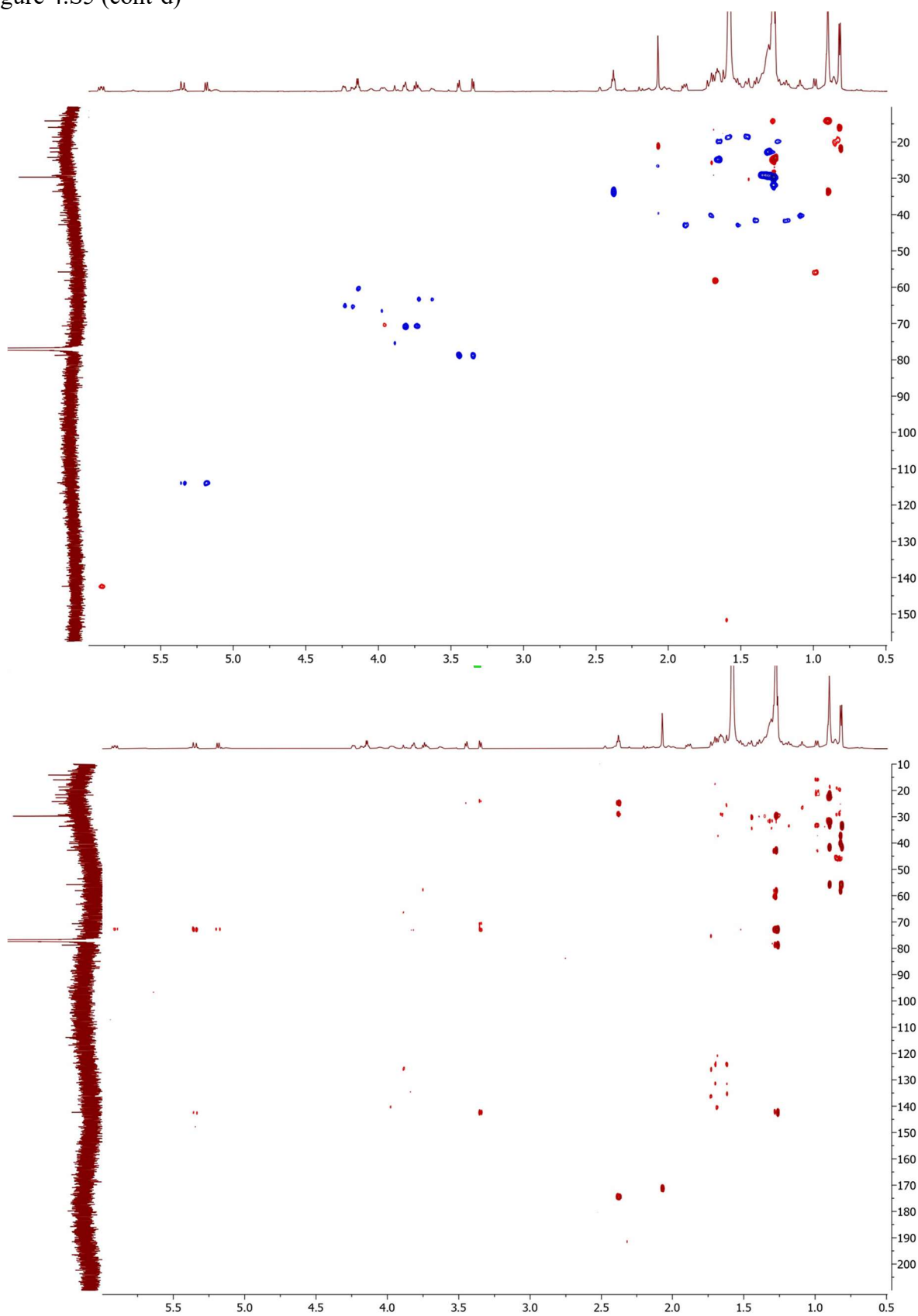


Figure 4.S5 (cont'd)

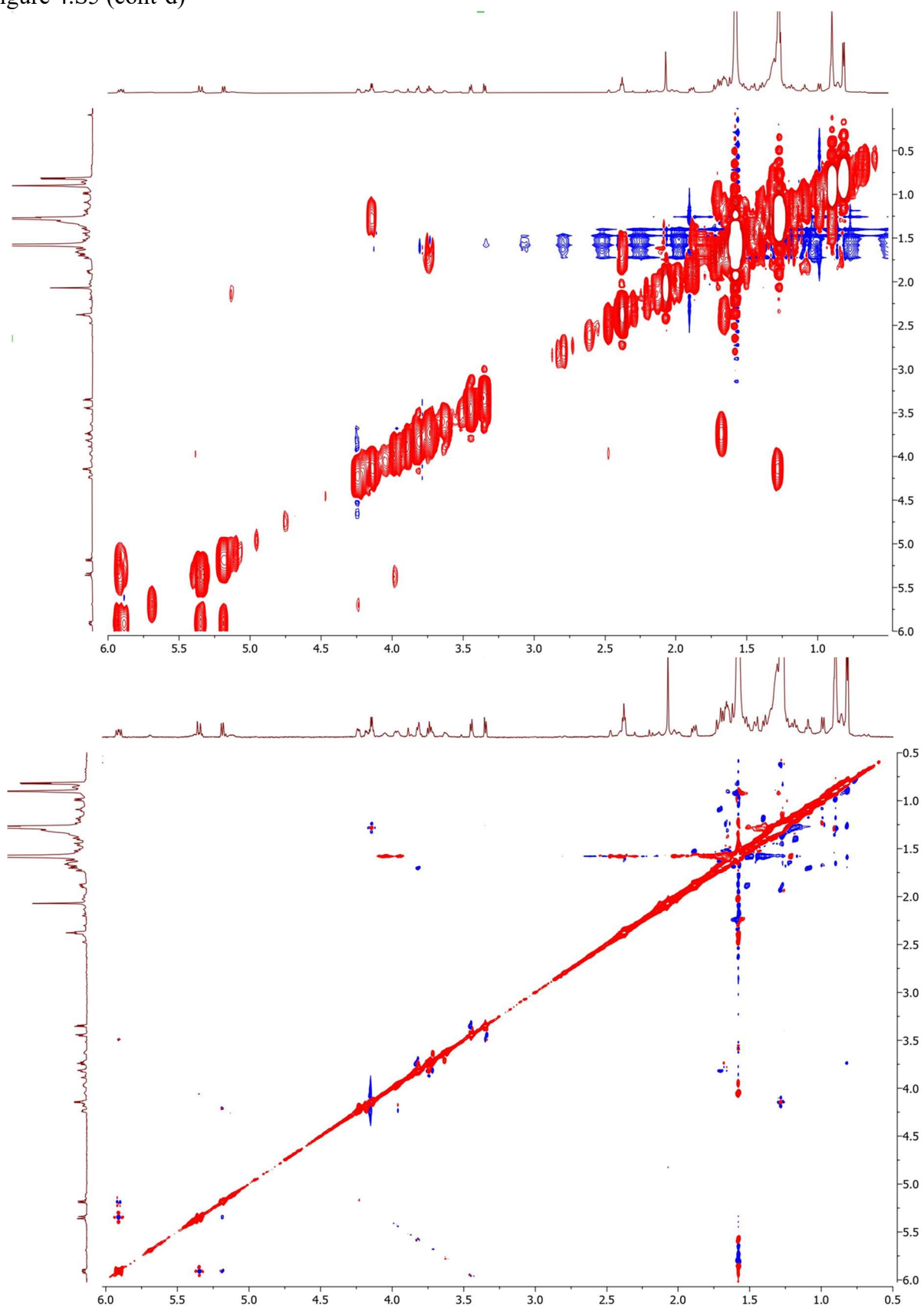
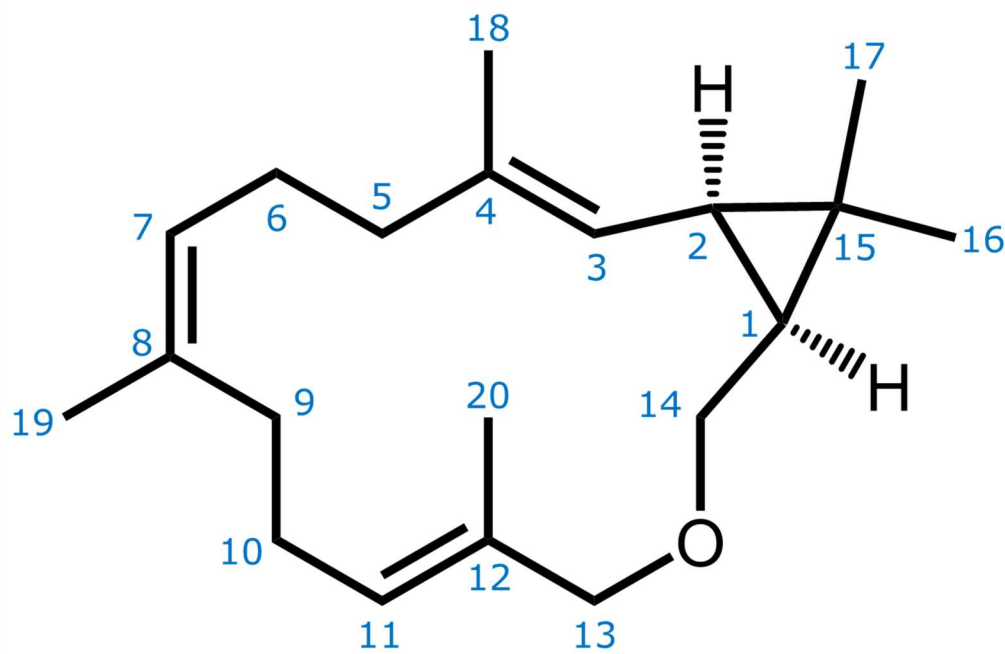


Table 4.S6: ^1H and ^{13}C chemical shifts for 13-oxo-casbene (**61**). CDCl_3 peaks were referenced to 7.26 and 77.00 ppm for ^1H and ^{13}C spectra, respectively. Absolute stereochemistry as drawn is based on the native mechanism of *El*CAS.

^{13}C NMR	δ	^1H NMR
C1	27.6	1.07 (ddd, $J = 6.5, 8.8, 9.5$ Hz, 1H)
C2	25.5	1.35 (m, 1H)
C3	119.7	4.94 (m, 1H)
C4	138.0	-
C5	38.1	2.13 (m, 2H)
C6	24.4	2.08 (m, 1H); 2.16 (m, 1H)
C7	125.3	4.98 (m, 1H)
C8	133.6	-
C9	39.2	2.08 (m, 1H); 2.15 (m, 1H)
C10	24.6	2.11 (m, 1H); 2.14 (m, 1H)
C11	128.7	5.29 (m, 1H)
C12	131.7	-
C13	77.3	3.74 (d, $J = 12.4$ Hz, 1H); 3.94 (d, $J = 12.3$ Hz, 1H)
C14	66.6	3.18 (d, $J = 6.4$ Hz, 1H); 3.21 (d, $J = 8.9$ Hz, 1H)
C15	19.8	-
C16	28.6	1.14 (s, 3H)
C17	15.6	1.02 (s, 3H)
C18	17.2	1.66 (s, 3H)
C19	15.6	1.61 (s, 3H)
C20	14.0	1.61 (s, 3H)



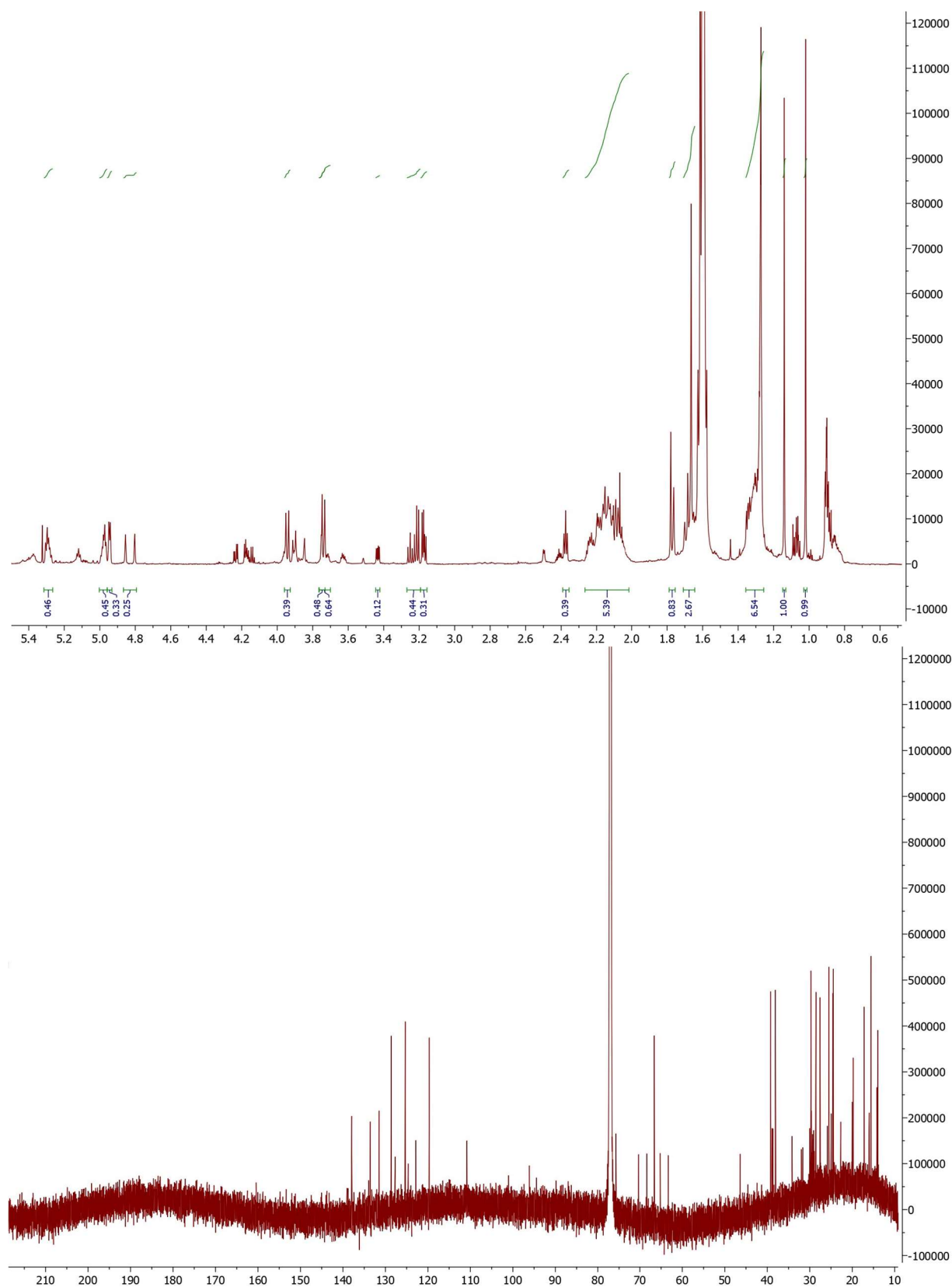


Figure 4.S6: NMR spectra (^1H , ^{13}C , HSQC, H2BC, HMBC, NOESY, COSY) for 13-oxo-casbene (61).

Figure 4.S6 (cont'd)

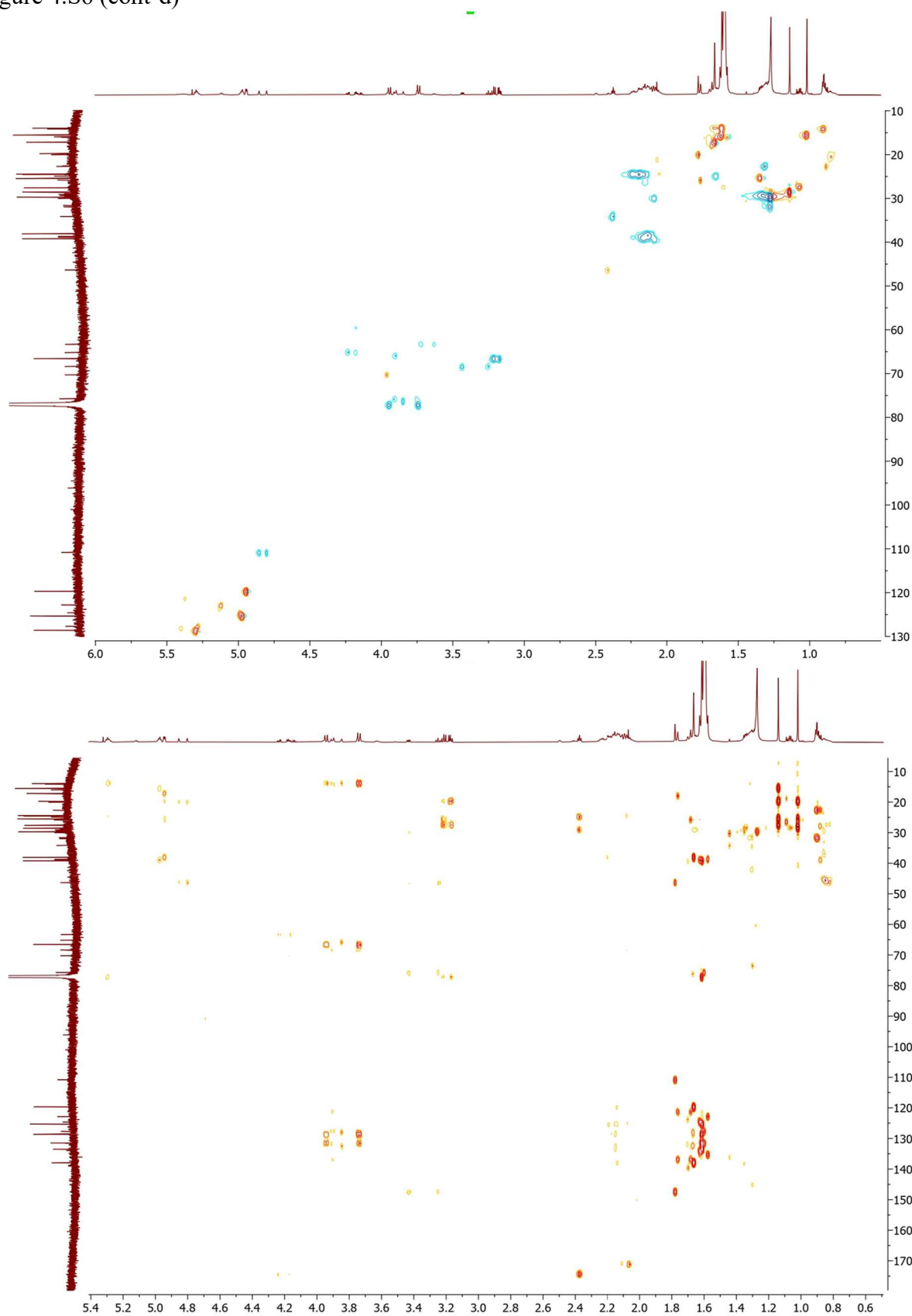
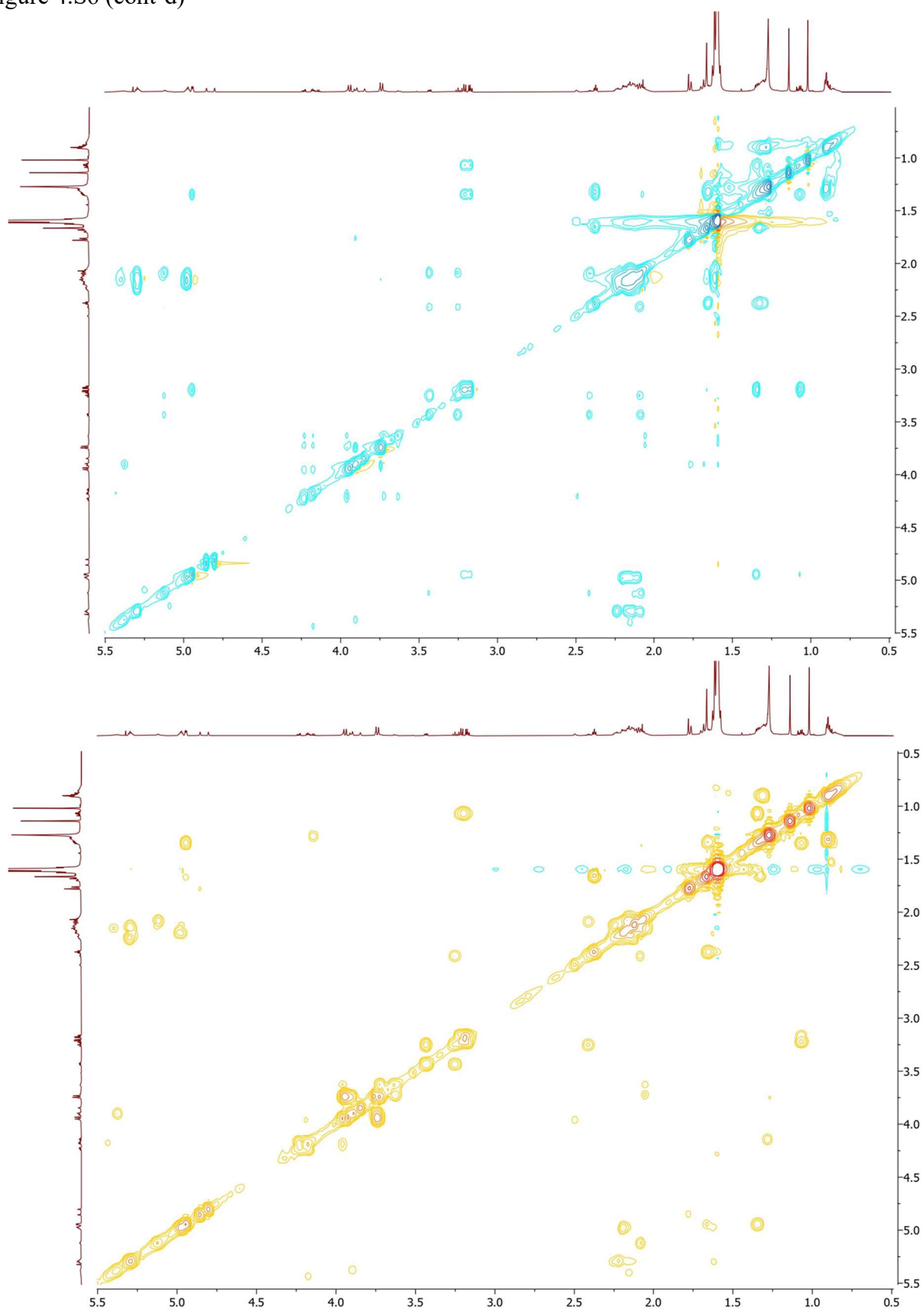


Figure 4.S6 (cont'd)



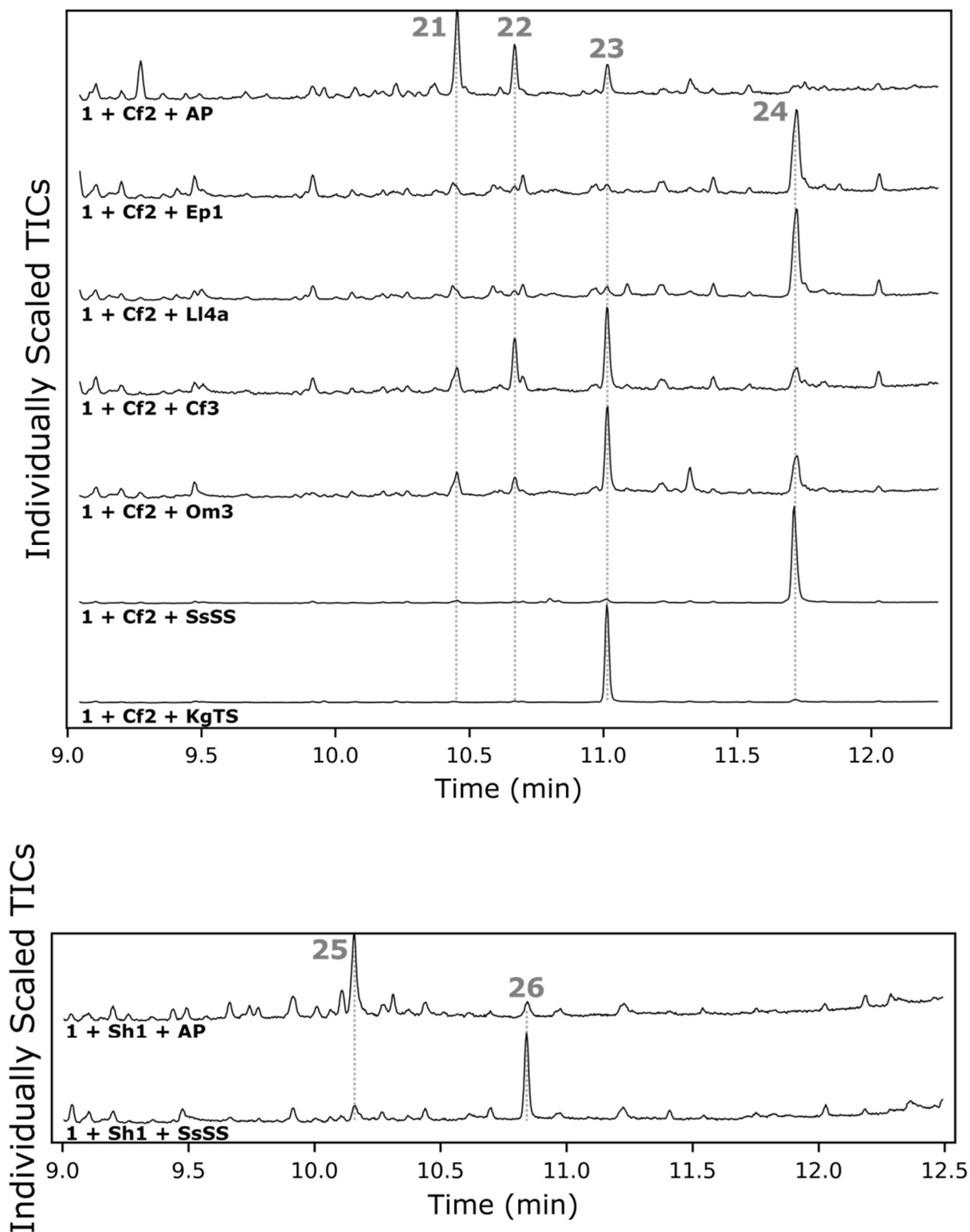


Figure 4.S7: GC-MS screening of class II/class I TPS combinations with hits from GC-FID screening. Mass spectra for numbered compounds are given in Figure 4.S10. Compounds labeled with a “b” are putative enantiomers of the same compound numbered without a letter.

Figure 4.S7 (cont'd)

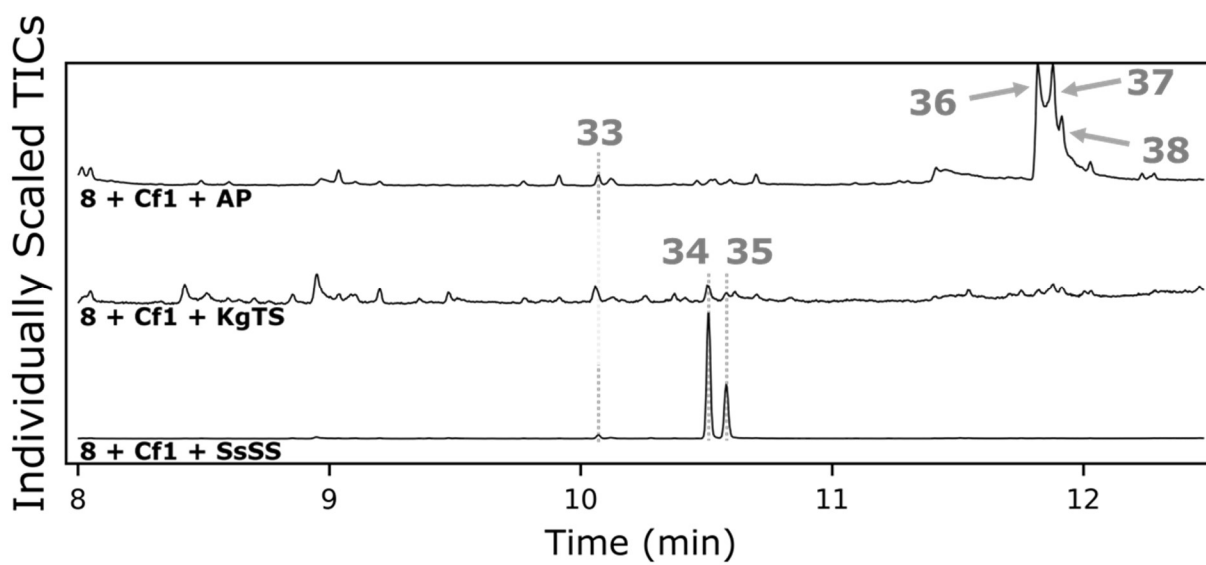
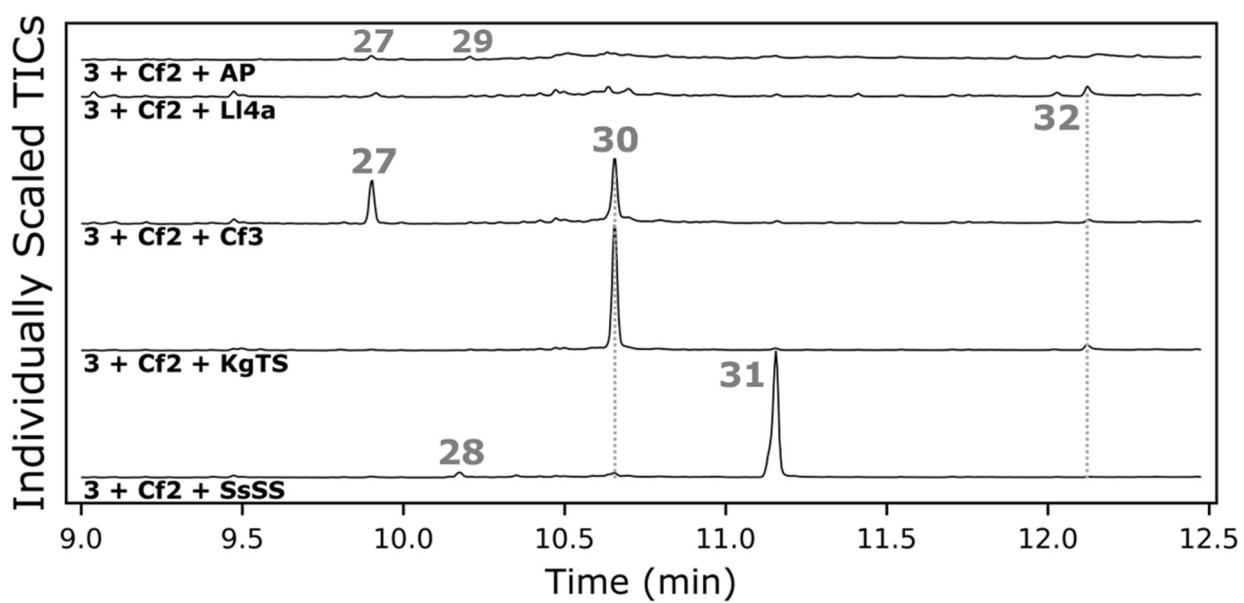
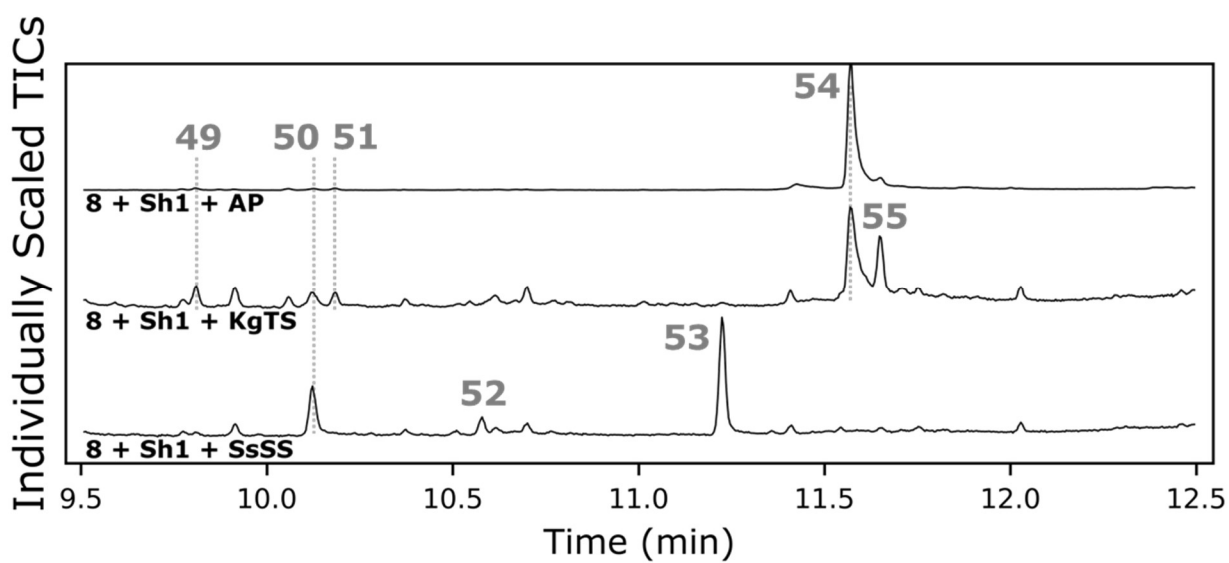
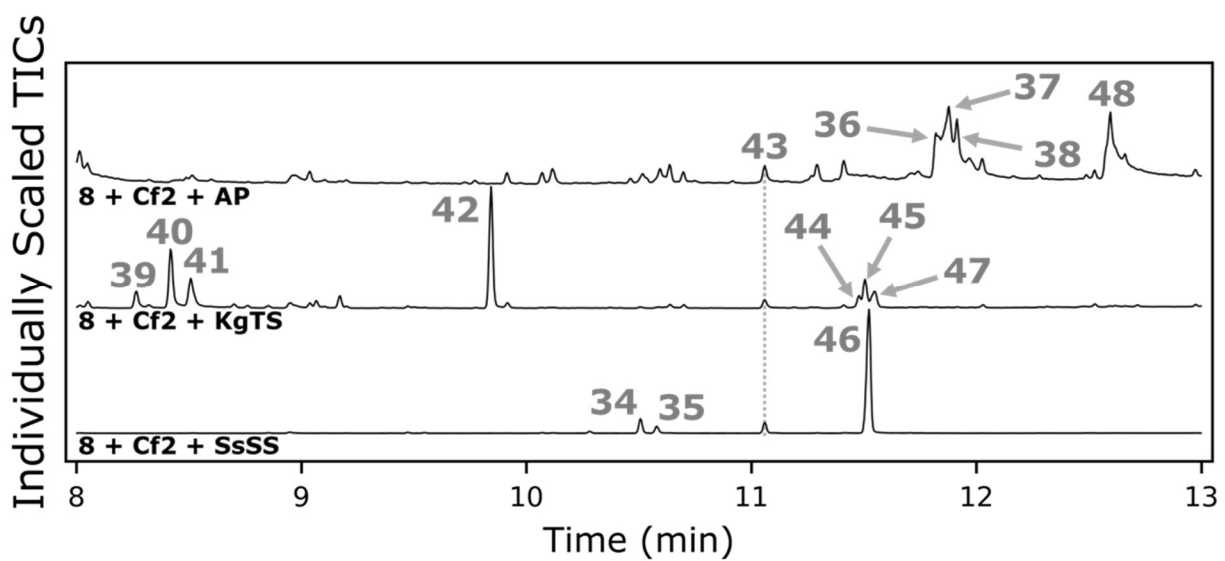


Figure 4.S7 (cont'd)



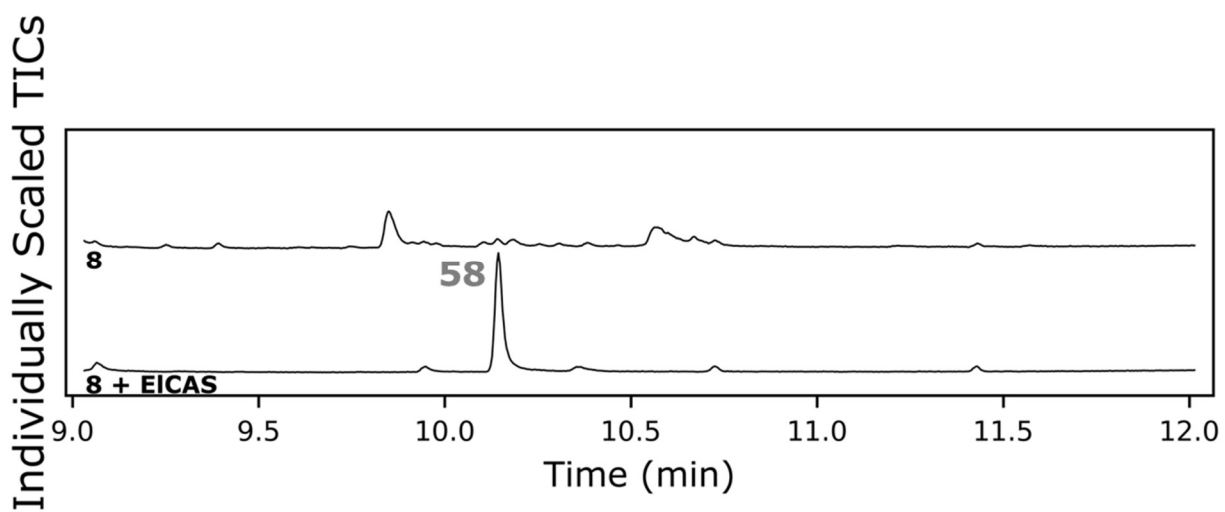
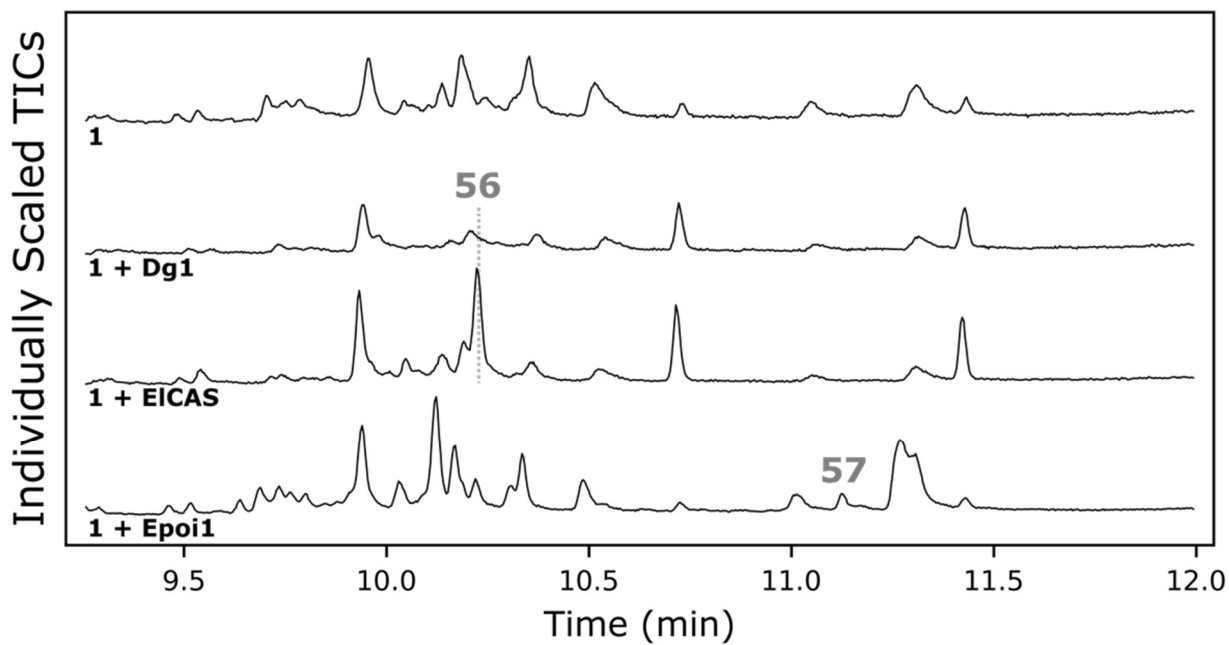


Figure 4.S8: GC-MS screening of single-step class I TPSs with hits from GC-FID screening. Mass spectra for numbered compounds are given in Figure 4.S10.

Figure 4.S8 (cont'd)

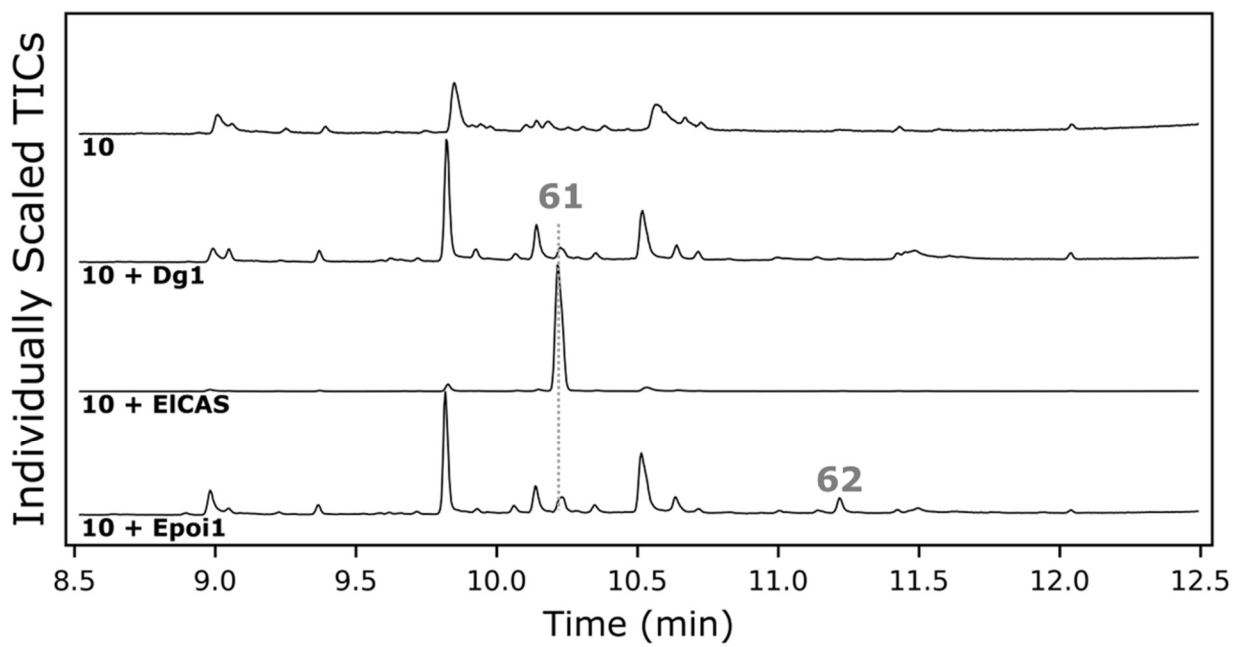
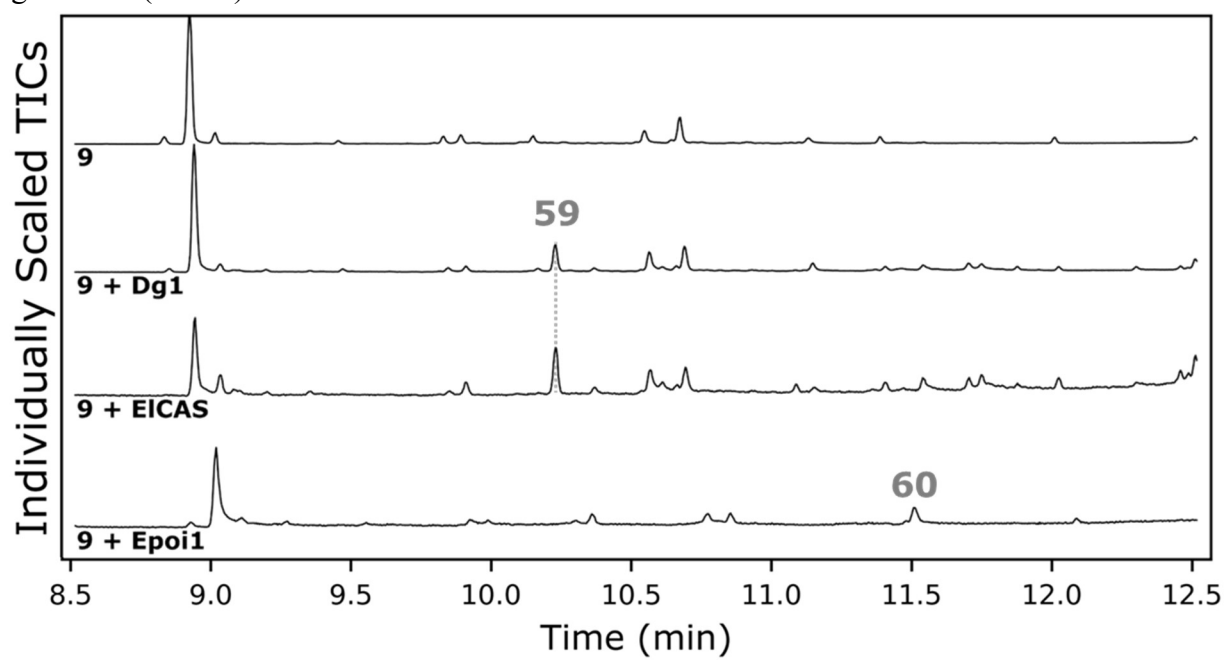
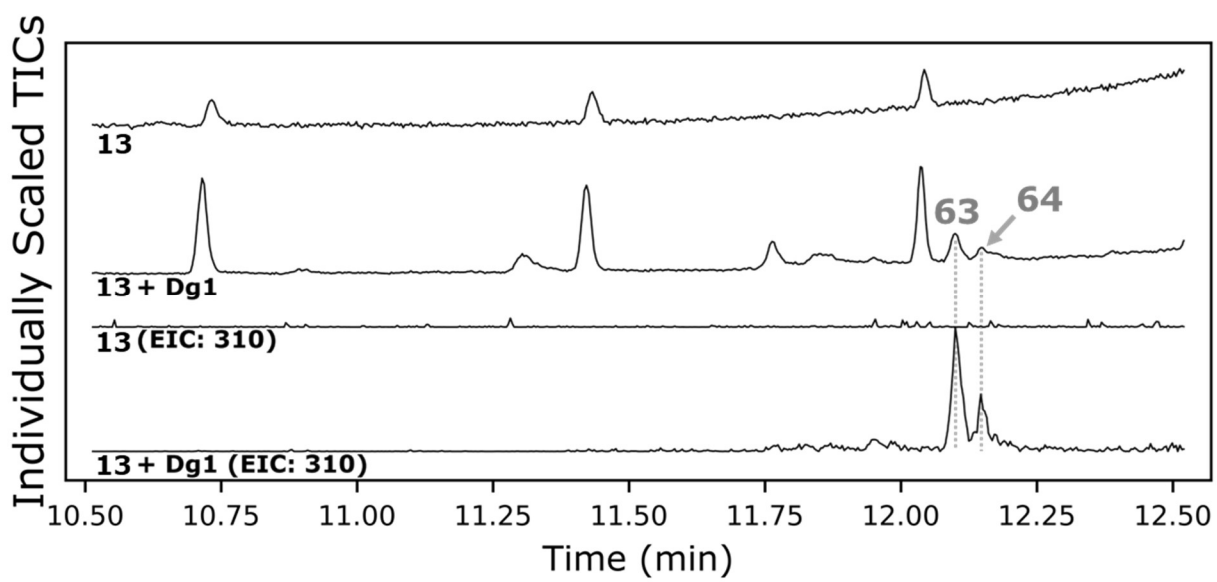


Figure 4.S8 (cont'd)



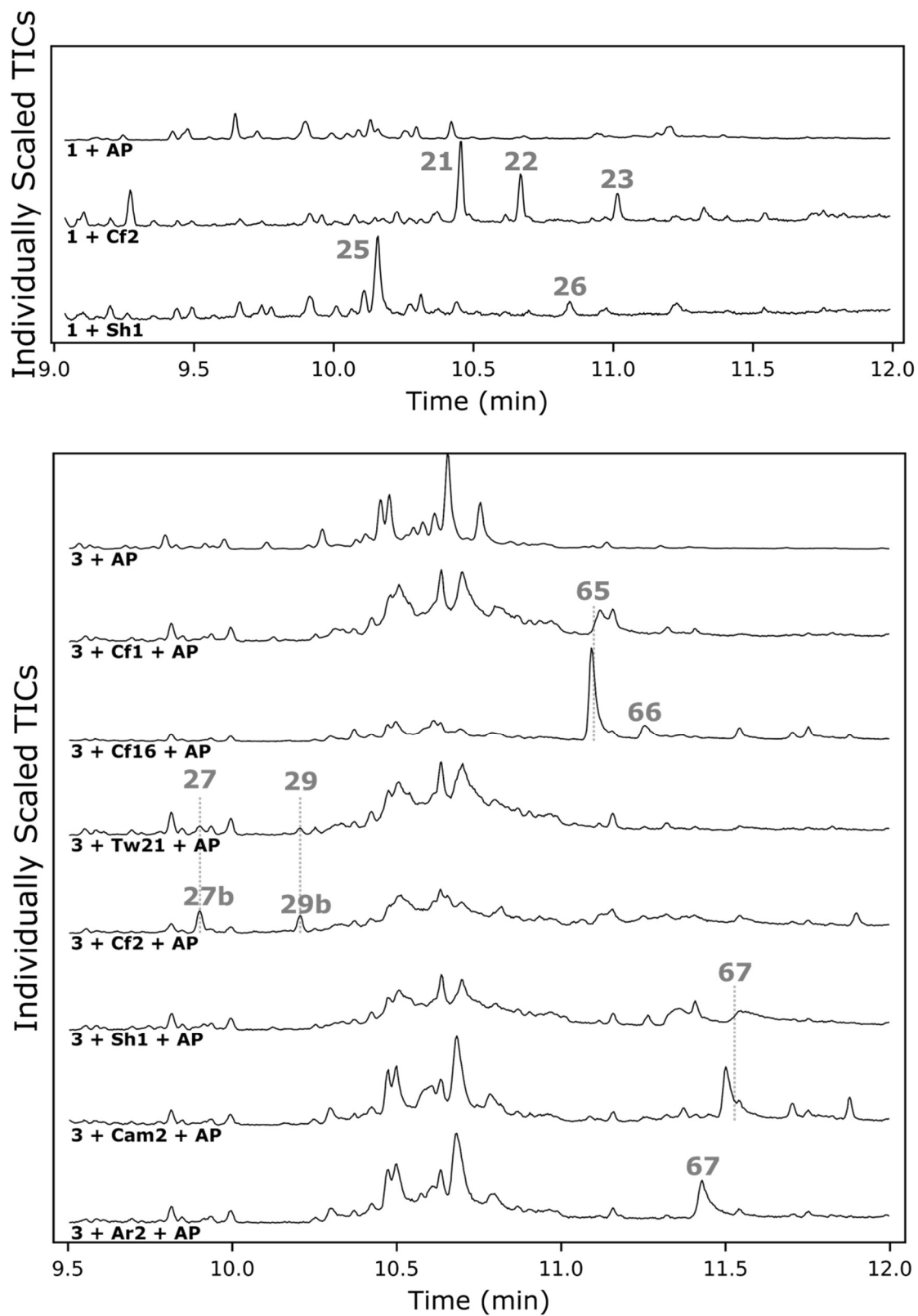


Figure 4.S9: GC-MS screening of class II TPSs with hits from GC-FID screening. Mass spectra for numbered compounds are given in Figure 4.S10.

Figure 4.S9 (cont'd)

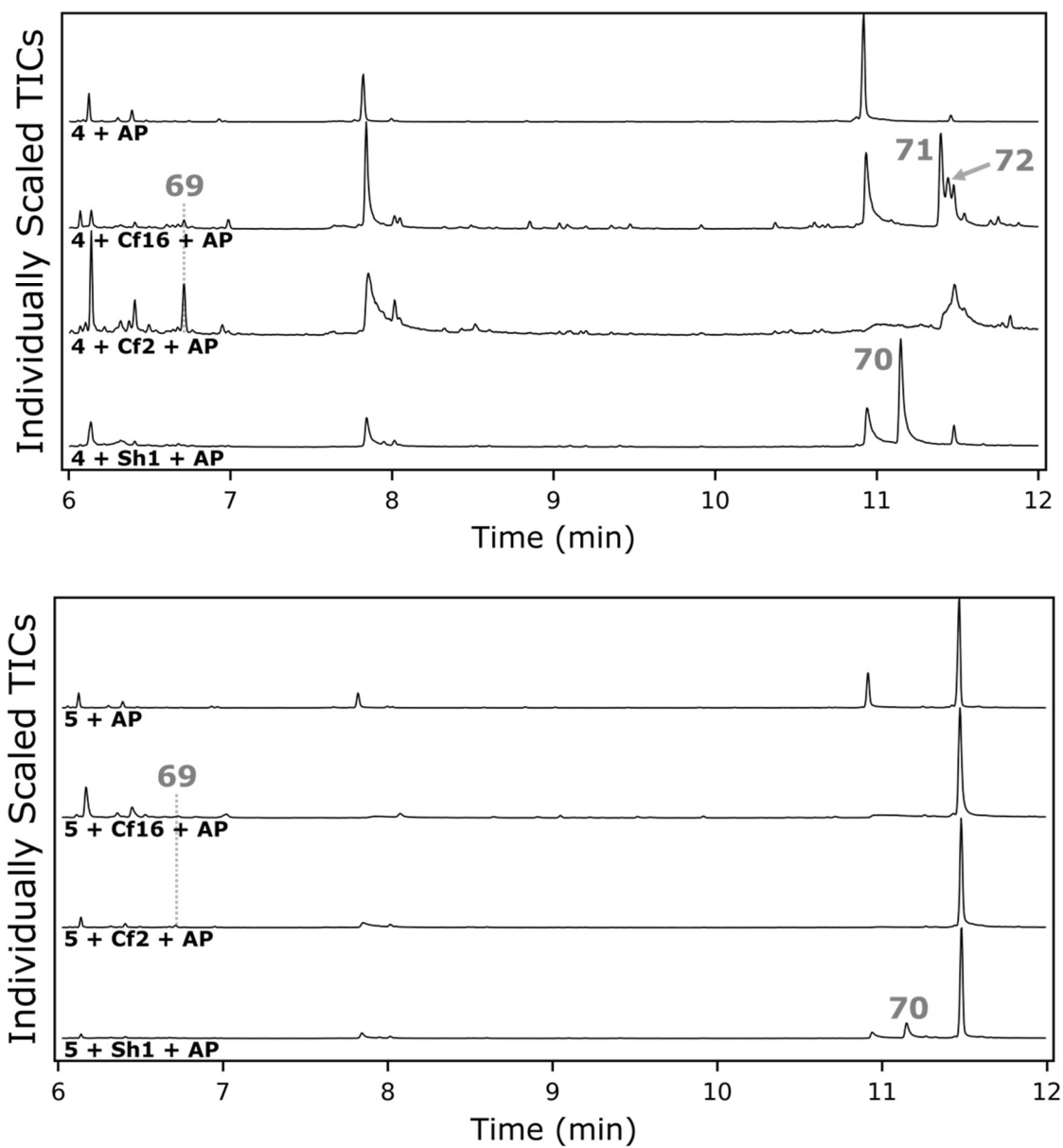


Figure 4.S9 (cont'd)

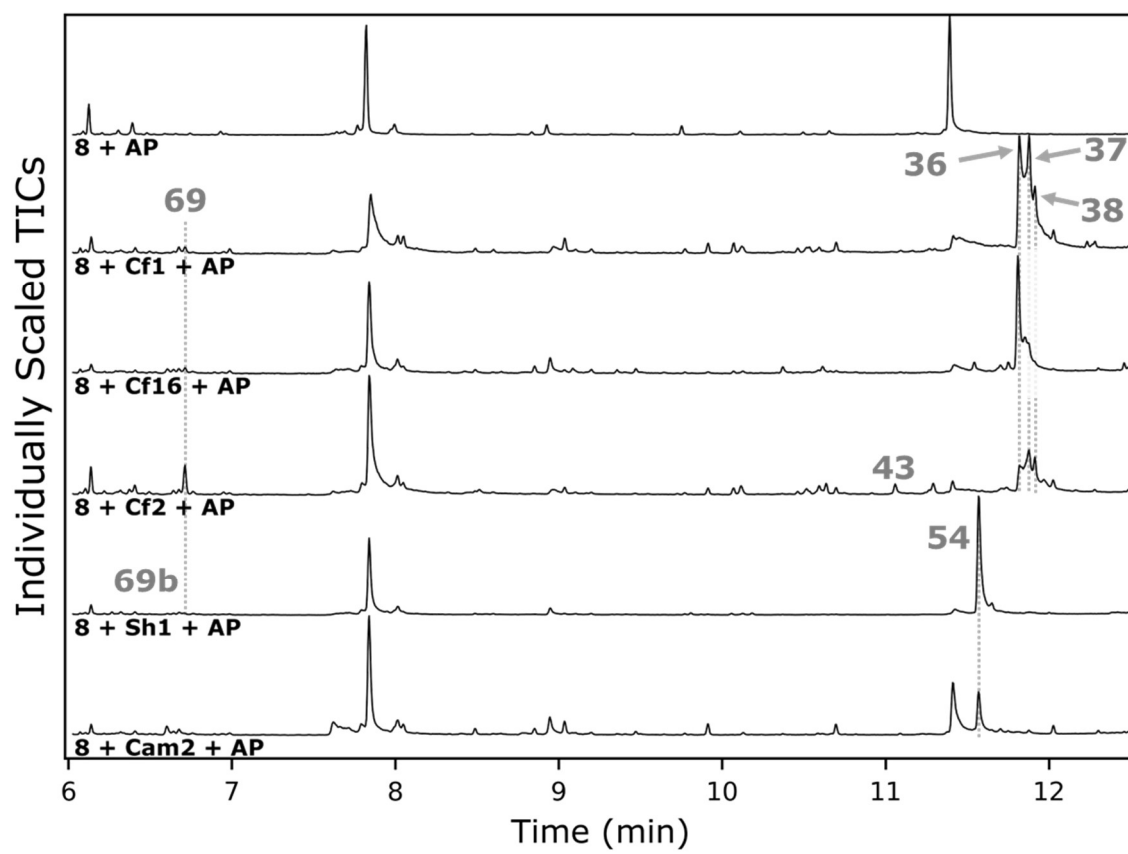
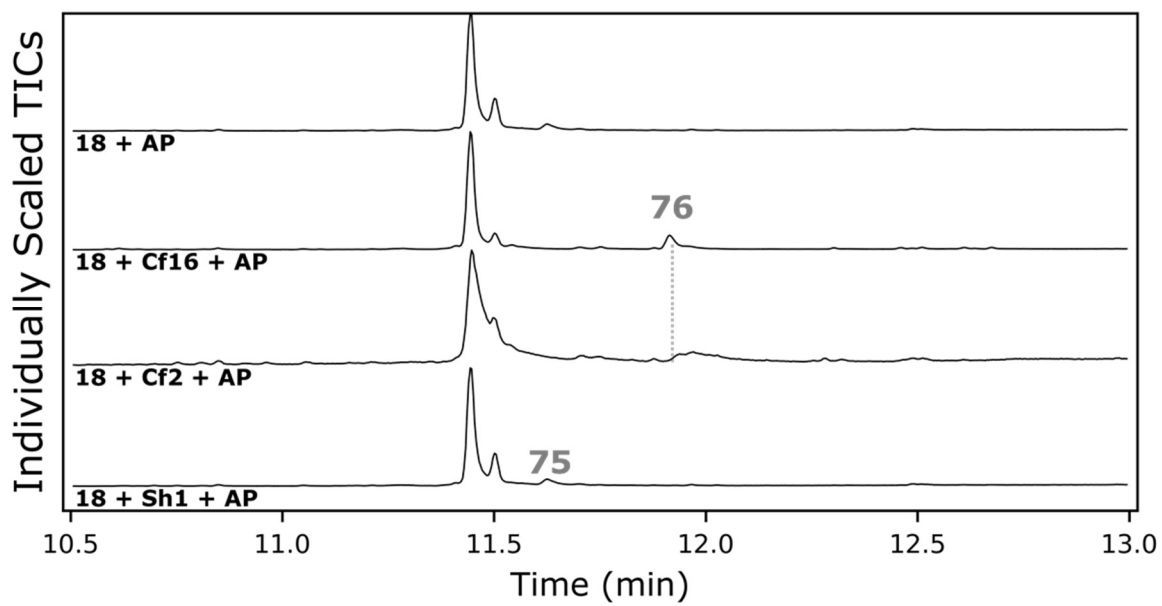
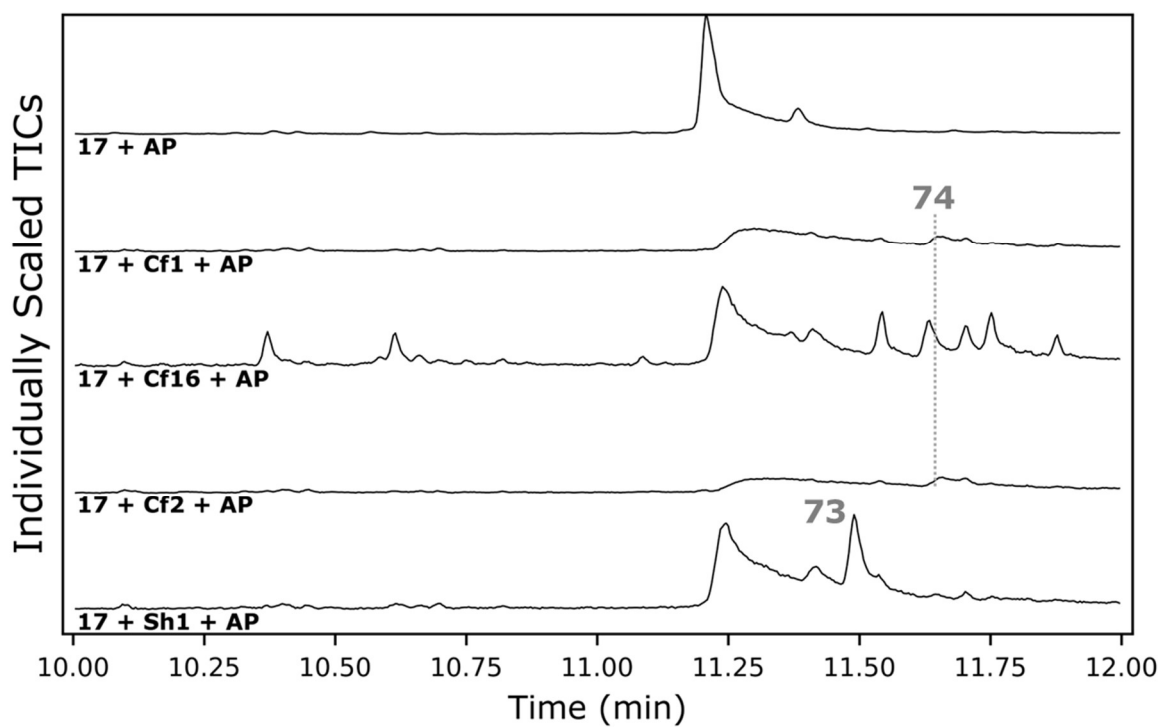


Figure 4.S9 (cont'd)



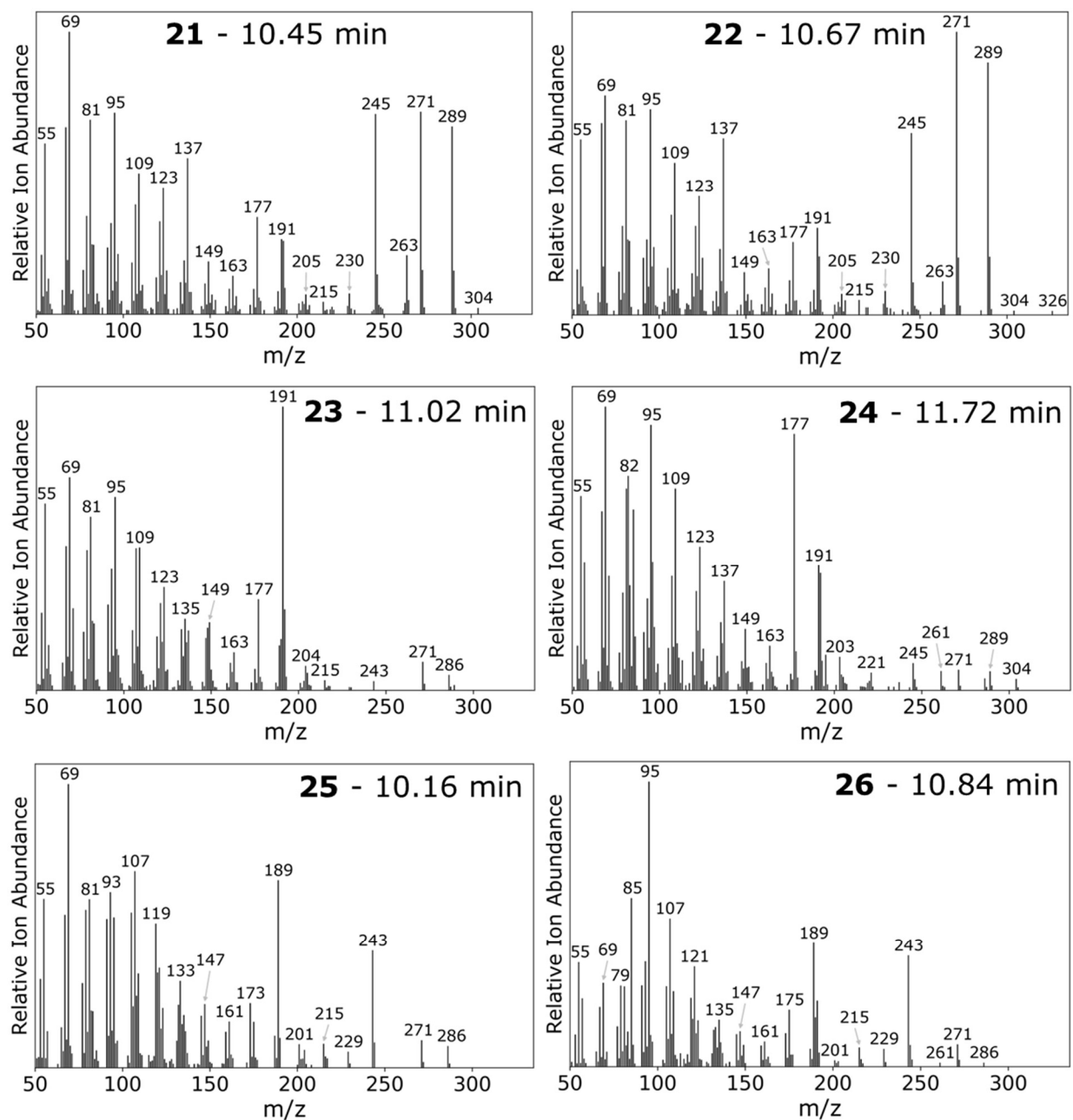


Figure 4.S10: Mass spectra for all compounds (21-74) shown in GC-MS screening. Retention time is listed next to compound number.

Figure 4.S10 (cont'd)

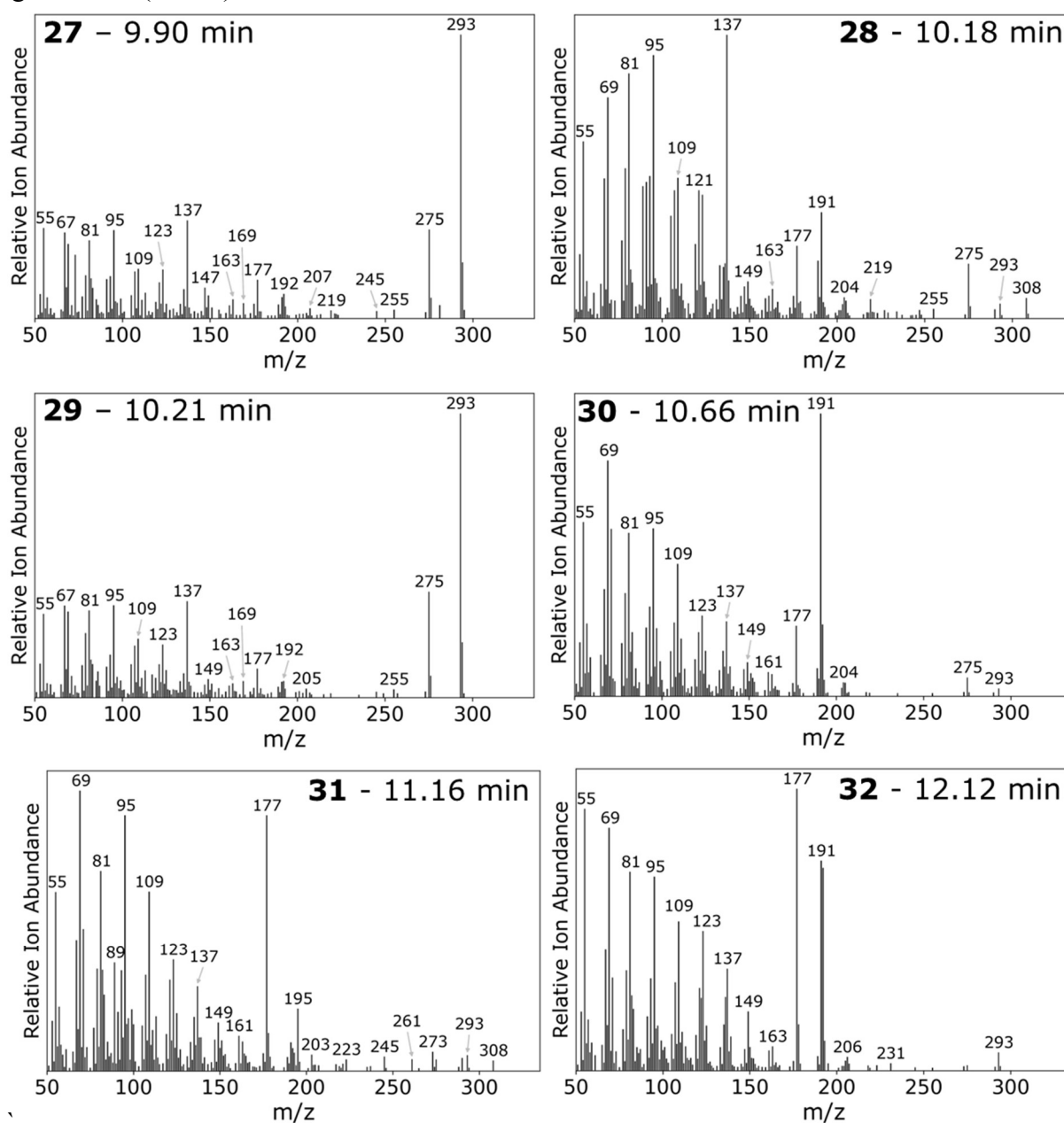


Figure 4.S10 (cont'd)

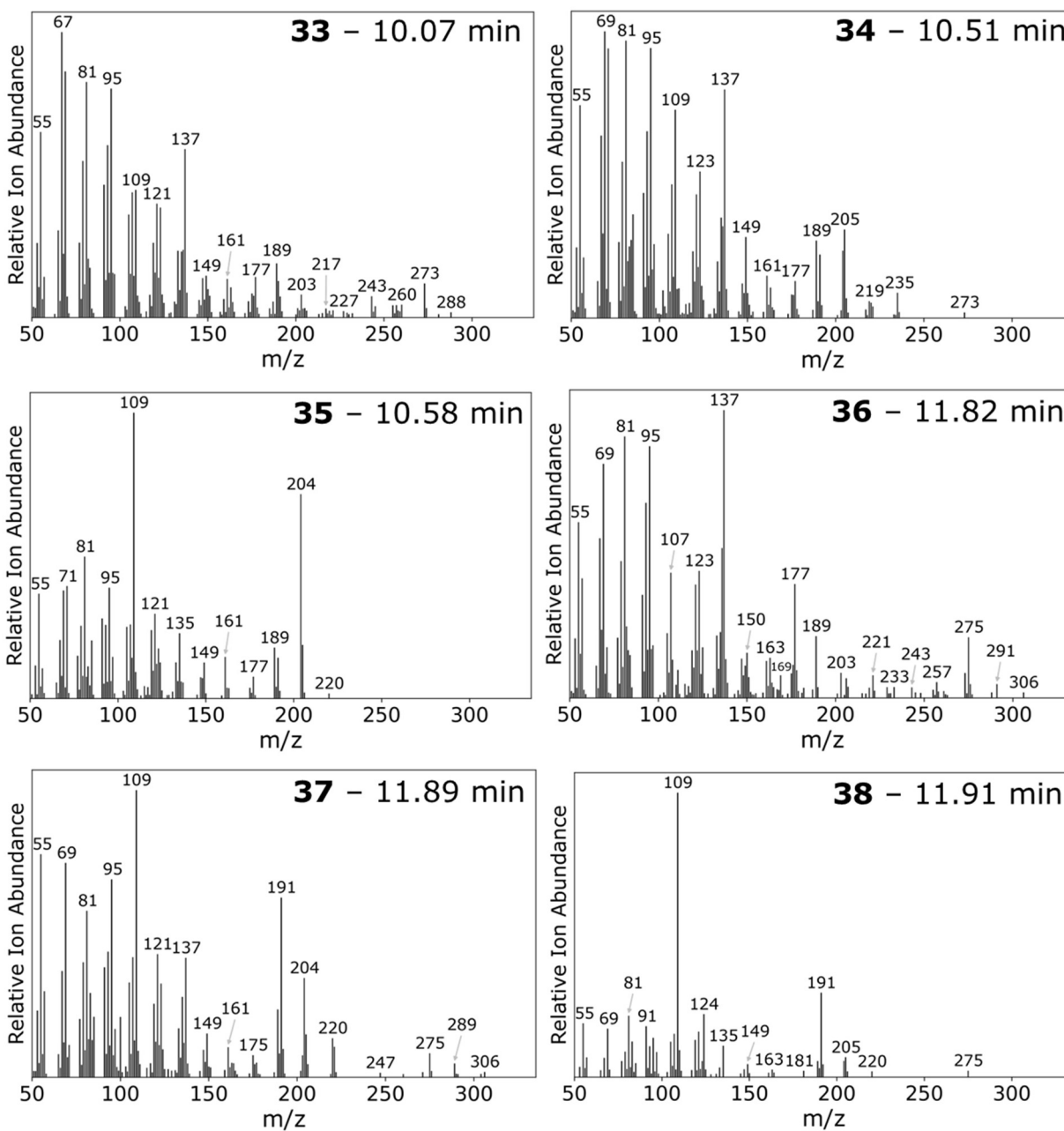


Figure 4.S10 (cont'd)

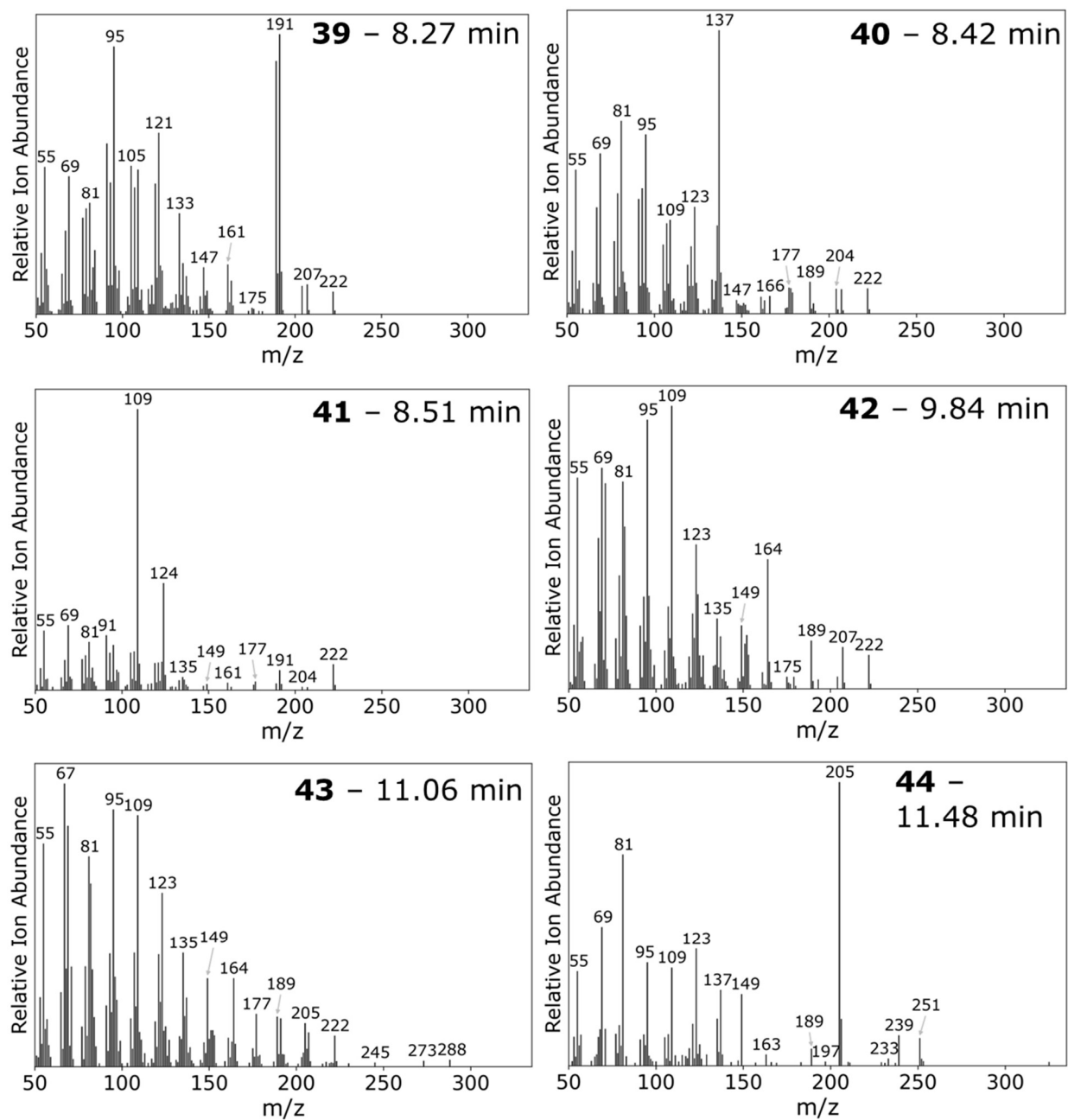


Figure 4.S10 (cont'd)

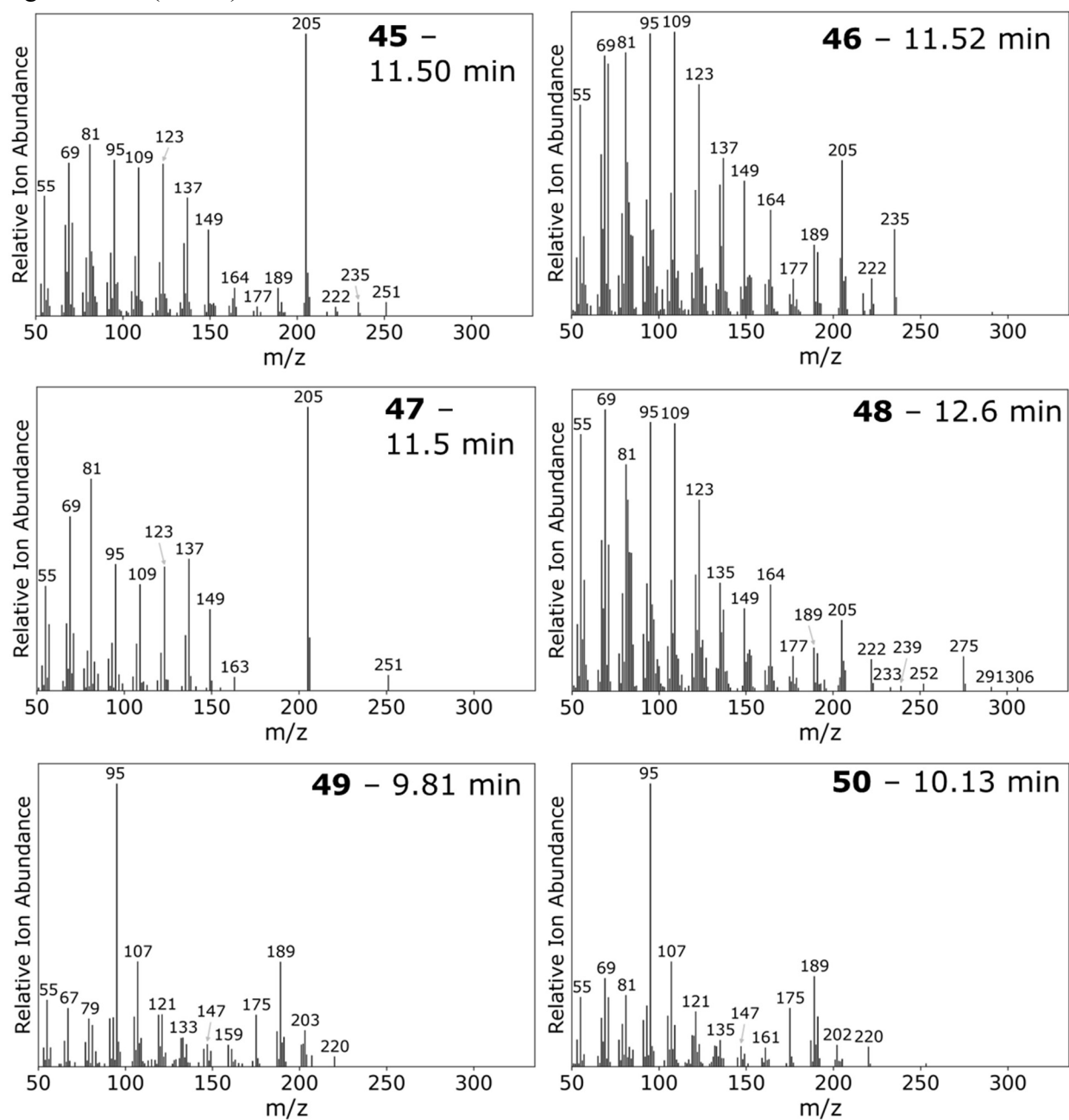


Figure 4.S10 (cont'd)

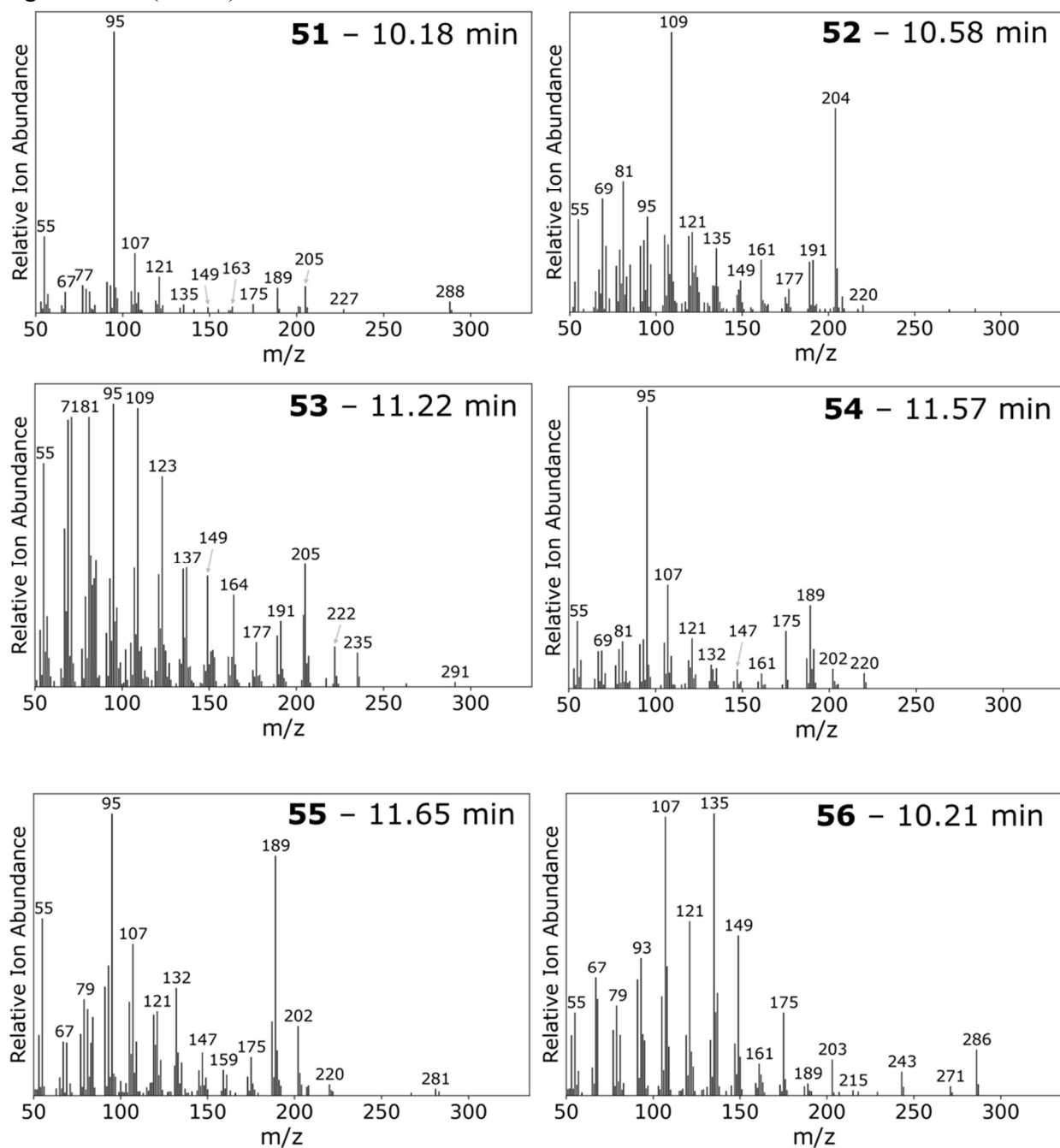


Figure 4.S10 (cont'd)

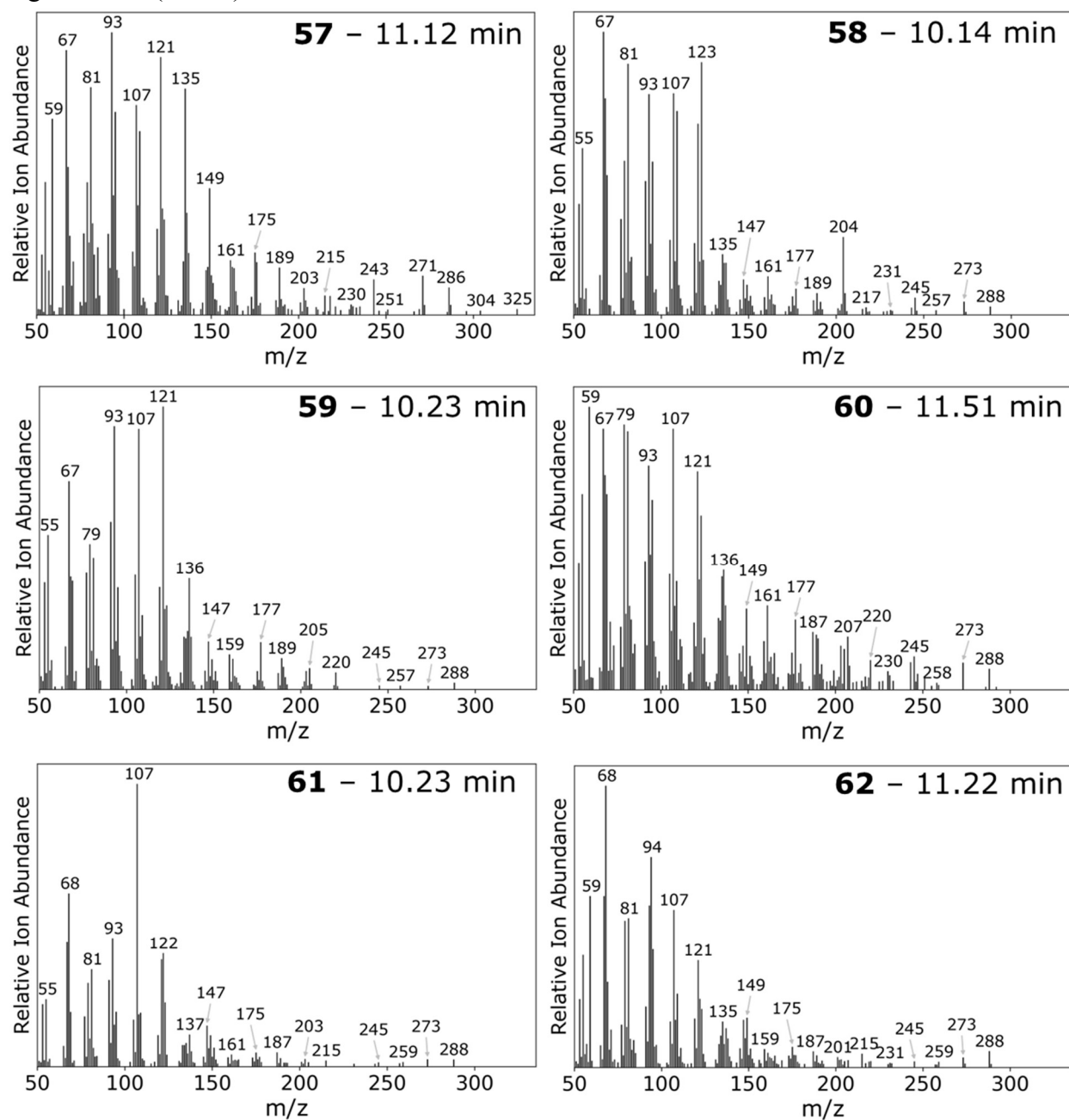


Figure 4.S10 (cont'd)

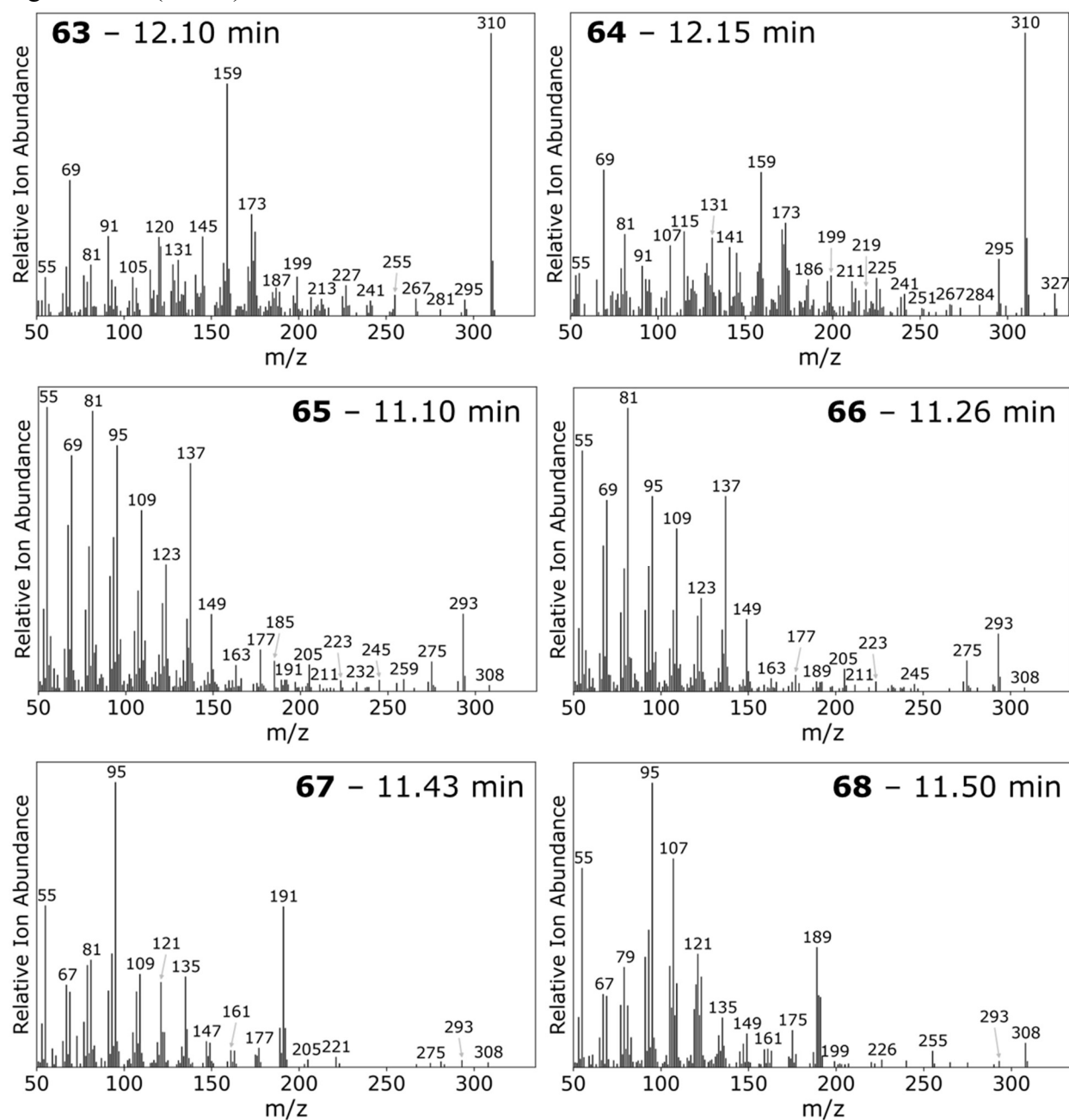


Figure 4.S10 (cont'd)

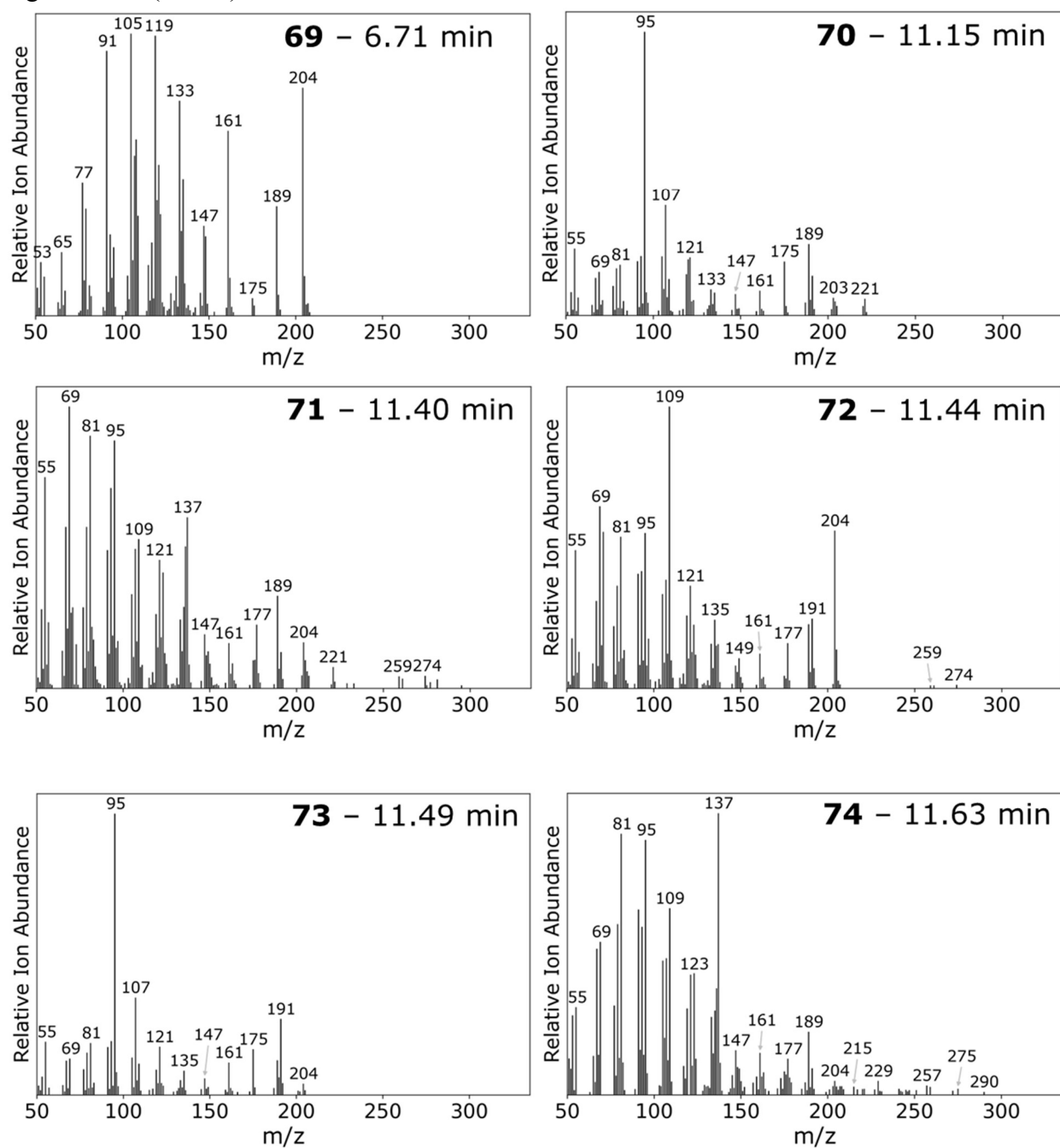
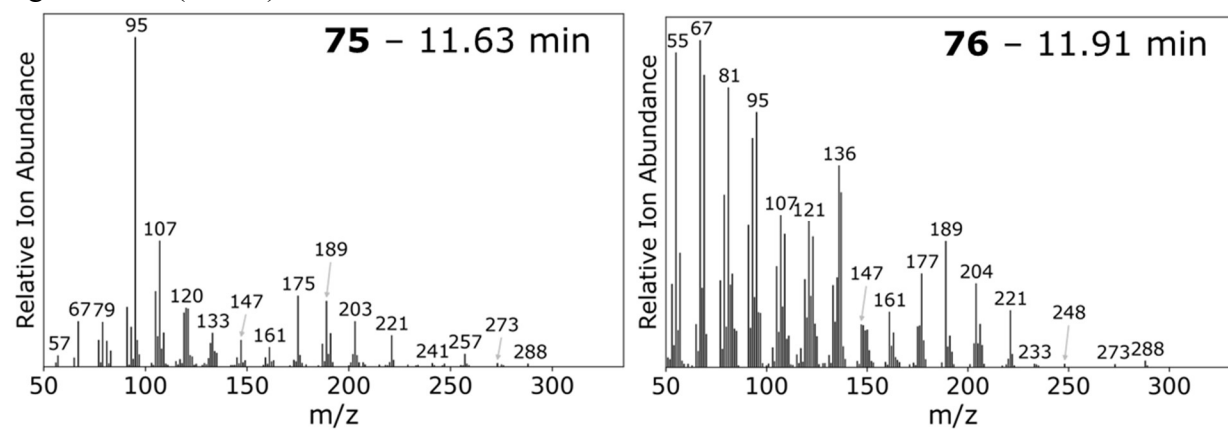


Figure 4.S10 (cont'd)



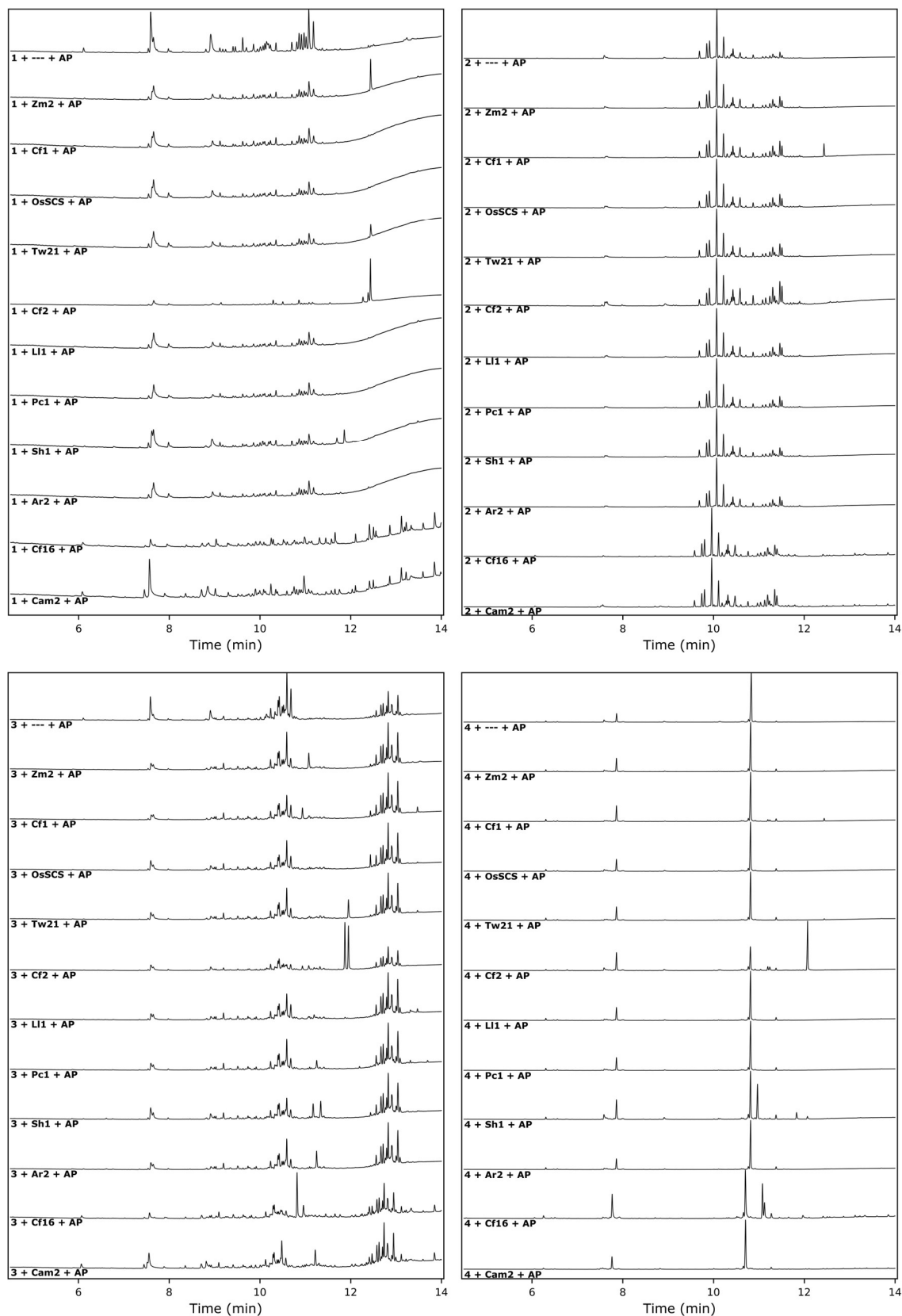


Figure 4.S11: Initial GC-FID screening of class II TPSs with substrates 1-20.

Figure 4.S11 (cont'd)

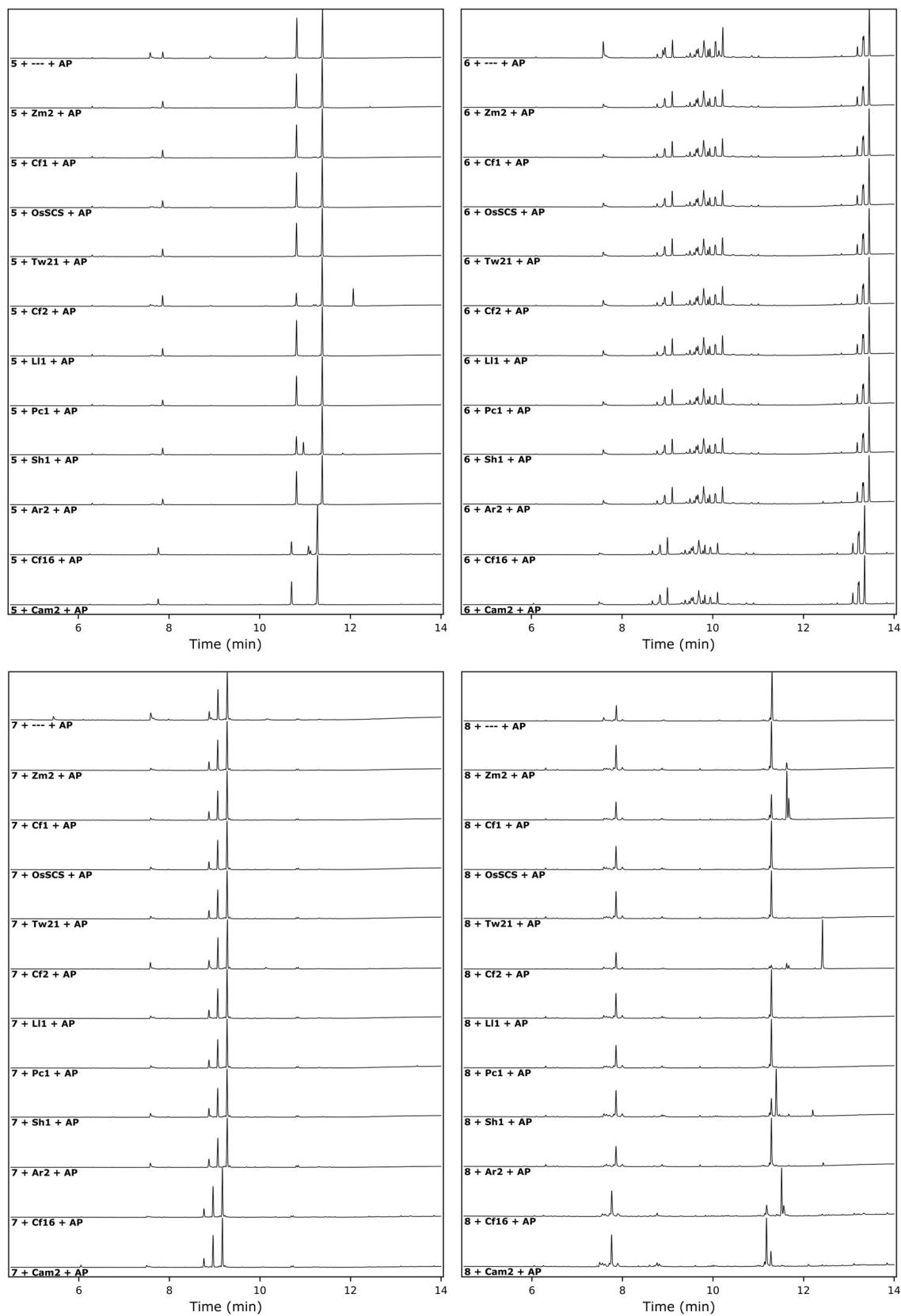


Figure 4.S11 (cont'd)

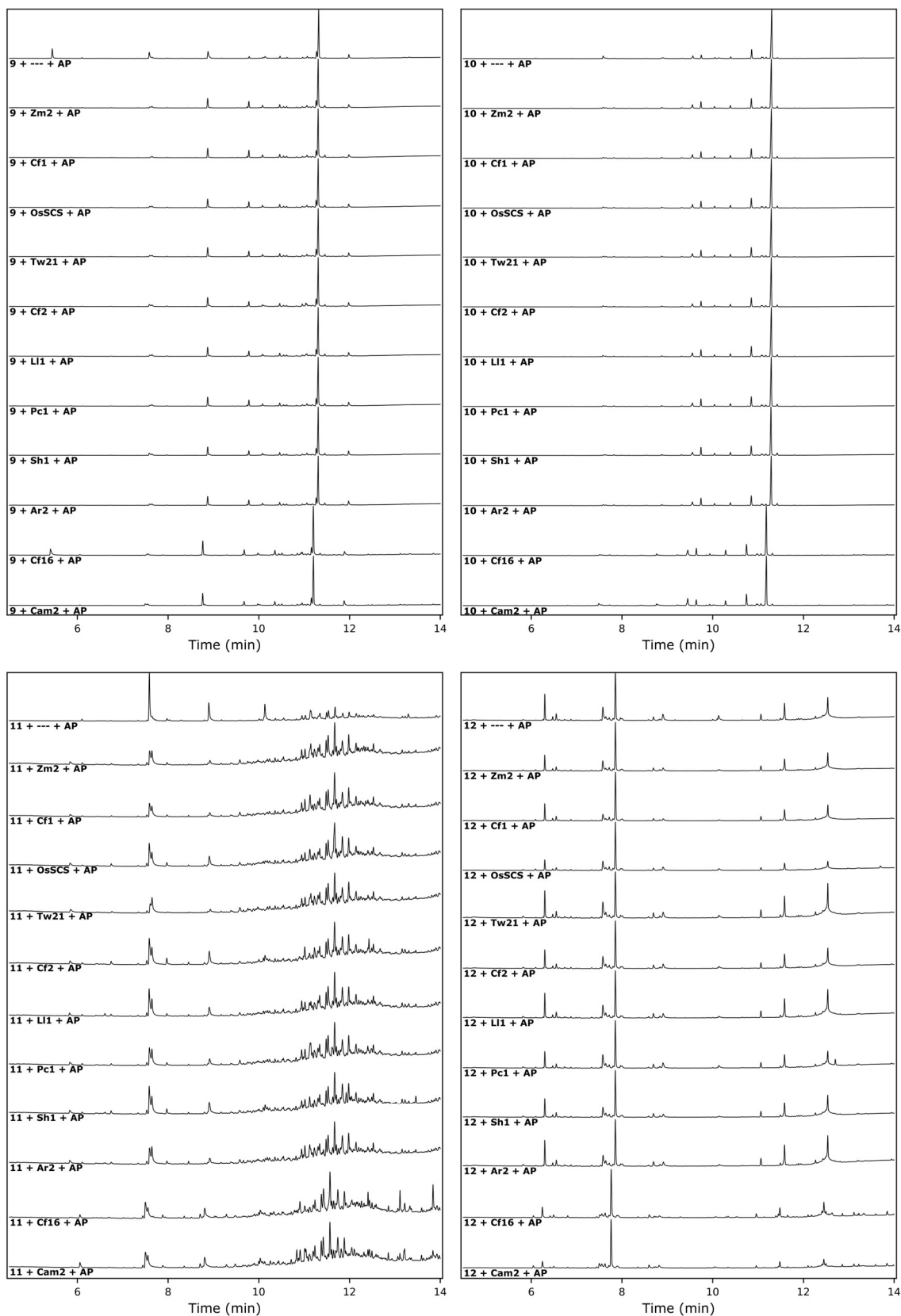


Figure 4.S11 (cont'd)

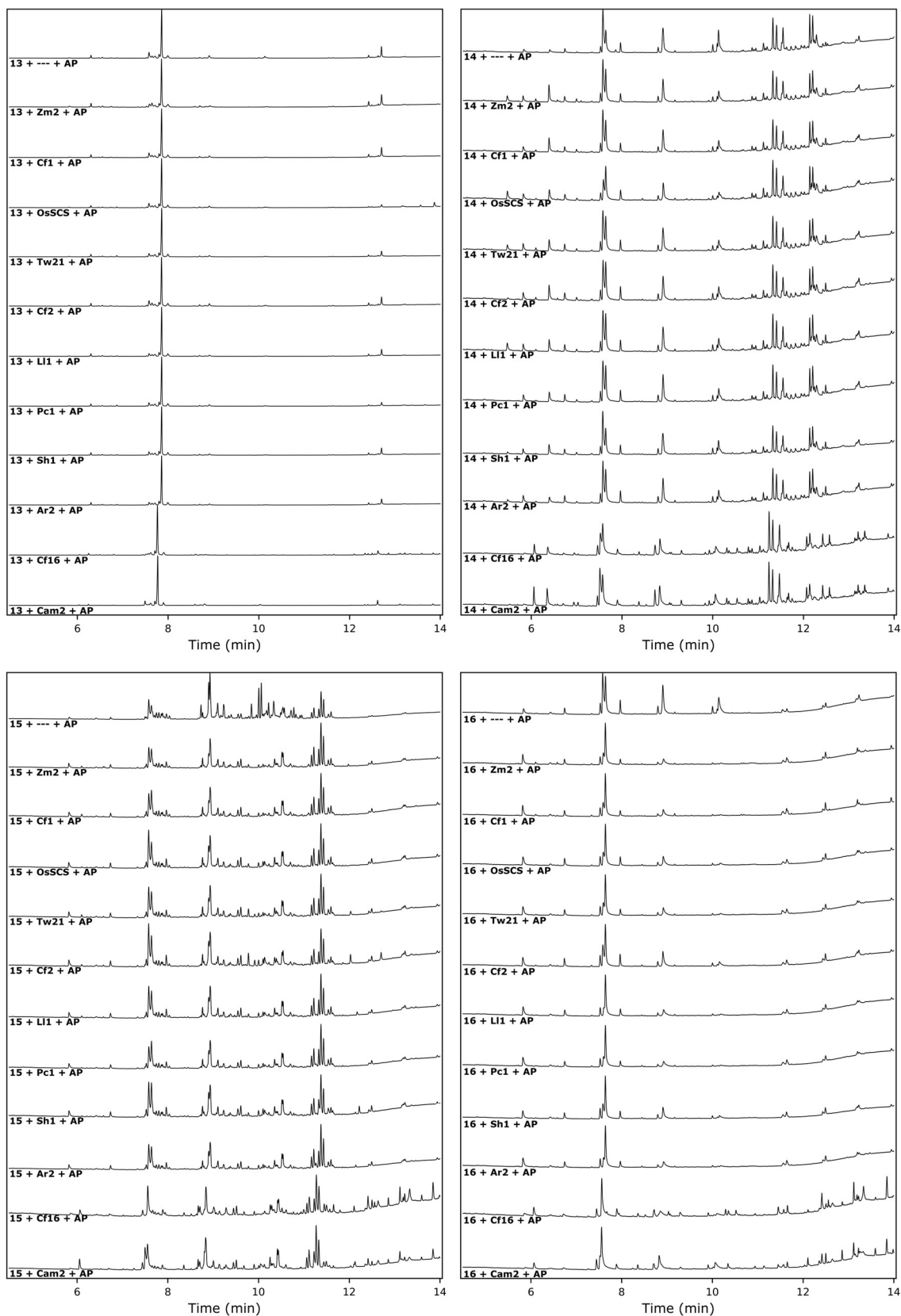
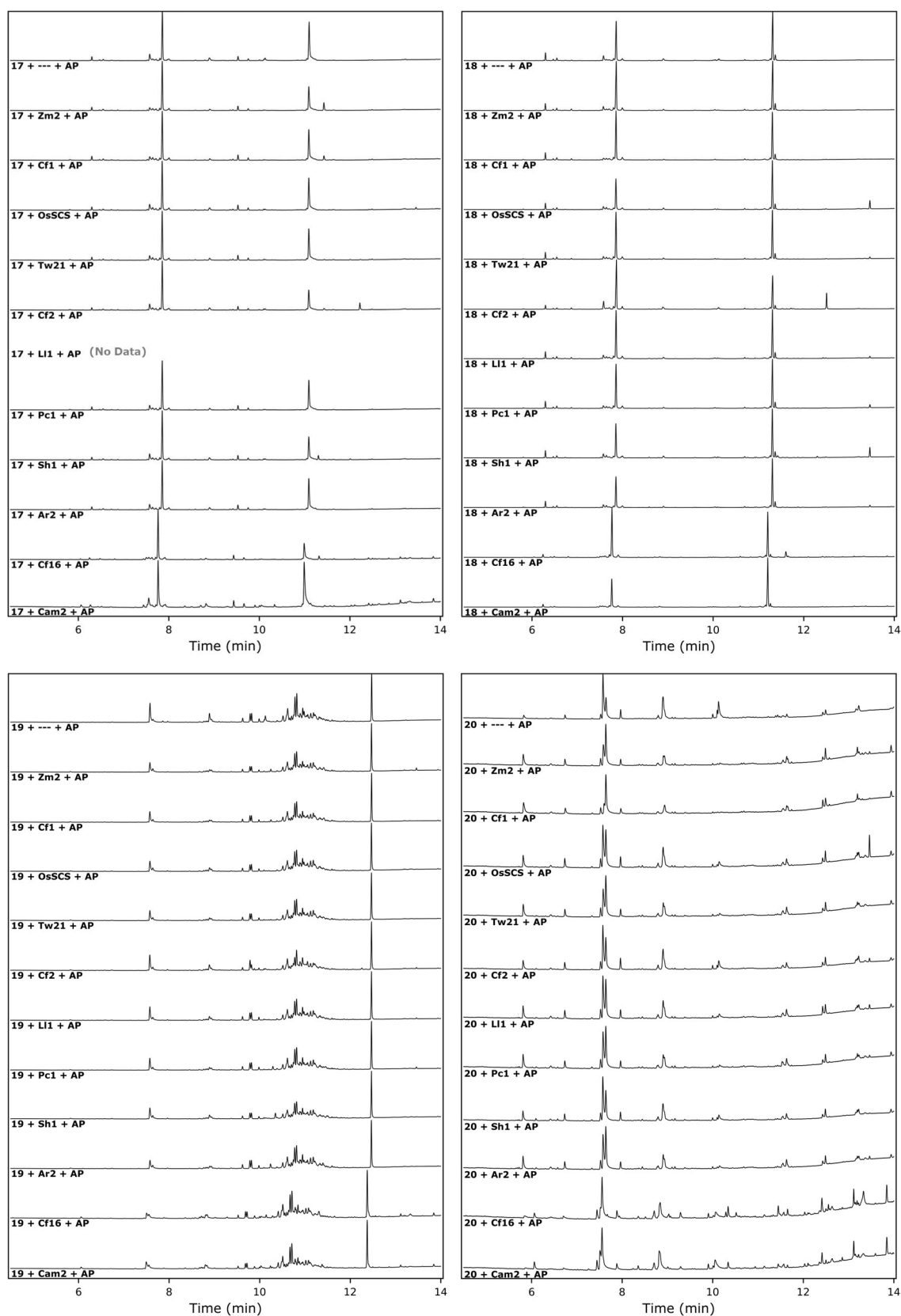


Figure 4.S11 (cont'd)



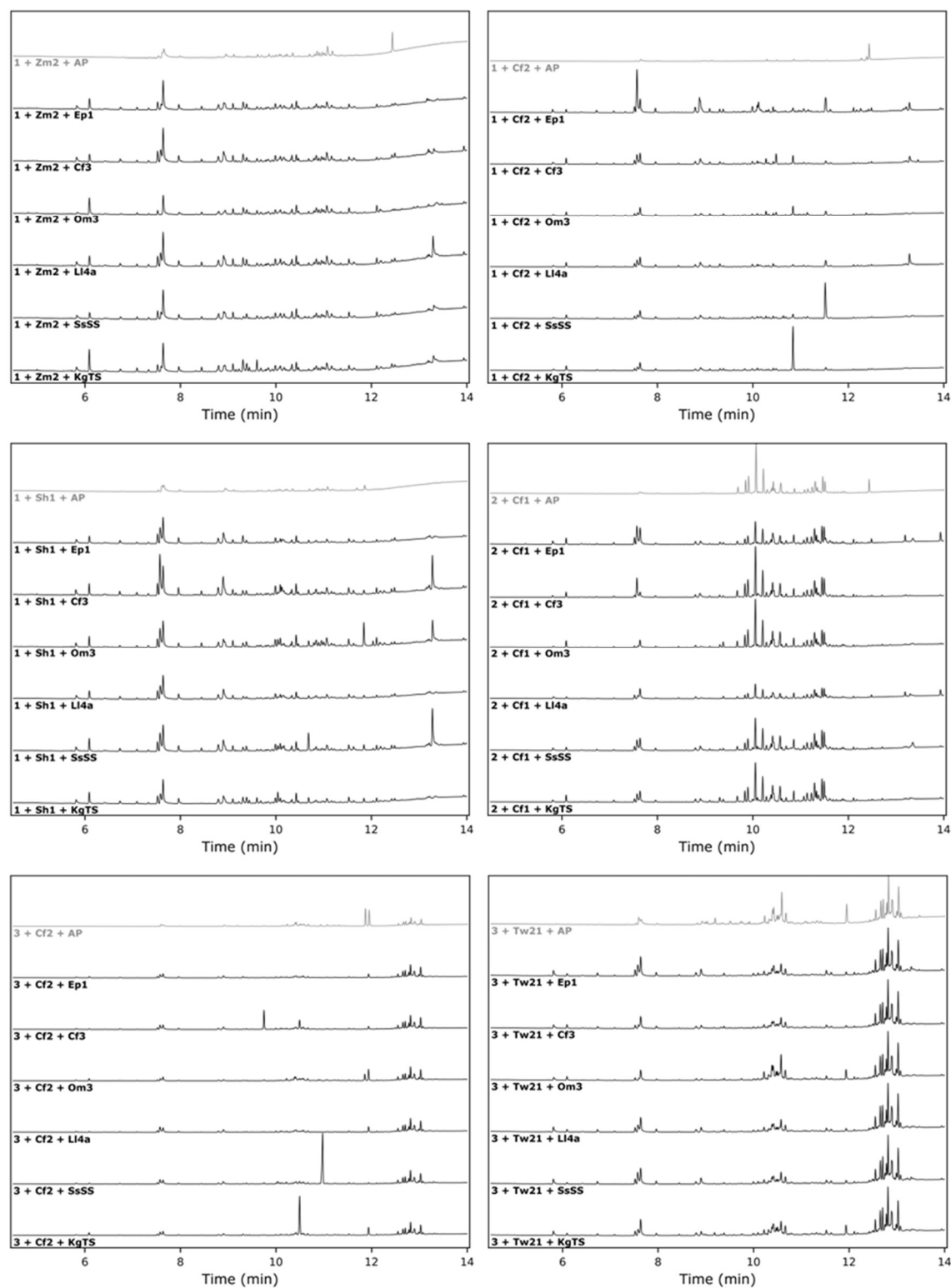


Figure 4.S12: *Initial GC-FID screening of class II/class I TPS combinations.* Corresponding class II sample (from screening shown in Figure 4.S11) is at the top of each set in gray.

Figure 4.S12 (cont'd)

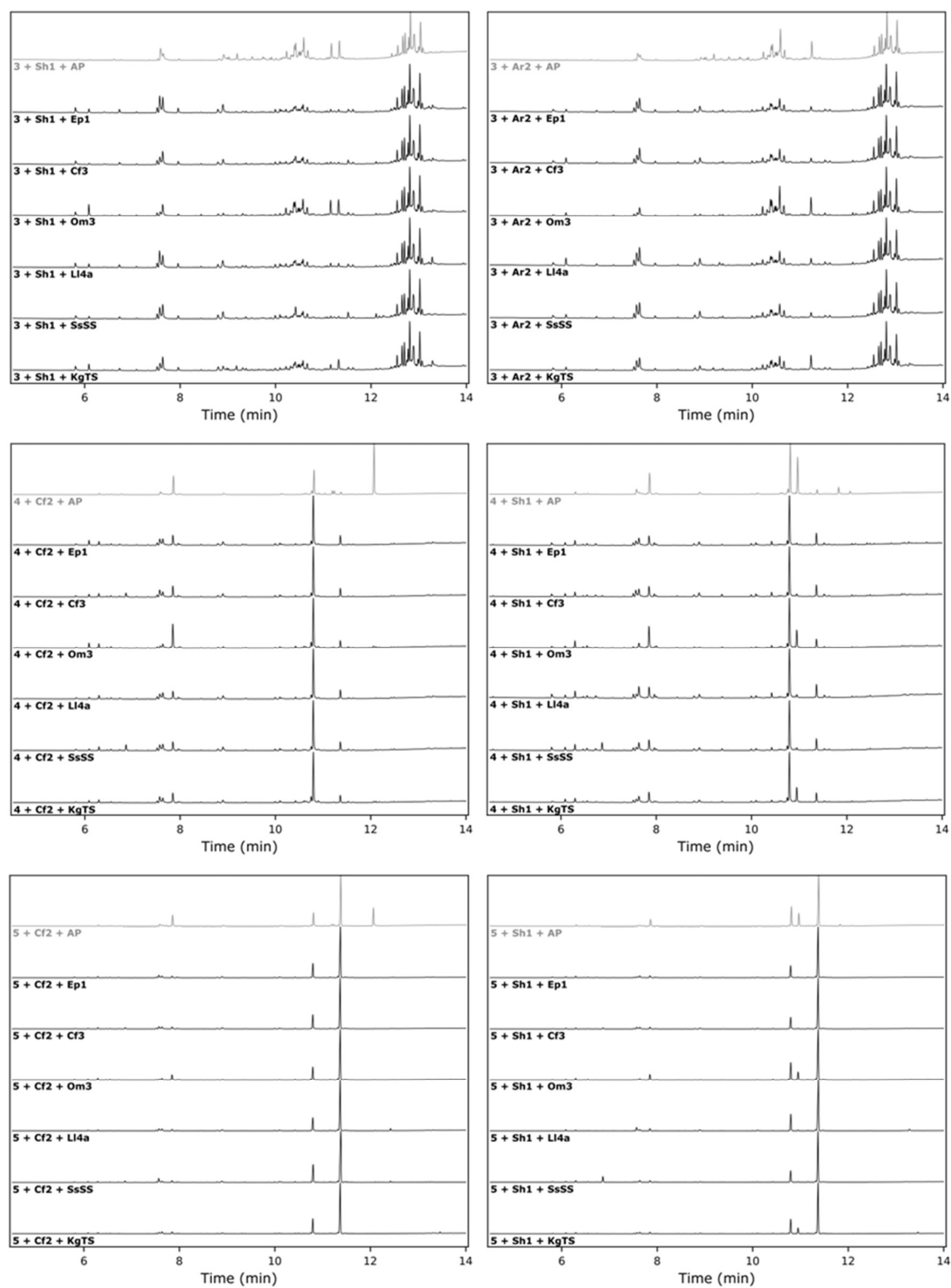
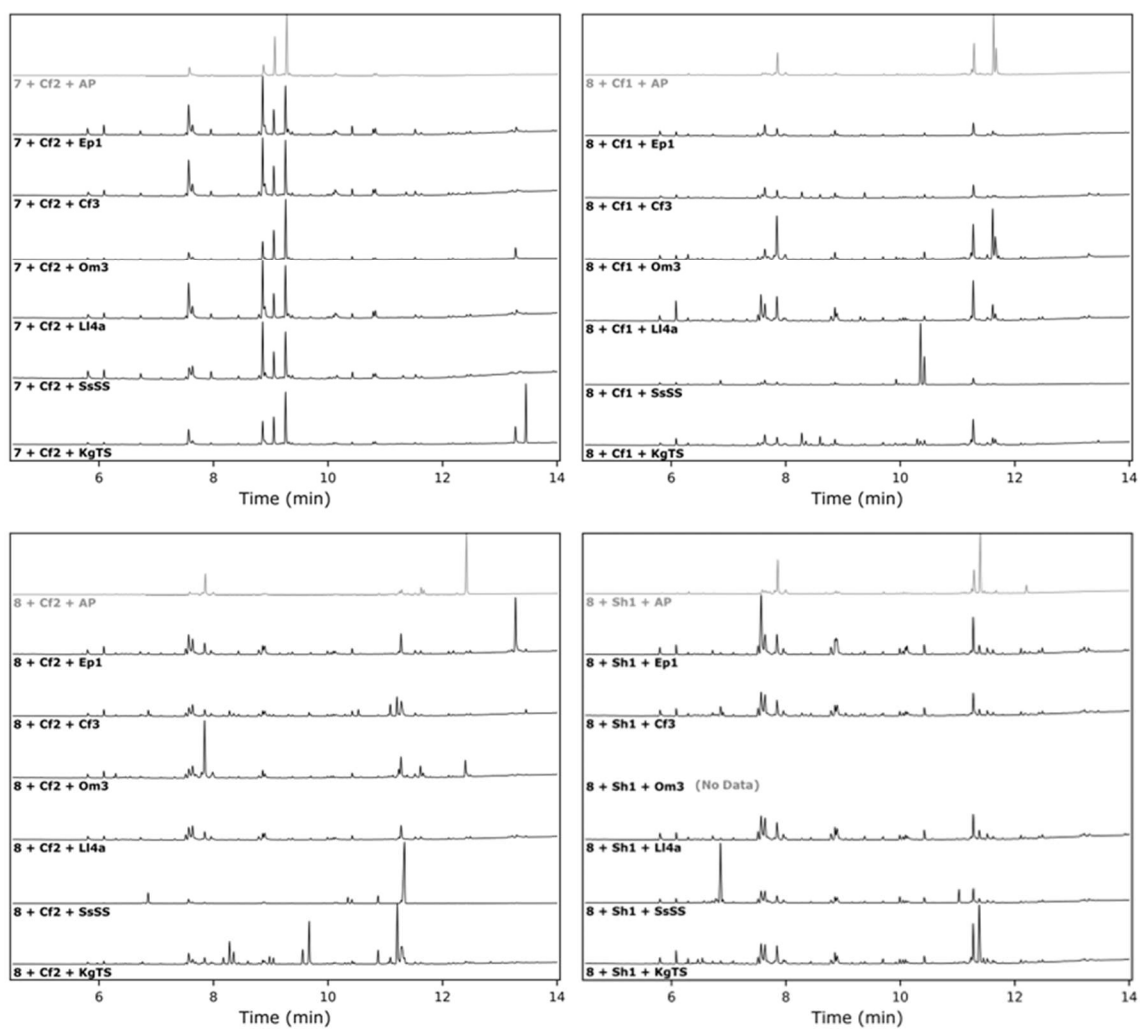


Figure 4.S12 (cont'd)



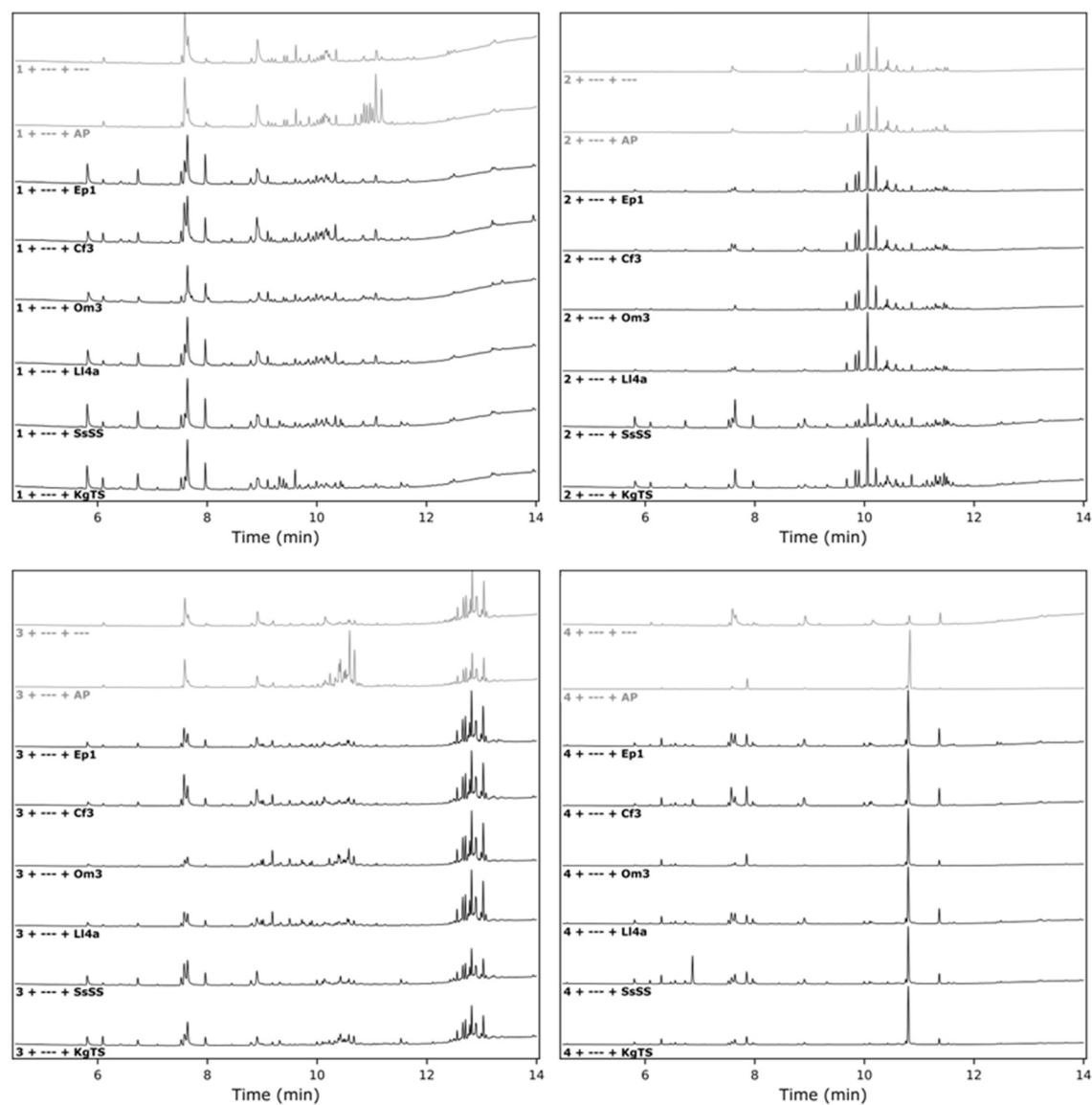
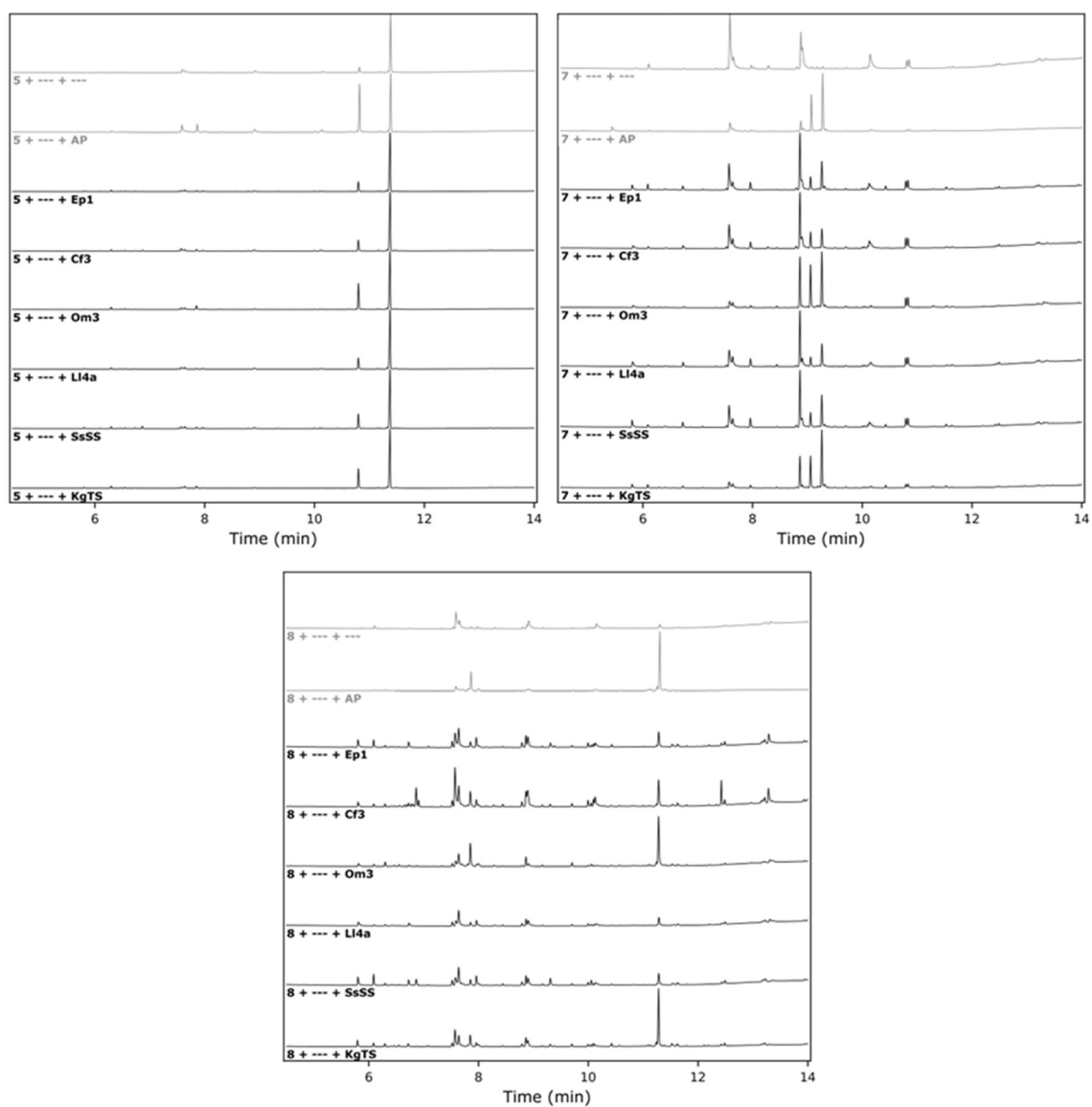


Figure 4.S13: *GC-FID chromatograms for second-step class I TPS negative controls.*

Figure 4.S13 (cont'd)



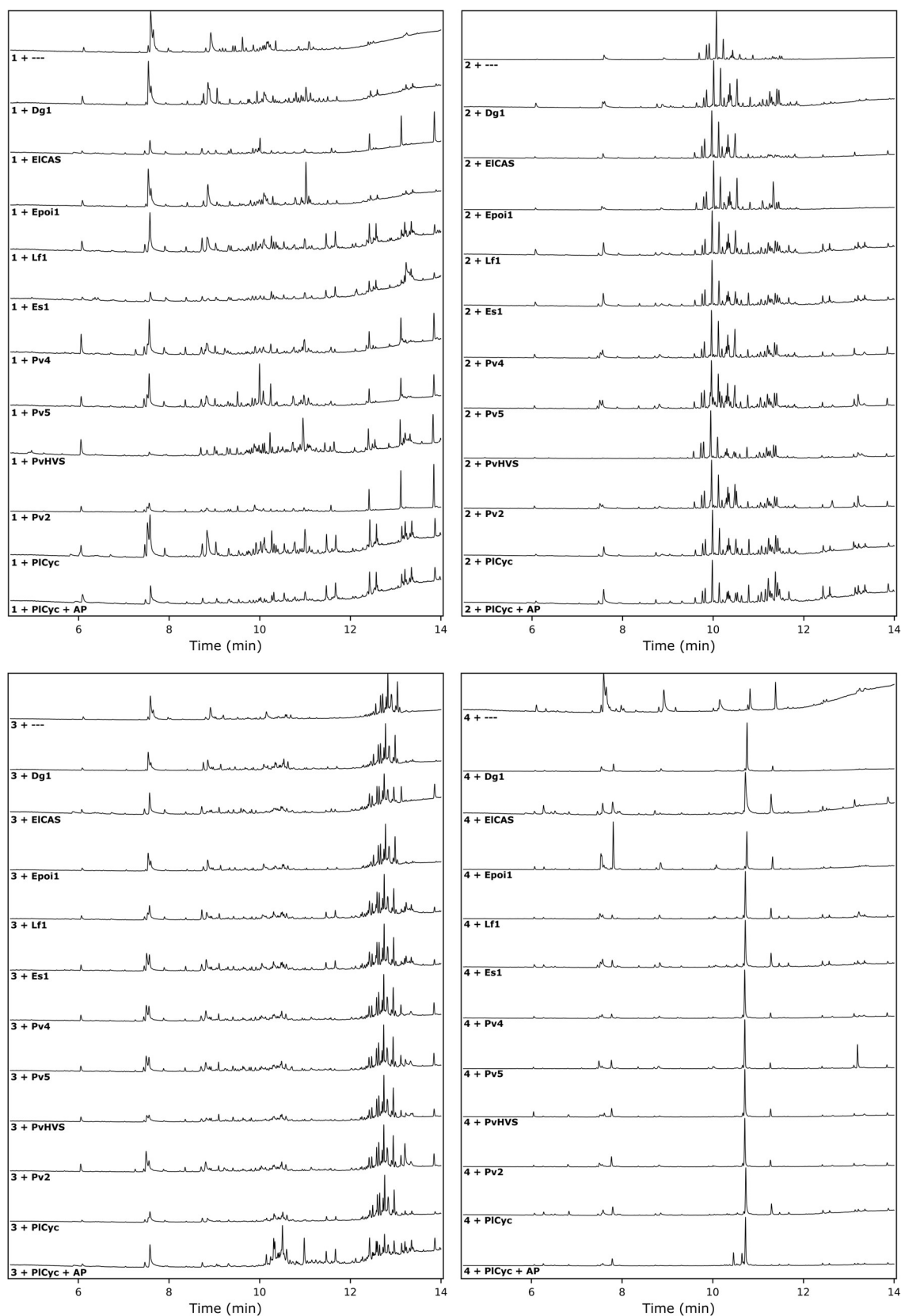


Figure 4.S14: Initial GC-FID screening of single-step class I TPSs with substrates 1-20.

Figure 4.S14 (cont'd)

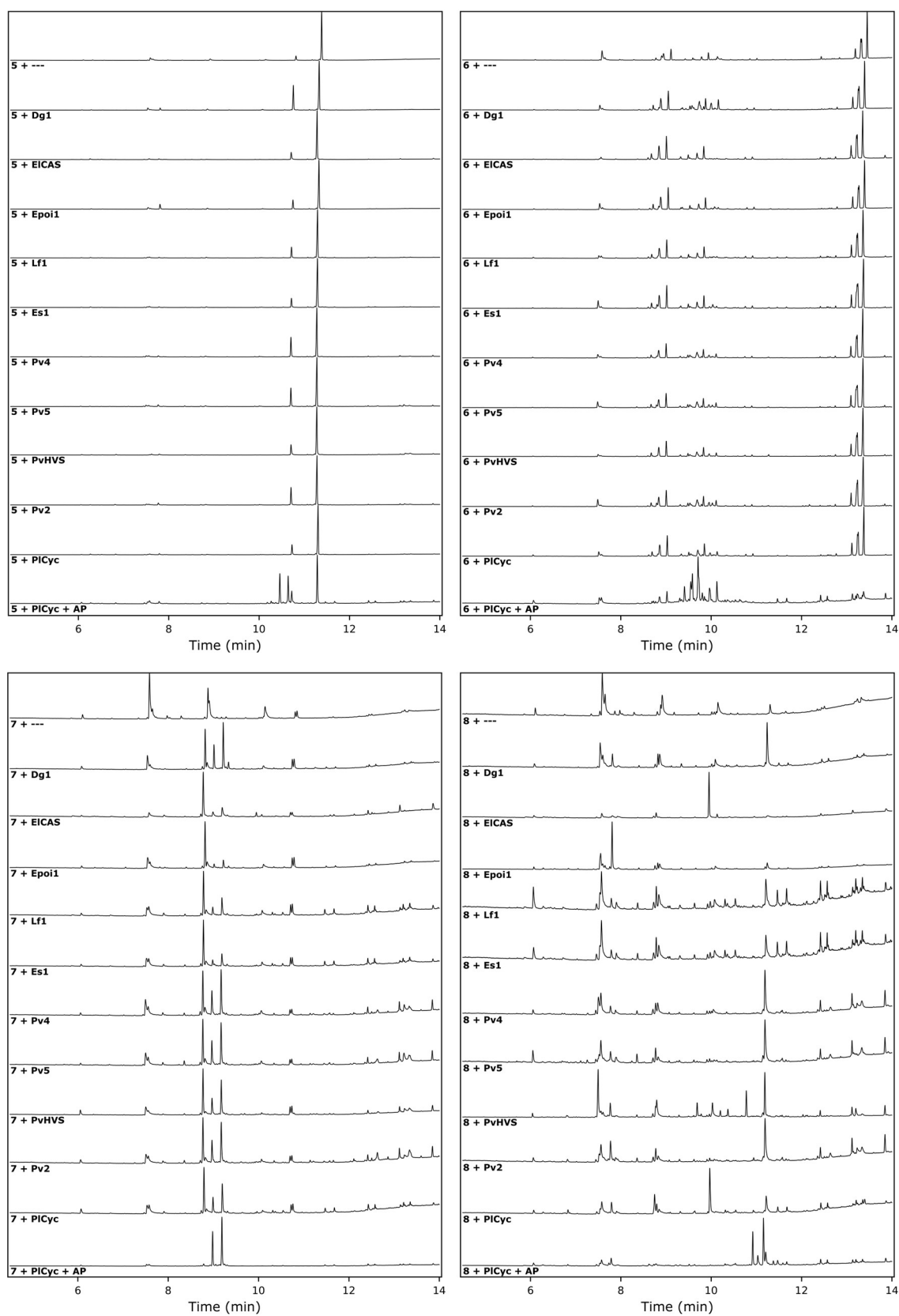


Figure 4.S14 (cont'd)

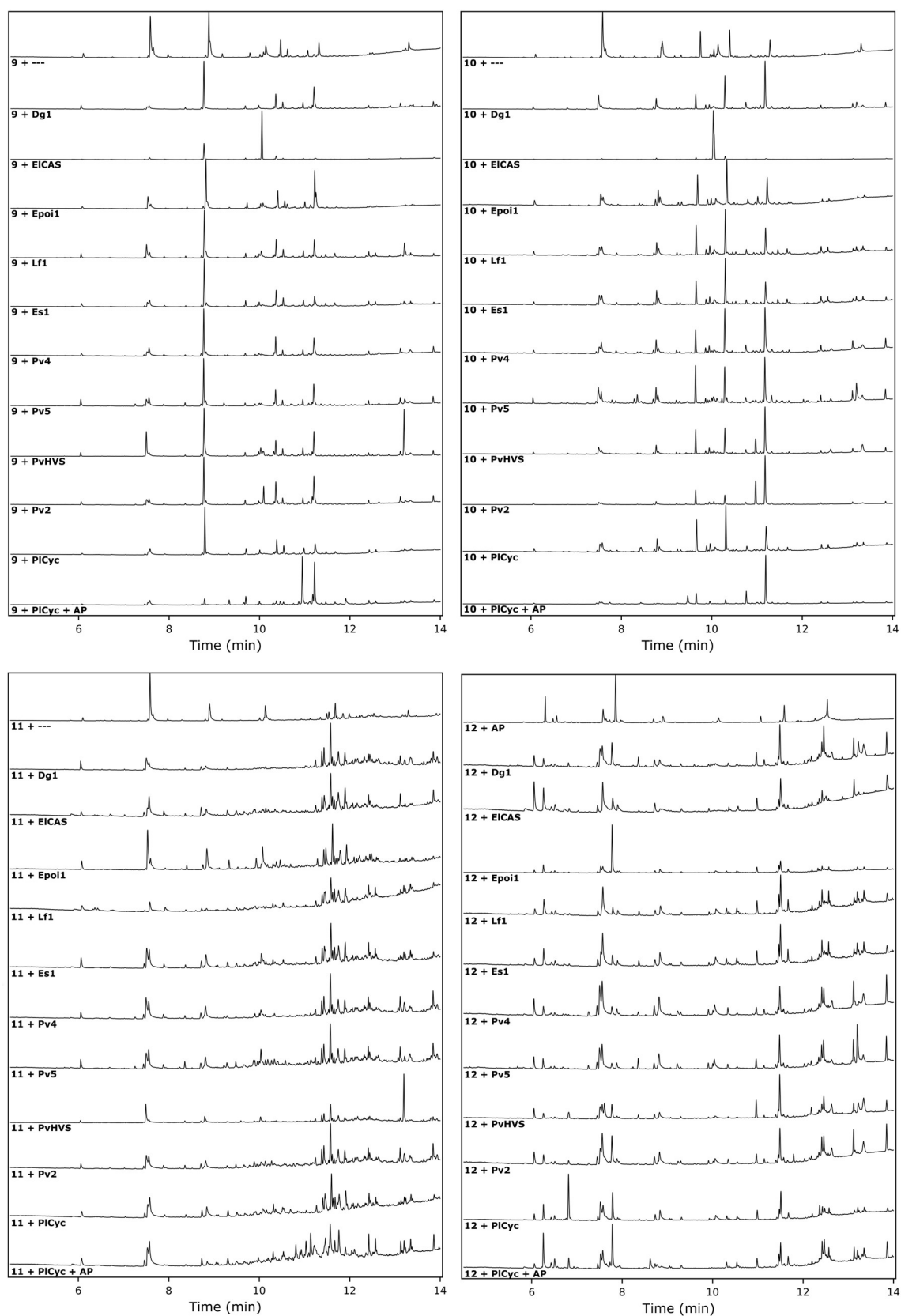


Figure 4.S14 (cont'd)

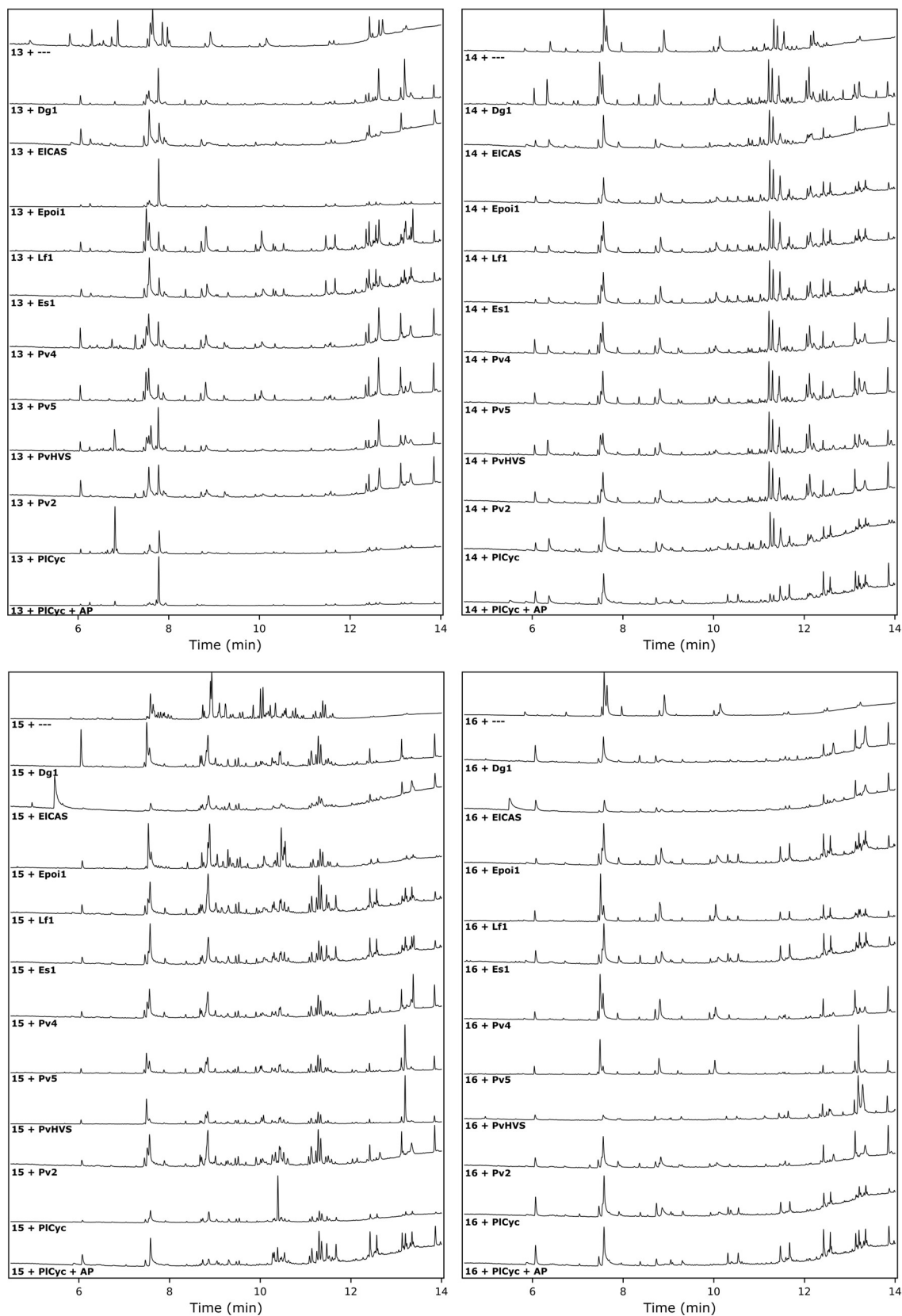
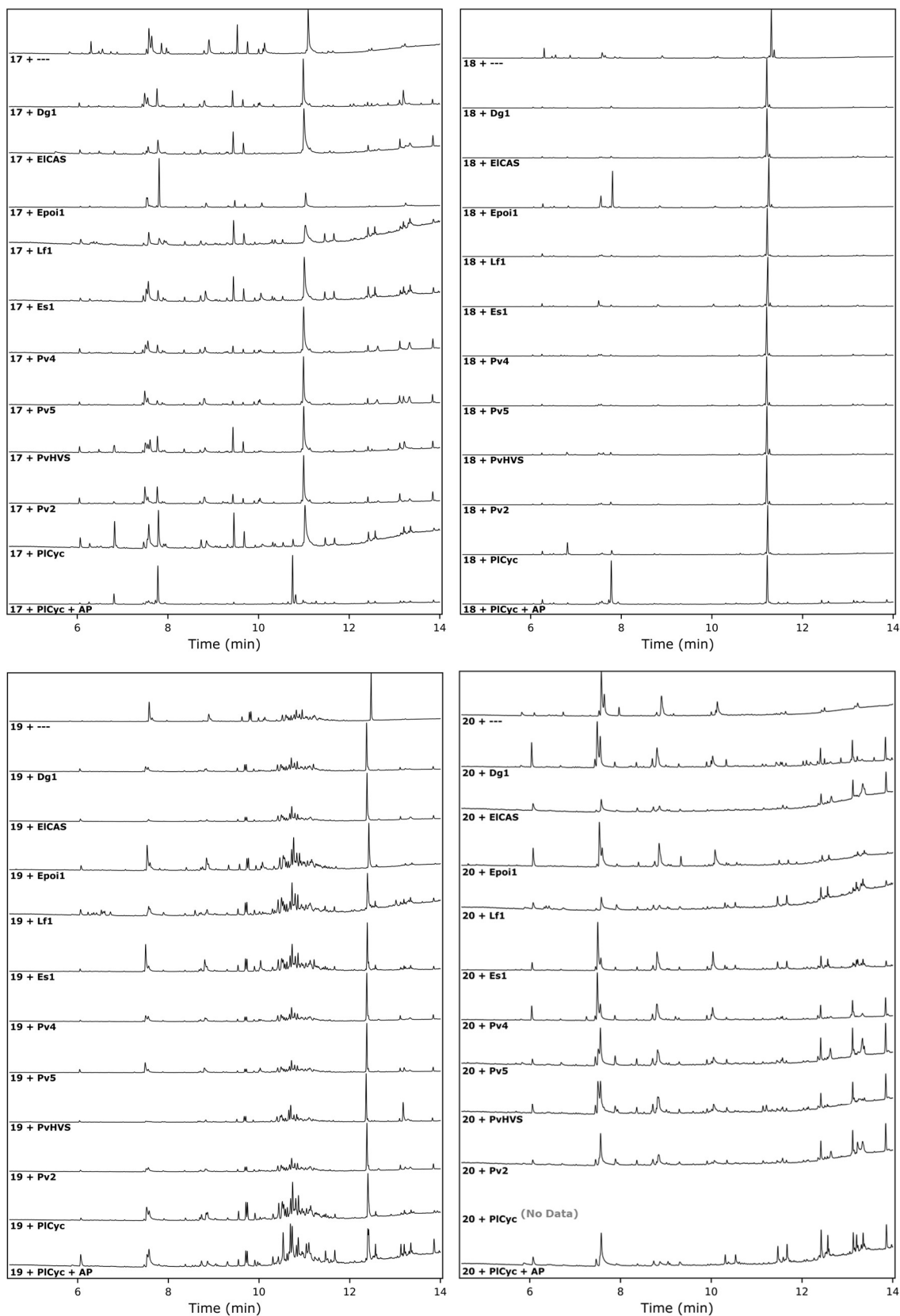


Figure 4.S14 (cont'd)



REFERENCES

REFERENCES

- (1) Zeng, T.; Liu, Z.; Liu, H.; He, W.; Tang, X.; Xie, L.; Wu, R. Exploring Chemical and Biological Space of Terpenoids. *J. Chem. Inf. Model.* **2019**, *59* (9), 3667–3678. <https://doi.org/10.1021/acs.jcim.9b00443>.
- (2) Karunanithi, P. S.; Zerbe, P. Terpene Synthases as Metabolic Gatekeepers in the Evolution of Plant Terpenoid Chemical Diversity. *Frontiers in Plant Science* **2019**, *10*.
- (3) Andersen-Ranberg, J.; Kongstad, K. T.; Nielsen, M. T.; Jensen, N. B.; Pateraki, I.; Bach, S. S.; Hamberger, B.; Zerbe, P.; Staerk, D.; Bohlmann, J.; Möller, B. L.; Hamberger, B. Expanding the Landscape of Diterpene Structural Diversity through Stereochemically Controlled Combinatorial Biosynthesis. *Angewandte Chemie International Edition* **2016**, *55* (6), 2142–2146. <https://doi.org/10.1002/anie.201510650>.
- (4) Peters, R. J. Two Rings in Them All: The Labdane-Related Diterpenoids. *Nat. Prod. Rep.* **2010**, *27* (11), 1521–1530. <https://doi.org/10.1039/C0NP00019A>.
- (5) Johnson, S. R.; Bhat, W. W.; Bibik, J.; Turmo, A.; Hamberger, B.; Hamberger, B. A Database-Driven Approach Identifies Additional Diterpene Synthase Activities in the Mint Family (Lamiaceae). *J Biol Chem* **2019**, *294* (4), 1349–1362. <https://doi.org/10.1074/jbc.RA118.006025>.
- (6) Jia, M.; Potter, K. C.; Peters, R. J. Extreme Promiscuity of a Bacterial and a Plant Diterpene Synthase Enables Combinatorial Biosynthesis. *Metab Eng* **2016**, *37*, 24–34. <https://doi.org/10.1016/j.ymben.2016.04.001>.
- (7) Jia, M.; Mishra, S. K.; Tufts, S.; Jernigan, R. L.; Peters, R. J. Combinatorial Biosynthesis and the Basis for Substrate Promiscuity in Class I Diterpene Synthases. *Metab Eng* **2019**, *55*, 44–58. <https://doi.org/10.1016/j.ymben.2019.06.008>.
- (8) Jia, M.; Peters, R. J. Cis or Trans with Class II Diterpene Cyclases. *Org Biomol Chem* **2017**, *15* (15), 3158–3160. <https://doi.org/10.1039/c7ob00510e>.
- (9) Johnson, S. R.; Bhat, W. W.; Sadre, R.; Miller, G. P.; Garcia, A. S.; Hamberger, B. Promiscuous Terpene Synthases from *Prunella Vulgaris* Highlight the Importance of Substrate and Compartment Switching in Terpene Synthase Evolution. *New Phytologist* **2019**, *223* (1), 323–335. <https://doi.org/10.1111/nph.15778>.
- (10) Demiray, M.; Tang, X.; Wirth, T.; Faraldos, J. A.; Allemann, R. K. An Efficient Chemoenzymatic Synthesis of Dihydroartemisinic Aldehyde. *Angewandte Chemie International Edition* **2017**, *56* (15), 4347–4350. <https://doi.org/10.1002/anie.201609557>.
- (11) Huynh, F.; Grundy, D. J.; Jenkins, R. L.; Miller, D. J.; Allemann, R. K. Sesquiterpene Synthase-Catalysed Formation of a New Medium-Sized Cyclic Terpenoid Ether from Farnesyl

Diphosphate Analogues. *ChemBioChem* **2018**, *19* (17), 1834–1838. <https://doi.org/10.1002/cbic.201800218>.

(12) Johnson, L. A.; Dunbabin, A.; Benton, J. C. R.; Mart, R. J.; Allemann, R. K. Modular Chemoenzymatic Synthesis of Terpenes and Their Analogues. *Angewandte Chemie International Edition* **2020**, *59* (22), 8486–8490. <https://doi.org/10.1002/anie.202001744>.

(13) Eiben, C. B.; de Rond, T.; Bloeszies, C.; Gin, J.; Chiniquy, J.; Baidoo, E. E. K.; Petzold, C. J.; Hillson, N. J.; Fiehn, O.; Keasling, J. D. Mevalonate Pathway Promiscuity Enables Noncanonical Terpene Production. *ACS Synth. Biol.* **2019**, *8* (10), 2238–2247. <https://doi.org/10.1021/acssynbio.9b00230>.

(14) Kschowak, M. J.; Wortmann, H.; Dickschat, J. S.; Schrader, J.; Buchhaupt, M. Heterologous Expression of 2-Methylisoborneol / 2 Methylenebornane Biosynthesis Genes in Escherichia Coli Yields Novel C11-Terpenes. *PLOS ONE* **2018**, *13* (4), e0196082. <https://doi.org/10.1371/journal.pone.0196082>.

(15) Oberhauser, C.; Harms, V.; Seidel, K.; Schröder, B.; Ekramzadeh, K.; Beutel, S.; Winkler, S.; Lauterbach, L.; Dickschat, J. S.; Kirschning, A. Exploiting the Synthetic Potential of Sesquiterpene Cyclases for Generating Unnatural Terpenoids. *Angewandte Chemie International Edition* **2018**, *57* (36), 11802–11806. <https://doi.org/10.1002/anie.201805526>.

(16) Harms, V.; Schröder, B.; Oberhauser, C.; Tran, C. D.; Winkler, S.; Dräger, G.; Kirschning, A. Methyl-Shifted Farnesyl diphosphate Derivatives Are Substrates for Sesquiterpene Cyclases. *Org. Lett.* **2020**, *22* (11), 4360–4365. <https://doi.org/10.1021/acs.orglett.0c01345>.

(17) Li, H.; S. Dickschat, J. Isotopic Labelling Experiments and Enzymatic Preparation of Iso-Casbenes with Casbene Synthase from Ricinus Communis. *Organic Chemistry Frontiers* **2022**, *9* (3), 795–801. <https://doi.org/10.1039/D1QO01707A>.

(18) Harris, L. J.; Saparno, A.; Johnston, A.; Pristic, S.; Xu, M.; Allard, S.; Kathiresan, A.; Ouellet, T.; Peters, R. J. The Maize An2 Gene Is Induced by Fusarium Attack and Encodes an Ent-Copalyl Diphosphate Synthase. *Plant Mol Biol* **2005**, *59* (6), 881–894. <https://doi.org/10.1007/s11103-005-1674-8>.

(19) Xu, M.; Hillwig, M. L.; Pristic, S.; Coates, R. M.; Peters, R. J. Functional Identification of Rice Syn-Copalyl Diphosphate Synthase and Its Role in Initiating Biosynthesis of Diterpenoid Phytoalexin/Allelopathic Natural Products. *The Plant Journal* **2004**, *39* (3), 309–318. <https://doi.org/10.1111/j.1365-313X.2004.02137.x>.

(20) Hansen, N. L.; Heskes, A. M.; Hamberger, B.; Olsen, C. E.; Hallström, B. M.; Andersen-Ranberg, J.; Hamberger, B. The Terpene Synthase Gene Family in Tripterygium Wilfordii Harbors a Labdane-Type Diterpene Synthase among the Monoterpene Synthase TPS-b Subfamily. *Plant J* **2017**, *89* (3), 429–441. <https://doi.org/10.1111/tpj.13410>.

- (21) Hamilton, J. P.; Godden, G. T.; Lanier, E.; Bhat, W. W.; Kinser, T. J.; Vaillancourt, B.; Wang, H.; Wood, J. C.; Jiang, J.; Soltis, P. S.; Soltis, D. E.; Hamberger, B.; Buell, C. R. Generation of a Chromosome-Scale Genome Assembly of the Insect-Repellent Terpenoid-Producing Lamiaceae Species, *Callicarpa Americana*. *GigaScience* **2020**, *9* (9), gaaa093. <https://doi.org/10.1093/gigascience/giaa093>.
- (22) Pateraki, I.; Andersen-Ranberg, J.; Hamberger, B.; Heskes, A. M.; Martens, H. J.; Zerbe, P.; Bach, S. S.; Møller, B. L.; Bohlmann, J.; Hamberger, B. Manoyl Oxide (13R), the Biosynthetic Precursor of Forskolol, Is Synthesized in Specialized Root Cork Cells in *Coleus forskohlii*. *Plant Physiology* **2014**, *164* (3), 1222–1236. <https://doi.org/10.1104/pp.113.228429>.
- (23) Luo, D.; Callari, R.; Hamberger, B.; Wubshet, S. G.; Nielsen, M. T.; Andersen-Ranberg, J.; Hallström, B. M.; Cozzi, F.; Heider, H.; Lindberg Møller, B.; Staerk, D.; Hamberger, B. Oxidation and Cyclization of Casbene in the Biosynthesis of Euphorbia Factors from Mature Seeds of *Euphorbia lathyris* L. *Proceedings of the National Academy of Sciences* **2016**, *113* (34), E5082–E5089. <https://doi.org/10.1073/pnas.1607504113>.
- (24) Miller, G. P.; Bhat, W. W.; Lanier, E. R.; Johnson, S. R.; Mathieu, D. T.; Hamberger, B. The Biosynthesis of the Anti-Microbial Diterpenoid Leubethanol in *Leucophyllum frutescens* Proceeds via an All-Cis Prenyl Intermediate. *The Plant Journal* **2020**, *104* (3), 693–705. <https://doi.org/10.1111/tbj.14957>.
- (25) Zerbe, P.; Hamberger, B.; Yuen, M. M. S.; Chiang, A.; Sandhu, H. K.; Madilao, L. L.; Nguyen, A.; Hamberger, B.; Bach, S. S.; Bohlmann, J. Gene Discovery of Modular Diterpene Metabolism in Nonmodel Systems. *Plant Physiology* **2013**, *162* (2), 1073–1091. <https://doi.org/10.1104/pp.113.218347>.
- (26) Caniard, A.; Zerbe, P.; Legrand, S.; Cohade, A.; Valot, N.; Magnard, J.-L.; Bohlmann, J.; Legendre, L. Discovery and Functional Characterization of Two Diterpene Synthases for Sclareol Biosynthesis in *Salvia sclarea* (L.) and Their Relevance for Perfume Manufacture. *BMC Plant Biol* **2012**, *12*, 119. <https://doi.org/10.1186/1471-2229-12-119>.
- (27) Dairi, T.; Hamano, Y.; Kuzuyama, T.; Itoh, N.; Furihata, K.; Seto, H. Eubacterial Diterpene Cyclase Genes Essential for Production of the Isoprenoid Antibiotic Terpentecin. *Journal of Bacteriology* **2001**, *183* (20), 6085–6094. <https://doi.org/10.1128/JB.183.20.6085-6094.2001>.
- (28) Lou, H.; Zheng, S.; Li, T.; Zhang, J.; Fei, Y.; Hao, X.; Liang, G.; Pan, W. Vulgarisin A, a New Diterpenoid with a Rare 5/6/4/5 Ring Skeleton from the Chinese Medicinal Plant *Prunella vulgaris*. *Org. Lett.* **2014**, *16* (10), 2696–2699. <https://doi.org/10.1021/ol5009763>.
- (29) Durairaj, J.; Di Girolamo, A.; Bouwmeester, H. J.; de Ridder, D.; Beekwilder, J.; van Dijk, A. DJ. An Analysis of Characterized Plant Sesquiterpene Synthases. *Phytochemistry* **2019**, *158*, 157–165. <https://doi.org/10.1016/j.phytochem.2018.10.020>.
- (30) Crowe, C.; Molyneux, S.; V. Sharma, S.; Zhang, Y.; S. Gkotsi, D.; Connaris, H.; M. Goss, R. J. Halogenases: A Palette of Emerging Opportunities for Synthetic Biology–Synthetic

Chemistry and C–H Functionalisation. *Chemical Society Reviews* **2021**, 50 (17), 9443–9481. <https://doi.org/10.1039/D0CS01551B>.

(31) Latham, J.; Brandenburger, E.; Shepherd, S. A.; Menon, B. R. K.; Micklefield, J. Development of Halogenase Enzymes for Use in Synthesis. *Chem. Rev.* **2018**, 118 (1), 232–269. <https://doi.org/10.1021/acs.chemrev.7b00032>.

(32) Liu, R.; Li, X.; Lam, K. S. Combinatorial Chemistry in Drug Discovery. *Curr Opin Chem Biol* **2017**, 38, 117–126. <https://doi.org/10.1016/j.cbpa.2017.03.017>.

(33) Kennedy, J. P.; Williams, L.; Bridges, T. M.; Daniels, R. N.; Weaver, D.; Lindsley, C. W. Application of Combinatorial Chemistry Science on Modern Drug Discovery. *J. Comb. Chem.* **2008**, 10 (3), 345–354. <https://doi.org/10.1021/cc700187t>.

(34) Seamon, K. B.; Padgett, W.; Daly, J. W. Forskolin: Unique Diterpene Activator of Adenylate Cyclase in Membranes and in Intact Cells. *Proceedings of the National Academy of Sciences* **1981**, 78 (6), 3363–3367. <https://doi.org/10.1073/pnas.78.6.3363>.

(35) Pateraki, I.; Andersen-Ranberg, J.; Jensen, N. B.; Wubshet, S. G.; Heskes, A. M.; Forman, V.; Hallström, B.; Hamberger, B.; Motawia, M. S.; Olsen, C. E.; Staerk, D.; Hansen, J.; Møller, B. L.; Hamberger, B. Total Biosynthesis of the Cyclic AMP Booster Forskolin from *Coleus forskohlii*. *eLife* **2017**, 6, e23001. <https://doi.org/10.7554/eLife.23001>.

(36) Gulakowski, R. J.; McMahon, J. B.; Buckheit, R. W.; Gustafson, K. R.; Boyd, M. R. Antireplicative and Anticytopathic Activities of Prostratin, a Non-Tumor-Promoting Phorbol Ester, against Human Immunodeficiency Virus (HIV)1Part 24 in the Series HIV-Inhibitory Natural Products; for Part 23, See Beutler, J.A. et al. (1995) *J. Nat. Prod.* 58, 1039–1046.1. *Antiviral Research* **1997**, 33 (2), 87–97. [https://doi.org/10.1016/S0166-3542\(96\)01004-2](https://doi.org/10.1016/S0166-3542(96)01004-2).

(37) Johnson, H. E.; Banack, S. A.; Cox, P. A. Variability in Content of the Anti-AIDS Drug Candidate Prostratin in Samoan Populations of *Homalanthus Nutans*. *J. Nat. Prod.* **2008**, 71 (12), 2041–2044. <https://doi.org/10.1021/np800295m>.

(38) Siller, G.; Gebauer, K.; Welburn, P.; Katsamas, J.; Ogbourne, S. M. PEP005 (Ingenol Mebutate) Gel, a Novel Agent for the Treatment of Actinic Keratosis: Results of a Randomized, Double-Blind, Vehicle-Controlled, Multicentre, Phase IIa Study. *Australasian Journal of Dermatology* **2009**, 50 (1), 16–22. <https://doi.org/10.1111/j.1440-0960.2008.00497.x>.

(39) Kedei, N.; Lundberg, D. J.; Toth, A.; Welburn, P.; Garfield, S. H.; Blumberg, P. M. Characterization of the Interaction of Ingenol 3-Angelate with Protein Kinase C. *Cancer Research* **2004**, 64 (9), 3243–3255. <https://doi.org/10.1158/0008-5472.CAN-03-3403>.

(40) Bathe, U.; Tissier, A. Cytochrome P450 Enzymes: A Driving Force of Plant Diterpene Diversity. *Phytochemistry* **2019**, 161, 149–162. <https://doi.org/10.1016/j.phytochem.2018.12.003>.

- (41) Hamberger, B.; Bak, S. Plant P450s as Versatile Drivers for Evolution of Species-Specific Chemical Diversity. *Philos Trans R Soc Lond B Biol Sci* **2013**, 368 (1612), 20120426. <https://doi.org/10.1098/rstb.2012.0426>.
- (42) Huang, G.-J.; Pan, C.-H.; Wu, C.-H. Sclareol Exhibits Anti-Inflammatory Activity in Both Lipopolysaccharide-Stimulated Macrophages and the λ -Carrageenan-Induced Paw Edema Model. *J. Nat. Prod.* **2012**, 75 (1), 54–59. <https://doi.org/10.1021/np200512a>.
- (43) Tsai, S.-W.; Hsieh, M.-C.; Li, S.; Lin, S.-C.; Wang, S.-P.; Lehman, C. W.; Lien, C. Z.; Lin, C.-C. Therapeutic Potential of Sclareol in Experimental Models of Rheumatoid Arthritis. *Int J Mol Sci* **2018**, 19 (5), 1351. <https://doi.org/10.3390/ijms19051351>.
- (44) Jia, M.; O'Brien, T. E.; Zhang, Y.; Siegel, J. B.; Tantillo, D. J.; Peters, R. J. Changing Face: A Key Residue for the Addition of Water by Sclareol Synthase. *ACS Catal* **2018**, 8 (4), 3133–3137. <https://doi.org/10.1021/acscatal.8b00121>.
- (45) Potter, K. C.; Zi, J.; Hong, Y. J.; Schulte, S.; Malchow, B.; Tantillo, D. J.; Peters, R. J. Blocking Deprotonation with Retention of Aromaticity in a Plant Ent-Copalyl Diphosphate Synthase Leads to Product Rearrangement. *Angewandte Chemie International Edition* **2016**, 55 (2), 634–638. <https://doi.org/10.1002/anie.201509060>.
- (46) Criswell, J.; Potter, K.; Shephard, F.; Beale, M. H.; Peters, R. J. A Single Residue Change Leads to a Hydroxylated Product from the Class II Diterpene Cyclization Catalyzed by Abietadiene Synthase. *Org. Lett.* **2012**, 14 (23), 5828–5831. <https://doi.org/10.1021/ol3026022>.
- (47) Hansen, N. L.; Nissen, J. N.; Hamberger, B. Two Residues Determine the Product Profile of the Class II Diterpene Synthases TPS14 and TPS21 of *Tripterygium Wilfordii*. *Phytochemistry* **2017**, 138, 52–56. <https://doi.org/10.1016/j.phytochem.2017.02.022>.
- (48) Emanuelsson, O.; Nielsen, H.; Brunak, S.; von Heijne, G. Predicting Subcellular Localization of Proteins Based on Their N-Terminal Amino Acid Sequence. *Journal of Molecular Biology* **2000**, 300 (4), 1005–1016. <https://doi.org/10.1006/jmbi.2000.3903>.
- (49) Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, Scalable Generation of High-Quality Protein Multiple Sequence Alignments Using Clustal Omega. *Mol Syst Biol* **2011**, 7, 539. <https://doi.org/10.1038/msb.2011.75>.
- (50) Hayashi, K.; Kawaide, H.; Notomi, M.; Sakigi, Y.; Matsuo, A.; Nozaki, H. Identification and Functional Analysis of Bifunctional Ent-Kaurene Synthase from the Moss *Physcomitrella Patens*. *FEBS Letters* **2006**, 580 (26), 6175–6181. <https://doi.org/10.1016/j.febslet.2006.10.018>.
- (51) Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics* **2014**, 30 (9), 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.

(52) Letunic, I.; Bork, P. Interactive Tree Of Life (ITOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res* **2021**, *49* (W1), W293–W296. <https://doi.org/10.1093/nar/gkab301>.