2012-08-31

Identification and characterization of diterpene synthases in the salvinorin A biosynthetic pathway

Mitchell, Rod

UNIVERSITY OF CALGARY

Identification and characterization of diterpene synthases in the salvinorin A biosynthetic pathway

by

Rod Mitchell

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF BIOLOGICAL SCIENCES

CALGARY, ALBERTA

AUGUST, 2012

Abstract

Salvia divinorum is a hallucinogenic plant that is used for divination and spiritual communion by the Mazatecs of Mexico. The active component of the plant is salvinorin A, a neoclerodane diterpenoid that selectively acts as a potent κ opioid receptor agonist. Salvinorin A's novel receptor binding profile makes derivatives of the compound potentially useful in the treatment of various psychiatric and mood disorders. In this work, next generation sequencing (Roche 454) was used to generate a *S. divinorum* transcript database. Five distinct putative diterpene synthase cDNAs (two type II and three type I) were identified in the database. We report here a recombinant type II diterpene synthase capable of catalyzing the rearrangement of GGPP into terpentedienyl diphosphate (a neoclerodane diphosphate), and a recombinant type I diterpene synthase that renders terpentedienyl diphosphate into kolavenol. These enzymatic products are potential intermediates in the salvinorin A biosynthetic pathway.

Acknowledgements

I would like to thank my graduate supervisor, Dr. Dae-Kyun Ro, for making this interesting, exciting, and frequently frustrating project possible, and for his invaluable guidance and support. Dr. Philipp Zerbe provided assistance in many aspects of this research, and his efforts are much appreciated. The members of my supervisory committee supplied me with comments and questions that helped me to steer this project in the correct direction. Every member of the Ro lab was of immense help to me, especially Gillian who is the heart and soul of our group. Finally, I would like to thank my family for always being there for me with kind words and encouragement. This work could not have been possible without the help of all of you.

Table of Contents

Abstract....ii

Acknowledgements....iii

Table of Contents....iv

List of Tables....vi

List of Figures....vii

List of Symbols, Abbreviations, and Nomenclature....ix

CHAPTER ONE: INTRODUCTION....1
1.1 Secondary metabolites....1
1.2 Isoprenoid biosynthesis....4
1.3 *Salvia divinorum*....10
1.4 Traditional and modern uses of *S. divinorum*....11
1.5 Salvinorin A....12
1.6 Pharmacology of salvinorin A....14
1.7 Pharmacokinetics of salvinorin A....17
1.8 Opioid receptor types and functions....18
1.9 Biosynthesis of salvinorin A by *S. divinorum*....19
1.10 Methyl jasmonate induction of glandular trichome density....22
1.11 Other *S. divinorum* secondary metabolites....22
1.12 Functional groups of salvinorin A....23
1.13 Potential medical uses of salvinorin A....24
1.14 Toxicology of salvinorin A....27
1.15 The salvinorin A biosynthetic pathway....27
1.16 Research rationale, goals, and objectives....37

CHAPTER TWO: MATERIALS AND METHODS....39
2.1 *S. divinorum* growth conditions....39
2.2 High performance liquid chromatography (HPLC) analyses....40
2.3 Gas Chromatography-Mass Spectrometry (GC-MS) analyses....40
2.4 Liquid Chromatography-Mass Spectrometry (LC-MS) analyses....40
2.5 *S. divinorum* organic solvent extractions....41
2.6 RNA isolation and cDNA library construction....42
2.7 Rapid amplification of cDNA ends (RACE) reactions....44
2.8 Methyl jasmonate treatment....47
2.9 Tissue specific real-time quantitative PCR (qPCR)....48
2.10 Subcloning into *E. coli* expression vectors....50
2.11 *E. coli* cultivation and protein extraction....55
2.12 *In vitro* activity assays....57

2.13 CPPSL2 product NMR analysis....58
2.14 Subcloning into yeast expression vectors....59
2.15 *S. cerevisiae* cultivation and protein extraction....62

CHAPTER THREE: RESULTS....64
3.1 Evidence of salvinorin A production by *S. divinorum*....64
3.2 Evidence of salvinorin A accumulation in peltate glandular trichomes....68
3.3 Preparation of *S. divinorum* cDNA....73
3.4 Characterization of *S. divinorum* cDNA libraries....75
3.5 Identification of candidate diterpene synthases....84
3.6 Methyl jasmonate induction of *S. divinorum*....90
3.7 Quantitative real-time PCR (qPCR) for candidate diterpene synthases....92
3.8 Production and characterization of recombinant CPPSL1/2....94
3.9 Production and characterization of recombinant KSL2/3....113
3.10 NMR analysis of CPPSL2 product....125
3.11 Yeast *in vivo* expression of candidate diterpene synthases....128

CHAPTER FOUR: DISCUSSION....136
4.1 Genomics of non-model medicinal plants....136
4.2 CPPSL2 catalyzes the synthesis of terpentedienyl diphosphate....138
4.3 Biochemical functions of KSL2 and KSL3....142
4.4 Development of a yeast *in vivo* expression system....144
4.5 Significance and future directions....146

Bibliography....148

APPENDIX A: NUCLEOTIDE AND PROTEIN SEQUENCES OF *S. divinorum* CANDIDATE DITERPENE SYNTHASES....163
A.1 *CPPSL1* nucleotide sequence....163
A.2 *CPPSL1* amino acid sequence....165
A.3 *CPPSL2* nucleotide sequence....165
A.4 *CPPSL2* amino acid sequence....168
A.5 *KSL1* nucleotide sequence....168
A.6 *KSL1* amino acid sequence....170
A.7 *KSL2* nucleotide sequence....170
A.8 *KSL2* amino acid sequence....172
A.9 *KSL3* nucleotide sequence....173
A.10 *KSL3* amino acid sequence....175

List of Tables

Table 1. Oligonucelotides used in RACE amplifications....44
Table 2. Oligonucleotides used in qPCR amplifications....50
Table 3. Oligonucelotides used in pDEST17 subclonings....52
Table 4. Oligonucelotides used in pET28b subclonings....53
Table 5. Oligonucelotides used in pESC-Leu subclonings....60
Table 6. Comparison of *S. divinorum* cDNA libraries....77
Table 7. MVA pathway representation in *S. divinorum* cDNA libraries....79
Table 8. DXP pathway representation in *S. divinorum* cDNA libraries....80
Table 9. Representation of the five most highly abundant unigenes in the *S. divinorum* standard cDNA library....82
Table 10. Representation of the five most highly abundant unigenes in the *S. divinorum* normalized cDNA library....83
Table 11. Representation of candidate diterpene synthases in *S. divinorum* cDNA libraries....85
Table 12. ^{13}C NMR chemical shifts (ppm) from CPPSL2 product and two kolavenol references....127

List of Figures

Figure 1. Chemical structure of isoprene....2
Figure 2. The MVA pathway of isoprenoid biosynthesis....7
Figure 3. The DXP pathway of isoprenoid biosynthesis....9
Figure 4. Chemical structures of *S. divinorum* secondary metabolites....14
Figure 5. Functional groups of salvinorin A....24
Figure 6. Diterpene synthase catalyzed reactions in the gibberellin biosynthetic pathway....29
Figure 7. Hypothetical salvinorin A biosynthetic pathway (generalized)....30
Figure 8. Hypothetical LPPS-like and sclareol synthase-like reactions in the salvinorin A biosynthetic pathway....32
Figure 9. Hypothetical TPPS and terpentriene synthase-like reactions in the salvinorin A biosynthetic pathway....34
Figure 10. Hypothetical *ent*-CPPS and kolavenol synthase-like reactions in the salvinorin A biosynthetic pathway....36
Figure 11. HPLC-DAD analysis of *S. divinorum* leaf extraction....66
Figure 12. LC-MS analysis of *S. divinorum* leaf extraction....67
Figure 13. Differential extraction of *S. divinorum* leaf tissue....70
Figure 14. Salvinorin A content of *S. divinorum* tissues....72
Figure 15. Integrity of *S. divinorum* young leaf total RNA....74
Figure 16. Qualitative comparison of *S. divinorum* young leaf standard and normalized cDNA....75
Figure 17. Diterpene synthase phylogenetic tree....89
Figure 18. *S. divinorum* peltate glandular trichome induction by methyl jasmonate....91
Figure 19. Tissue specific abundance of candidate diterpene synthase transcripts in *S. divinorum*....94
Figure 20. Chloroplast transit peptide predictions for candidate diterpene synthases....96
Figure 21. Diterpene synthase catalyzed reactions in the gibberellin and hypothesized salvinorin A biosynthetic pathways....97
Figure 22. Effect of induction time and temperature on production of soluble CPPSL1 and CPPSL2 protein in *E. coli*....100
Figure 23. Production of CPPSL1 and CPPSL2 protein in *E. coli* cultures induced for 24 hr at 16°C....102
Figure 24. GC-MS analysis of CPPSL1 *in vitro* assays....104
Figure 25. GC-MS analysis of CPPSL2 *in vitro* assays....108
Figure 26. LC-MS analysis of CPPSL2 *in vitro* assays....111
Figure 27. Chemical structures of kolavelool derivatives....112
Figure 28. Purification of CPPSL2, KSL2 and KSL3 protein in *E. coli* cultures induced for 24 hr at 16°C....114
Figure 29. GC-MS analysis of CPPSL2 and KSL2 co-incubation *in vitro* assays....116

Figure 30. GC-MS analysis of CPPSL2 and KSL3 co-incubation *in vitro* assays....118
Figure 31. GC-MS analysis of *ent*-CPPS and (+)-CPPS *in vitro* assays....121
Figure 32. GC-MS analysis of (+)-CPPS and KSL3 co-incubation *in vitro* assays....124
Figure 33. GC-MS mass spectra of KSL3 enzymatic substrates and products…125
Figure 34. GC-MS analysis of *LVPS* and "GGPP helper" plasmid *in vivo* yeast expression assays....131
Figure 35. Assay of *CPPSL2*, *KSL1/2/3,* and *BTS1 in vivo* yeast expression by anti-cMYC and anti-FLAG immunoblot....134
Figure 36. Oxygen geometry of kolavenol and kolavelool....141
Figure 37. Potential mechanisms of salvinorin A furan ring formation....142

List of Symbols, Abbreviations, and Nomenclature

Symbol	Definition
BLAST	Basic local alignment search tool
BTS	Bet two suppressor
CDP-ME	C-methyl-D-erythritol-methyl-erythritol
CDP-MEP	4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate
CMP	Cytidine monophosphate
CoA	Coenzyme A
CPP	Copalyl diphosphate
CPPS	Copalyl diphosphate synthase
CPPSL	Copalyl diphosphate synthase-like
CTP	Cytidine triphosphate
DAD	Diode array detector
DMAPP	Dimethylallyl diphosphate
DOR	Delta (δ) opioid receptor
DSN	Duplex specific nuclease
DXP	Deoxy-xylulose-5-phosphate
EI-MS	Electron ionization-mass spectrum
FPP	Farnesyl diphosphate
G3P	Glyceraldehyde-3-phosphate
GC-MS	Gas chromatography-mass spectrometry
GPP	Geranyl diphosphate
GGPP	Geranylgeranyl diphosphate
HMBPP	(E)-4-Hydroxy-3-methyl-but-2-enyl diphosphate
HMG	Hydroxyl-methyl-glutaryl
HPLC	High performance liquid chromatography
IPP	Isopentenyl diphosphate
KOR	Kappa (κ) opioid receptor
KS	Kaurene synthase
KSL	Kaurene synthase-like
LC-MS	Liquid chromatography-mass spectrometry

Symbol	Definition
LPP	Labdiendol diphosphate
LPPS	Labdiendol diphosphate synthase
LVP	Levopimaradiene
LVPS	Levopimaradiene synthase
MEcPP	2-C-methyl-D-erythritol 2,4-cyclopyrophosphate
MEP	2-C-methyl-D-erythritol 4-phosphate
MOR	Mu (μ) opioid receptor
MVA	Mevalonate
NCBI	National Centre for Biotechnology Information
NGS	Next generation sequencing
NMR	Nuclear magnetic resonance
ORF	Open reading frame
PCR	Polymerase chain reaction
PPM	Parts per million
RACE	Rapid amplification of cDNA ends
SDS-PAGE	Sodium dodecyl sulfate-poly-acrylamide gel electrophoresis
TPP	Terpentedienyl diphosphate
TPPS	Terpentedienyl diphosphate synthase
qPCR	Quantitative real-time PCR

CHAPTER ONE: INTRODUCTION

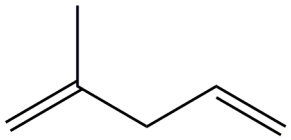
1.1 Secondary metabolites

Secondary (or specialized) metabolites are metabolic products that are not strictly necessary for the survival of an organism, but that influence the fitness of organisms in diverse ecological niches. Secondary metabolites were once thought to be waste products, but it is now known that they have a wide range of eco-physiological functions. These metabolites are widespread in kingdom *plantae*, as defensive compounds, signaling molecules, and pigments. There are four major classes of secondary metabolites in nature - isoprenoids (or terpenoids), polyketides, phenylpropanoids, and alkaloids [1]. These four types of metabolites are briefly outlined in the following pages.

Isoprenoids are defined as natural products derived from isopentenyl diphosphate (IPP). Pioneering research into isoprenoid metabolism was performed by Wallach in 1887 [2]. Wallach is credited for the “isoprene rule” which states that isoprenoids are biosynthesized by the condensations of isoprene (C_5H_8) units (Figure 1). In 1924, Robinson proposed that these isoprene units are condensed by head to tail addition. Two independent groups [3, 4] later discovered that IPP and its isomer dimethylallyl diphosphate (DMAPP) are the true biological precursors of all isoprenoids. These groups found that IPP can be produced by decarboxylating mevalonic acid, a primary metabolite that was first discovered in bacterial and fungal cultures. The isoprenoid biosynthetic pathway is discussed in more detail in

section 1.2 of this thesis.

Figure 1. Chemical structure of isoprene



Polyketides are biosynthesized by the condensation of acetyl-CoA units. Acetyl-CoA is carboxylated with biotin as a carrier, forming malonyl-CoA. The decarboxylative condensation of these malonyl-CoA units creates a poly-β-keto chain of a desired length, which is further modified by cyclization or by the addition of various functional groups such as alcohols and allylic and saturated alkanes. Notable polyketides include prostaglandins, macrolide antibiotics (e.g., erythromycin), large aromatics (e.g., tetracyclines), and small aromatics (e.g., aflatoxin) [5].

Phenylpropanoids are primarily derived from the aromatic amino acid phenylalanine. The skeleton of these secondary metabolites is composed of a six carbon phenolic ring and a three carbon propane tail that are supplied by the phenylalanine backbone. At the entry point of phenylpropanoid metabolism, phenylalanine ammonia lyase catalyzes the non-oxidative deamination of phenylalanine to yield cinnamic acid. Cinnamic acid is further oxidized and activated by cinnamate-4-hydroxylase and coenzyme A (CoA) ligase, respectively,

to supply a carbon source for various downstream phenolic compounds. One notable phenolic compound in nature is lignin, which constitutes the second most abundant bio-polymer after cellulose. Other phenolic metabolites, such as flavonoids, isoflavonoids, and tannins, function as defenses against herbivory, floral pigments, and scent compounds [6].

The alkaloids are the largest group of amino acid derived secondary metabolites, with many thousands of members. Winterstein and Trier [7] were the first researchers to suggest that amino acids function as alkaloid precursors. Robinson [8] expanded on this suggestion, hypothesizing several alkaloid biosynthetic pathways that used amino acids as a substrate. Support for this hypothesis came in the form of tracer studies that were performed mainly in the 1960's, more than 50 years later [9]. It is now known that the products of the shikimate and polyketide pathways are used as precursors in alkaloid biosynthetic pathways [1]. Nitrogen is introduced into these metabolites by transamination, or is provided by an amino acid. Plants are believed to have evolved the ability to produce alkaloids in order to protect themselves from herbivory and insect attack.

Many alkaloids are used as medications and drugs. Morphine is a potent analgesic that is also the most abundant alkaloid in opium poppy (*Papaver somniferum)* sap [10]. The coca plant (*Erythoxylum coca*), produces the alkaloid cocaine that acts as a central nervous system stimulant [11]. Morphine is derived from the amino acid L-tyrosine, and cocaine from the amino acid L-glutamine,

respectively. There are also several groups of non-amino acid alkaloids, like the central nervous system stimulant caffeine that is biosynthesized using the purine xanthosine [12]. Alkaloids are mainly biosynthesized by plants, but some novel alkaloids are also observed in animals and microbes. *Mechanitis polymnia* butterflies, for example, are able to convert the alkaloids rinderine and senecioylretronecine into their N-oxides and other alkaloid derivatives [13].

Specialized metabolites have a rich history of chemical investigation and *in vivo* evaluation as health products (e.g., pharmaceuticals and nutraceuticals). However, due to the complexity of these metabolites and their metabolic pathways, significant progress in biochemical and molecular genetic studies of specialized metabolisms has only been made in the last decade. In particular, high-throughput sequencing improvements have recently played critical roles in understanding specialized metabolisms in many non-model medicinal plants.

1.2 Isoprenoid biosynthesis

Isoprenoids are an extremely large and diverse family of plant metabolites with over fifty thousand known structures [14]. Isoprenoids have a wide range of functions, with some acting as essential primary metabolites and others acting as secondary metabolites. Isoprenoids play roles as membrane components (sterols), protein regulators (prenylation), photosythetic pigments (carotenoids), phytoalexins, and phytohormones (giberrellins, brassinosteroids, abscisic acid, and cytokinins). Members of this metabolite family have a wide range of medical uses,

including as anti-microbials, anti-cancer drugs (paclitaxel), anti-malarials (artemisinin), anti-HIV drugs (cananolide A), and analgesics [15]. Isoprenoids contribute to many natural smells and tastes, and are widely used as commercial flavor and smell enhancers [16]. Some isoprenoids have also been shown to have an effect on the central nervous system. Isoprenoids isolated from guava plants, for example, were found to have a relaxant effect on mice [17].

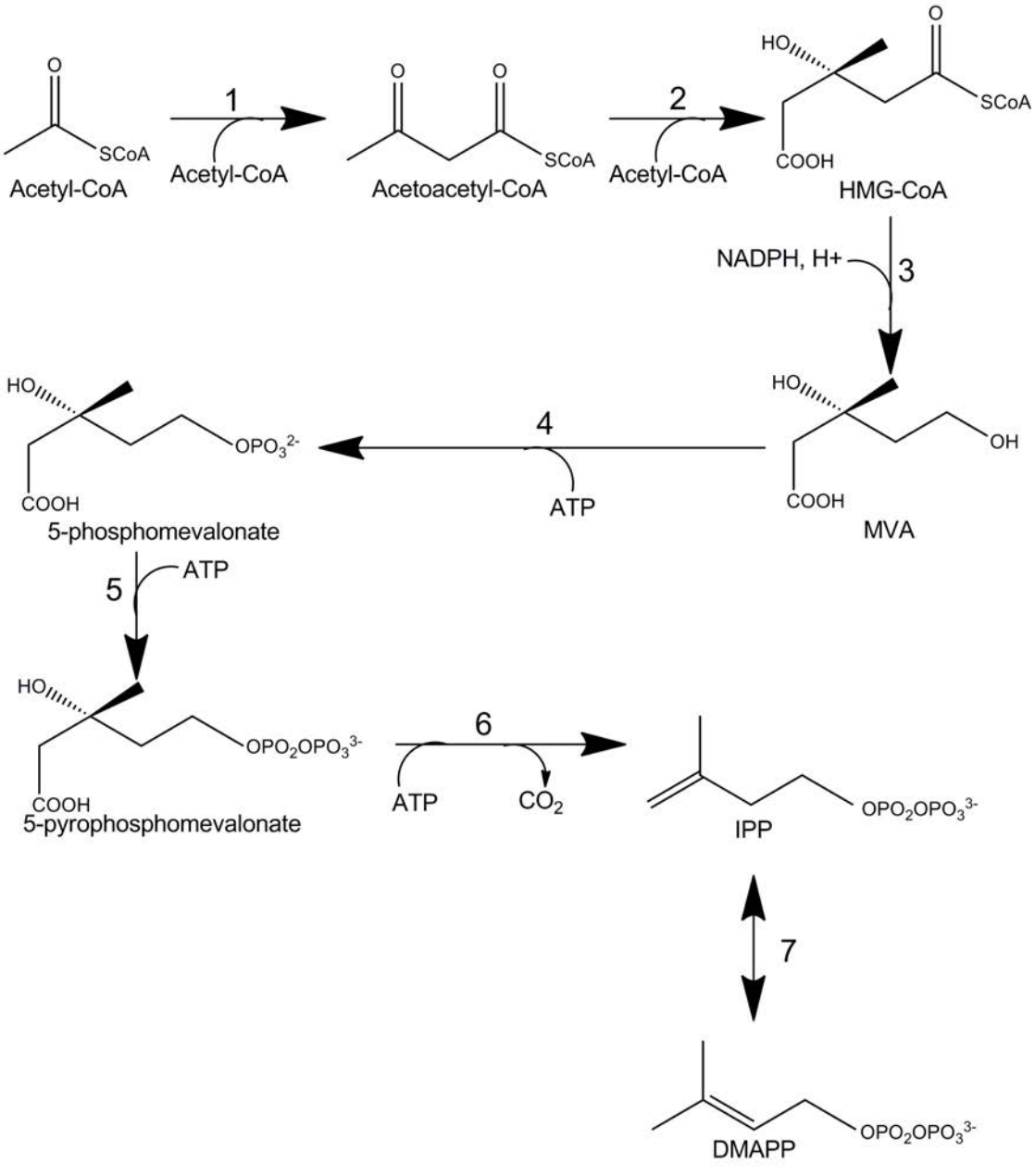
The structure of all isoprenoids is composed of a series of homologous five carbon isoprene units. These compounds are classified by the number of carbons in their skeleton, with monoterpenes having ten carbons, sesquiterpenes fifteen carbons, diterpenes twenty carbons, triterpenes thirty carbons and so on. Geranyl diphosphate (GPP), the ten carbon precursor of the monoterpenes, is biosynthesized by the head to tail condensation of IPP to its isomer DMAPP [18]. The addition of IPP to enzyme bound GPP forms the larger fifteen carbon sesquiterpene precursor farnesyl diphosphate (FPP). Addition of IPP to FPP is then used to create the twenty carbon diterpene precursor, geranylgeranyl diphosphate (GGPP). GPP, FPP, and GGPP are cyclized by monoterpene, sesquiterpene or diterpene synthases, producing the ten carbon monoterpene, fifteen carbon sesquiterpene and twenty carbon diterpene skeletons, respectively. Structural modification of these skeletons by enzymes *in planta* creates the wide array of isoprenoids in nature.

Folkers *et al.* [3] and Tamura [4] were the first groups to discover the

mevalonate (MVA) pathway, a now well-established biosynthetic pathway that uses the decarboxylation of mevalonic acid to produce IPP and DMAPP. In the MVA pathway, three acetyl-CoA are condensed, forming hydroxyl-methyl-glutaryl-CoA (HMG-CoA) [19]. Reduction of HMG-CoA by HMG-CoA reductase then forms MVA. HMG-CoA reductase is a highly conserved enzyme that catalyzes the rate limiting step in the MVA pathway [20]. MVA is phosphorylated twice and decarboxylated once to form IPP. The enzyme IPP isomerase then reversibly converts IPP into DMAPP, maintaining a balance between the two compounds. The MVA pathway is shown in detail in Figure 2.

Figure 2. The MVA pathway of isoprenoid biosynthesis

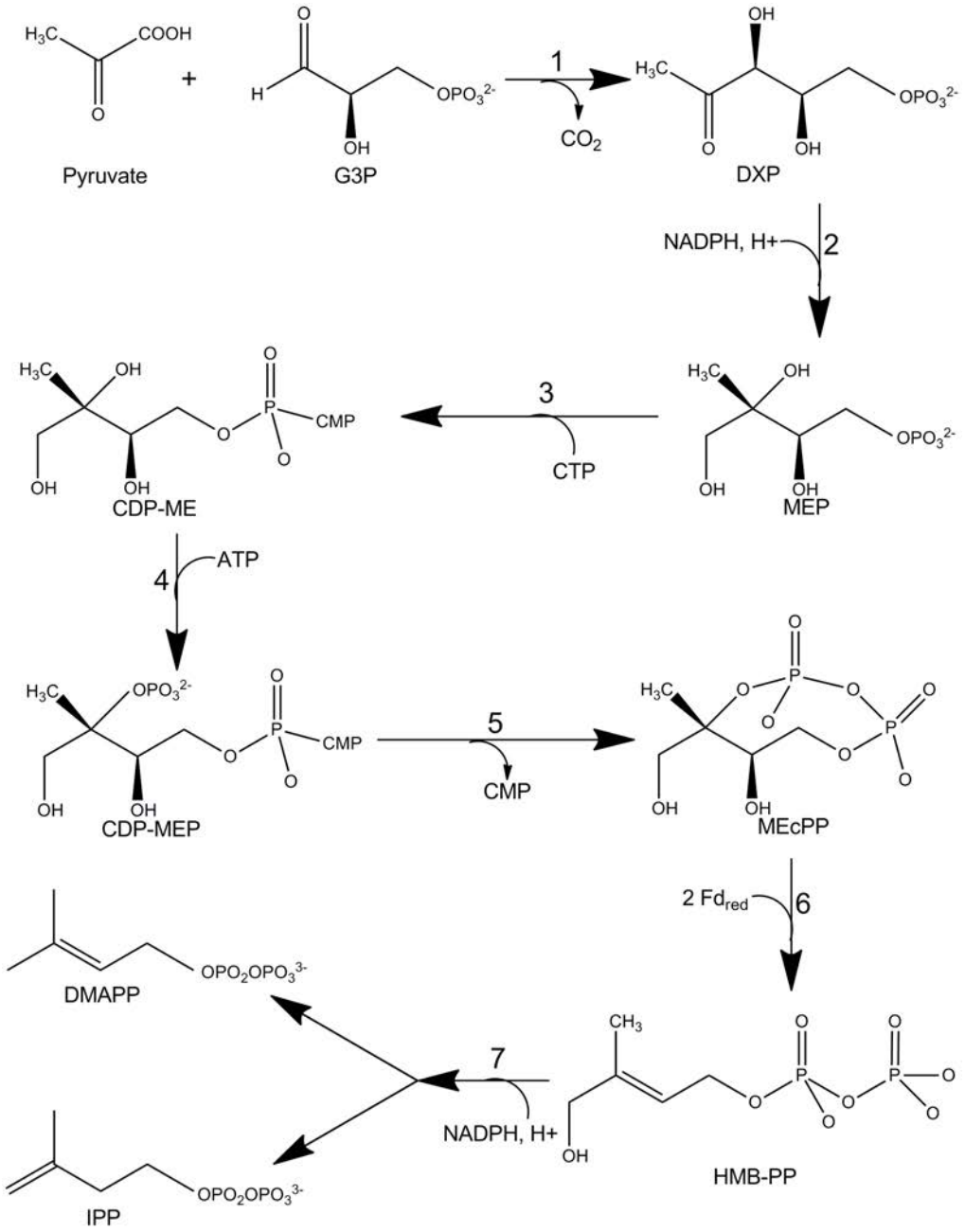
1 = Acetoacetyl-CoA ligase, 2 = HMG-CoA synthase, 3 = HMG-CoA reductase, 4 = MVA kinase, 5 = Phosphomevalonate kinase, 6 = Pyrophosphomevalonate decarboxylase, 7 = IPP isomerase.



The MVA pathway that these groups discovered is not the exclusive pathway for IPP supply. Rohmer *et al.* [21] demonstrated that a MVA-independent pathway is present in bacteria. His group used ^{13}C labeled glucose, acetate and pyruvate to show that bacterial hopanoids were not synthesized from MVA pathway-derived IPP. There are in fact two biosynthetic routes of IPP production - the MVA pathway and the deoxy-xylulose-5-phosphate (DXP) pathway. The DXP pathway is also referred to as the methyl erythritol phosphate (MEP), non-mevalonate, or Rohmer pathway. The DXP pathway was only fully deciphered in 2000, and it will be perhaps the last hidden metabolic pathway conserved across different kingdoms. The metabolic details of the DXP pathway are depicted in Figure 3. In this metabolic pathway, glyceraldehyde-3-phosphate (G3P) is condensed with activated acetaldehyde derived from pyruvate by DXP synthase, forming DXP [22]. DXP is then rearranged and reduced by DXP reductoisomerase, forming MEP. C-methyl-D-erythritol-ME (CDP-ME) synthase in the presence of cytidine triphosphate converts MEP into CDP-ME, and CDP-ME kinase phosphorylates the compound, forming 4-diphosphocytidyl-2-C-methyl-D-erythritol 2-phosphate (CDP-MEP). CDP-MEP is cyclized into 2-C-methyl-D-erythritol 2,4-cyclopyrophosphate (MEcPP) by MEcPP synthase, and MEcPP is then converted into (E)-4-Hydroxy-3-methyl-but-2-enyl diphosphate (HMBPP) by HMBPP synthase. Finally, HMBPP reductase catalyzes the conversion of HMBPP into IPP and DMAPP in a five to one ratio. As DMAPP is produced by this step, IPP isomerase is not necessary and a large number of bacterial species lack an IPP isomerase gene as a result.

Figure 3. The DXP pathway of isoprenoid biosynthesis

1 = DXP synthase, 2 = DXP reductase, 3 = CDP-ME synthase, 4 = CDP-ME kinase, 5 = MeCPP synthase, 6 = HMB-PP synthase, 7 = HMB-PP reductase. CTP = Cytidine triphosphate, CMP = Cytidine monophosphate.



Both the MVA and DXP pathways are present in plants, but the pathways are compartmentalized and entirely different precursor pools are used to produce IPP and DMAPP. MVA pathway enzymes are located in the cytosol of plant cells, whereas DXP pathway enzymes function in the plastid [23]. Isoprenoid biosynthesis in plants is also compartmentalized, with monoterpenes and diterpenes usually produced by the plastidic DXP pathway, and sesquiterpenes and triterpenes usually produced by the cytosolic MVA pathway. However, recent evidence suggests that there is metabolic cross talk between the two pathways. Isolated plastids are able to efficiently transport IPP, DMAPP, GPP, and FPP, suggesting the existence of a plastidic proton symport [24]. Transport is uni-directional from the plastid to the cytoplasm, and no transport of GGPP or MVA was observed. Metabolic flux between the MVA and DXP pathways has also been observed *in planta* in *Arabidopsis thaliana.* 48 hr after the addition of an MVA pathway inhibitor, sterol triterpenoid concentrations in inhibited plants were revealed to be higher than in control plants [25]. Isoprenoid biosynthetic pathway transcript levels did not change in the MVA pathway inhibited plants, indicating that some transport of sterol precursors from the plastid to the cytosol must be occuring.

1.3 *Salvia divinorum*

Salvia divinorum, or diviner's sage [26], is a hallucinogenic perennial herb that is indigenous to the Oaxaca region of Mexico [27]. *S. divinorum* prefers a humid environment and grows primarily in the moist forest ravines of the Oaxaca region. *S. divinorum* is a member of the *Lamiaceae* family, which is in order *Lamiales*.

There are approximately seven to nine hundred known species of *Salvia,* all of which are annuals, herbaceaous perennials, or shrubs [28]. Of these *Salvia* species, *S. divinorum* is the only known hallucinogen. Characteristic features of the plant include large thick green leaves, hollow square stems, and white flowers with purple calyces. *S. divinorum* reproduces primarily by vegetative cloning, as the plant rarely flowers or produces seeds [29]. Larger plants collapse under their own weight (a trait called lodging), and offspring sprout where the collapsed plants meet the forest floor. This method of reproduction is rarely observed in other species of *Salvia*.

1.4 Traditional and modern uses of *S. divinorum*

S. divinorum is used by the Mazatecs of Mexico in spiritual divination [30]. The mazatecs use several different hallucinogens in divination rituals, including *S. divinorum, Psilocybe cubensis* mushrooms, and morning glory seeds [29]. Mazatec folk healers primarily use *S. divinorum* to facilitate spiritual communion. Religion is united closely with healing in Mazatec culture, and this spiritual communion is used to find cures to diseases, and to foretell the future [27] . The mazatecs also use *S. divinorum* as an analgesic, and to treat several ailments including rheumatism, diarrhea, headache, and “swollen belly”, a disease that is supposedly caused by a sorceror's curse [31]. *S. divinorum* is very sacred to the Mazatecs, as they believe that the plant is an earthly incarnation of the Virgin Mary. Reflecting this, common names for *S. divinorum* include ska Maria Pastoris and la Maria. Traditionally, Mazatecs consume the plant by chewing fresh leaves as a quid, or by drinking a

crushed leaf extract. When consumed in this way, dosages range from twenty to two hundred leaves [32].

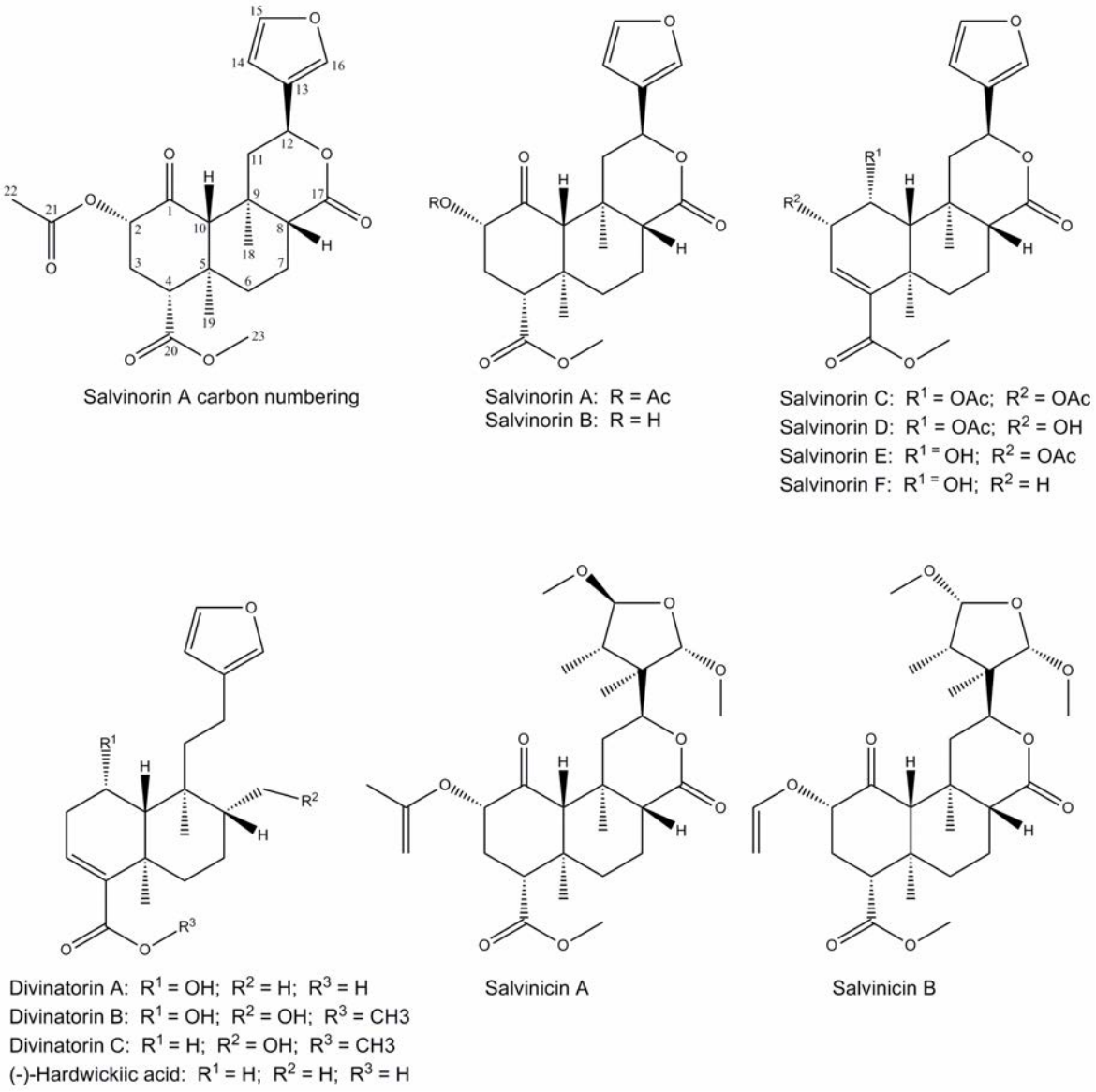
In the United States and Canada, *S. divinorum* is mainly used as a recreational hallucinogen [30]. Leaves are dried, concentrated, and pyrolized, and the resulting smoke is inhaled. *S. divinorum* is currently not a controlled substance in Canada, and in many American states and European countries. Delaware, Illinois, Lousiana and Missouri, however, have passed laws regulating the use of *S. divinorum*, and several other states have proposed legislation against the plant [33].

1.5 Salvinorin A

The diterpenoid salvinorin A is the active hallucinogenic component of *S. divinorum.* The structure of Salvinorin A was first elucidated in 1982 [34]. The hallucinogen was purified from *S. divinorum* leaf extracts, and X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy techniques were utilized to elucidate its structure. Salvinorin A was found to be a neoclerodane diterpene. Structural diagrams of salvinorin A and other *S. divinorum* secondary metabolites are shown in Figure 4. Neoclerodane compounds are rearranged labdane molecules that minimally contain a bicyclic hydrocarbon ring (clerodane moiety) and quarternary carbon (neo- prefix). Pure salvinorin A was first used in bioassays 1 year later, in 1983 [31] . Mice were exposed to fractionated *S. divinorum* extracts, and the fraction containing salvinorin A was revealed to have a sedative-like effect on the animals. The effects

of the compound on humans were first recorded in a 1994 study that had test subjects inhale the pyrolized vapors of pure salvinorin A [35]. Vivid and realistic hallucinations were reported by all of the test subjects. Total chemical synthesis of salvinorin A was a relatively recent accomplishment, as the compound has several oxygenated functionalities and chiral carbons that make it difficult to synthesize. An asymmetric approach that involves thirty three steps was eventually developed in 2007 [36].

Figure 4. Chemical structures of *S. divinorum* secondary metabolites



1.6 Pharmacology of salvinorin A

The hallucinogenic effects of *S. divinorum* have an extremely quick onset, and can last from 30 min to 2 hr [35]. Salvinorin A was administered to baboons to evaluate its *in vivo* efficacy, and it was found that the drug has an extremely rapid

brain uptake time when smoked. ~3% of the dosage used was able to pass the blood brain barrier within 40 sec [37]. Salvinorin A was also found to clear the brain quickly, with a half-life of 8 min. These results directly mirror the visual hallucination time-course of the drug that is reported by human users [35]. Salvinorin A is an extremely potent hallucinogen, as less than 10 µg in the human brain can have a significant psychoactive effect [37]. When pyrolized salvinorin A is inhaled, the threshold dose is 200 µg, the same range as lysergic acid diethylamide [35].

Common effects of *S. divinorum* use include the sensation of movement, uncontrollable laughter, and visions of people and places [35]. When doses higher than 500 µg are pyrolized and inhaled, depersonalization often occurs, and users become unaware of their surroundings. The drug also has an incapacitating effect, and generally causes some degree of anxiety as a result [38]. When users of psychedelics were polled on their experiences, higher perception and reality modifying effects were reported for salvinorin A than for other hallucinogenic drugs [38]. In general, salvinorin A users report that they experience a modified perception of reality and an inability to interact with themselves or their surroundings. This experience is often described as a type of “spatiotemporal dislocation”.

Salvinorin A has also been shown to have mood altering effects. Administration of the drug to rats significantly and rapidly sedated the animals, resulting in motor incoordination and a disruption in climbing behaviour [39]. The

forced swim and intracranial self-stimulation tests (common methods of measuring depression) were used to show that these depressive effects are dose-dependent [40]. The addictiveness of the drug in rats has also been assayed using conditioned place aversion tests (a common method of assaying addiction/aversion). Rats that were exposed to high doses of salvinorin A were significantly more likely to avoid it in the future, indicating that the drug is aversive in nature [41]. This increase in conditioned place aversion has also been reported in mice that were treated with salvinorin A [42]. Salvinorin A was found to lower striatal dopamine levels in the mice, and cause sedative-like symptoms. This decrease in striatal dopamine is likely the reason that high doses of salvinorin A result in aversive behavior.

The mood altering effects of salvinorin A are not entirely negative; some positive effects have also been noted. An increase in swimming speed was observed when a threshold dose of salvinorin A was administered to zebrafish, indicating that low doses of the drug can have a rewarding effect [43]. A similar effect was also observed in rats [41]. Rats that were treated with threshold doses of salvinorin A experienced a significant increase in extracellular dopamine levels. The rats were in turn more likely to seek out salvinorin A than they were to avoid it. Pre-treating the zebrafish and rats with a cannabinoid type 1 receptor antagonist significantly reduced this stimulatory effect in both cases. These positive mood effects are therefore likely a result of cannabinoid type 1 receptors directly modulating the release of dopamine [43]. Salvinorin A's method of activating the cannabinoid type 1 receptor is thus far poorly understood.

1.7 Pharmacokinetics of salvinorin A

Salvinorin A causes these hallucinogenic and mood-altering effects primarily by selectively binding to and activating the κ (kappa) opioid receptor (KOR), acting as an extremely potent agonist. The KOR is a G protein-coupled receptor in the opioid receptor family. Salvinorin A`s binding affinity was screened against an array of fifty recombinant human G protein-coupled receptors, transporters, and ligand-gated ion channels, and the drug was found to significantly bind to only the recombinant KOR [44]. Ligand binding assays using KOR expressed in human embryonic kidney cells and *in situ* in guinea pig brains were also performed, and salvinorin A again acted only as a potent KOR agonist in both cases [44]. Salvinorin A's receptor binding profile is not observed in other hallucinogens such as psilocybin and mescaline. These classical hallucinogens bind a wide array of receptors, with their primary molecular target being the 5-hydroxytryptamine 2A receptor [45]. To date no interaction between the 5-hydroxytryptamine 2A receptor and salvinorin A has been reported.

Salvinorin A is metabolized and excreted primarily by the renal and biliary systems [37]. These systems hydrolyze the salvinorin A C2 acetyl group, converting the compound into salvinorin B (Figure 4). When rat brain and liver organelle fractions were incubated with salvinorin A, salvinorin B was the only metabolic product observed [46]. These results indicate that conversion into salvinorin B is the primary method of salvinorin A metabolism in animals. Monkey plasma was used to perform salvinorin A metabolism assays, and it was found that blood esterases are

responsible for this rapid hydrolyzation of salvinorin A [47]. Siebert [48] pyrolized and inhaled pure salvinorin B, and reported that the compound does not cause any psychotomimetic effects in humans. The short half-life of salvinorin A in the human body therefore appears to be due to the rapid inactivation of the drug by hydrolysis of its C2 acetyl group.

1.8 Opioid receptor types and functions

There are three major types of opioid receptor in the human body – κ (kappa), μ (mu), and δ (delta). Salvinorin A is not the only exogenous compound that binds the KOR; several opioids also strongly bind this type of receptor. Opioids are defined as compounds that cause physiological or pharmacological effects that are similar to those caused by morphine. Opioids are not necessarily opiates, and do not necessarily have a morphine-like structure. These compounds mediate physiological effects by causing neurotransmitters to be released into the synapse, or by inhibiting the electrical activity of the neuron. Major effects of opioids include analgesia, respiratory depression, and sedation. Medically, opioids are primarily used as analgesics [49].

KOR activation in humans can result in a wide array of effects, including dysphoria, hallucinations, analgesia, and sedation [50]. Several opioids, including endogenous dynorphin and exogenous salvinorin A bind the KOR. Major effects of μ opioid receptor (MOR) activation include euphoria, analgesia, and respiratory depression [50]. Endogenous β-endolphine and exogenous morphine are among

the opioids that primarily bind the MOR. Activation of δ opioid receptors (DOR) can cause analgesia and the inhibition of smooth muscle [50]. DOR binding opioids include endogenous met (methionine containing)-enkephalin and exogenous etrophine. Salvinorin A is highly specific to the KOR, and does not bind significantly to the DOR or MOR [44].

In humans, emotional situations have been found to cause the activation of opioid receptors. The MOR and KOR work together to regulate emotional reactions; activation of the MOR causes euphoric effects, whereas activation of the KOR results in dysphoria [51]. Modulation of KOR signaling can also regulate human perception. Activation of the receptor by the selective agonist enadoline results in visual distortions, feelings of dissociation, and dysphoria [52]. Some opioids, such as the benzomorphans cyclazocine and N-allyl-normetazocine, mediate psychotomimetic and dysphoric effects by binding and activating the KOR [51]. The hallucinogenic effects of known KOR agonists indicate that salvinorin A likely also mediates hallucinations by activating the KOR. Salvinorin A is unique in that it is the only natural non-nitrogenous KOR agonist to cause psychotomimetic effects.

1.9 Biosynthesis of salvinorin A by *S. divinorum*

Salvinorn A is stored in the peltate glandular trichomes of *S. divinorum.* Trichomes are small, microscopic structures that emerge from the epidermis of many plant organs. There are two major types of trichome – non-glandular and

glandular. Both types of trichome are found on many plants in the *Lamiaceae* family, including *S. divinorum* [53]. Non-glandular trichomes are hair-like structures that cover a large number of plant organs, including the stem and the abaxial (bottom) and adaxial (top) surface of the leaves. These trichomes are composed of four to ten unbranched cells that extend linearly from a basal epidermal cell. *S. divinorum* non-glandular trichomes can be curved or straight, and are usually 50 to 200 μM long [48]. Non-glandular trichomes likely serve to limit the mechanical movement of insects on the surface of the plant [54].

Glandular trichomes are plant epidermal organs that biosynthesize and secrete secondary metabolites. Terpenes are usually a major component of this secondary metabolite secretion [55]. There are two types of glandular trichome – capitate and peltate. Capitate glandular trichomes are made up of a basal epidermal cell, one or more stalk cells, and one or two head cells. Long stalked capitate glandular trichomes also have a neck cell that separates the stalk cell from the head cells. In *S. divinorum, c*apitate glandular trichomes are found primarily on pedicles, bracts, calyces and leaves [48]. Capitate glandular trichomes secrete the secondary metabolites they produce into the surrounding environment [56].

Peltate glandular trichomes are typically composed of six cells; a basal cell that connects the trichome to the leaf epidermis, four secretory cells that biosynthesize and secrete secondary metabolites, and a stalk cell that connects the secretory cells to the basal cell. Secondary metabolites are secreted into the

cuticular sac that surrounds the peltate glandular trichomes [53]. In *S. divinorum,* this cuticular sac has four lobes, and is usually 35 to 45 μM in diameter when fully developed [48]. Peltate glandular trichomes develop on *S. divinorum* stems, rachises, bracts, pedicles, calyces, and the abaxial surface of the leaves. This type of glandular trichome is notably absent from the adaxial surface of the leaves. New trichomes initiate and mature as *Salvia spp.* leaves age, but most of a leaf's trichomes initiate when the leaf is very young. Peltate glandular trichome density is highest in young *S. divinorum* leaves as a result. The abaxial surface of mature *S. divinorum* leaves and stems has an average density of 1,600 peltate glandular trichomes per square inch [48].

The metabolite profile of *S. divinorum* peltate glandular trichomes was analyzed, and it was revealed that these structures are the storage site for salvinorin A. As salvinorin A content per g of leaf tissue remains relatively constant as leaves age and peltate glandular trichome density decreases, it is likely that salvinorin A accumulates in peltate glandular trichome subcuticular sacs as time passes [48]. Salvinorin A concentrations ranging from 0.89 to 3.70 mg/g were observed in leaves from a selection of common cloned and seed grown *S. divinorum* plants [57]. Salvinorin A is the most abundant secondary metabolite in *S. divinorum* leaves, composing up to 0.3% of each leaf's dry weight. *S. divinorum* likely produces salvinorin A as an anti-herbivory defense. Herbivory by the deer that reside in the Oaxaca region would rupture the peltate glandular trichome subcuticular sacs of the plant, releasing stored salvinorin A.

1.10 Methyl jasmonate induction of glandular trichome density

Methyl jasmonate is a volatile ester that is derived from the phytohormone jasmonic acid [89]. The chemical serves to regulate several plant developmental processes, such as seed germination and fruit development. Methyl jasmonate also upregulates plant secondary metabolite and defense protein biosynthetic pathways [89]. Abiotic and biotic stresses like pathogens and drought cause plants to produce the chemical. The methyl jasmonate then diffuses through the air, signalling the activation of defense pathways in the original plant and its neighbours [89]. Methyl jasmonate can also upregulate genes that are involved in trichome biosynthesis. Peltate glandular trichome density was found to be ninefold higher on the leaves of methyl jasmonate treated tomato plants than on control plants 14 days after exposure to the chemical [90].

1.11 Other *S. divinorum* secondary metabolites

Salvinorin A is not the only diterpenoid stored in *S. divinorum* peltate glandular trichomes. Salvinorins C to F, divinatorins A to C, (-)-hardiwickiic acid, and salvinicins A and B are also found in the herbs leaves (Figure 4) [48, 58]. Branch points in the salvinorin A pathway and the non-enzymatic degradation of the diterpenoid likely creates these other secondary metabolites. Of these salvinorin A derivatives, salvinorins B, C, and D are most abundant. The receptor binding profile of these other metabolites was assayed, and it was revealed that only salvinorin C is able to bind the KOR, and it does so very weakly [59] . Salvinorin C has thus far not been shown to cause any psychotomimetic effects in humans.

The concentration of salvinorin C in *S. divinorum* leaves increases significantly as the leaves age, indicating that the compound may play a role in anti-herbivory [48].

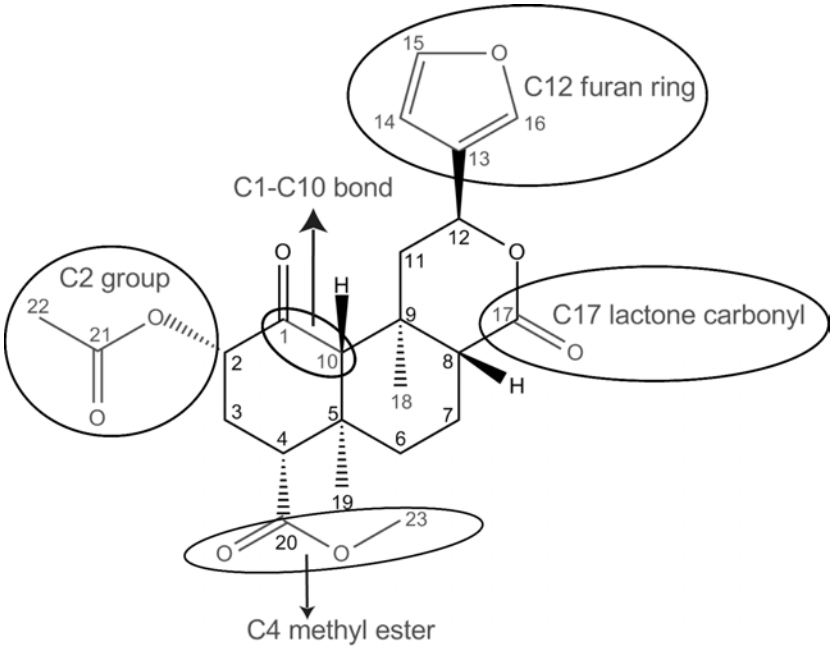
1.12 Functional groups of salvinorin A

A wide array of chemical modifications of salvinorin A have been performed, in an effort to better understand the structure-function relations of the drugs side-groups. The carbon numbering of salvinorin A is depicted in Figure 4. Salvinorin A's C4 carbomethoxy and C12 furan ring appear to be critical for KOR activity; removal or modification of either of the groups significantly decreases the compounds KOR binding affinity [60]. The size of salvinorin A's C2 substituent is also an important factor in determining the potency of the compound. When side-groups smaller or larger than a three atom branched chain are substituted into this position, a significant loss of activity is observed [61]. The metabolites C17 lactone carbonyl also aids in KOR affinity, but not as stringently. Removal or modification of this group only weakly decreases KOR binding. These key functional groups are shown in Figure 5.

Given that salvinorin A metabolism involves the hydrolyzation of the compounds C2 acetoxy moiety, it follows that the metabolic stability of the drug can be increased by substituting different groups at this position. Substituting the C2 group with a carbamate, for example, results in a compound that is nearly as potent as salvinorin A, and significantly more resistant to hydrolyzation [61]. Interestingly, substitution of aromatic esters at the C2 position greatly increases

affinity for the MOR, but also greatly decreases affinity for the KOR [60]. Some substitutions were found to raise affinity at the DOR as well; KOR, MOR and DOR antagonist activity is significantly increased when a 1,10-alkene is introduced at the C1 position [60]. Modification of salvinorin A can therefore be used to increase not only KOR affinity, but affinity and antagonistic activity at the DOR and MOR.

Figure 5. Functional groups of salvinorin A



1.13 Potential medical uses of salvinorin A

Salvinorin A's unique receptor binding profile and singular potency makes it potentially useful as a medicine, and the drugs applications in this regard have been extensively studied. Salvinorin A's affinity for the KOR is similar to that of

dynorphin A, an endogenous KOR agonist that is a known analgesic [62]. KOR agonists are potent regulators of antinociception, but the severe side effects they cause (sedation, diuresis, and psychotomimosis) usually limit their use in medicine. Salvinorin A's use as an analgesic was investigated by administering it to mice, and using tail flick response to gauge its effects on the perception of pain. The drug was observed to cause time- and dose-dependent antinociception that diminished rapidly over the course of 20 min [63]. Given that salvinorin B is not an analgesic [64], this short duration is likely a result of the rapid metabolic conversion of salvinorin A into salvinorin B [61]. In addition to its antinociceptive activity, salvinorin A also exhibits body temperature reducing effects [64]. No analgesic or temperature reducing effects were noted in KOR knockout mice that were treated with salvinorin A, indicating that these effects are largely a result of the drugs interaction with the KOR [64]. It is possible that modification of salvinorin A can be used to decrease its psychotomimetic side effects, making it a useful fever treatment.

In addition to its potential as an analgesic, salvinorin A's receptor binding profile grants the drug several other pharmacological properties. Studies in rodents, for example, demonstrate that the drug can be used as a diarrhea treatment. Administration of salvinorin A was shown to prevent contraction of the guineau pig ileum [65], and to inhibit intestinal motility in mice [66]. Inflammation of the gut was also revealed to significantly increase salvinorin A's ability to reduce intestinal motility, whereas the potency of a synthetic KOR agonist was not increased under the same conditions [66]. Salvinorin A likely has targets other than

the KOR in the gut, making it an especially potent diarrhea treatment. Behavioural studies involving salvinorin A show that the drug is also useful in the treatment of drug addictions. Monkeys that were physically dependent on cocaine experienced less cravings for the drug when treated with salvinorin A [67]. A similar effect was observed in rats, where administration of salvinorin A resulted in a reduction in the locomotor-stimulant effects of cocaine [68]. It is possible then that certain salvinorin A derivatives can be used to reduce the symptoms of cocaine addiction in humans.

Salvinorin A's receptor binding profile also makes it an excellent candidate for the treatment of mood alteration and hallucination associated disorders. Modulation of KOR signalling can have a profound effect on mood. Rats that are exposed to KOR antagonists or KOR agonists experience anti-depressive [69] or depressive symptoms [70], respectively . No large scale studies of the effects of salvinorin A on human mood disorders have been performed, but in one case study a woman was able to successfully treat her refractory depression by self-medicating with *S. divinorum* leaves [71]. Modulation of KOR signaling can also be used to treat disorders that are characterized by vivid hallucinations, like schizophrenia and Alzheimer's disease. Schizophrenic patients that were treated with a KOR antagonist experienced significant improvements in mental deterioration and social withdrawal with respect to patients who received a placebo [72]. KORs are significantly more abundant in the brains of patients with Alzheimer's disease, and KOR antagonists may therefore be useful in reducing the symptoms of this disease as well [73]. Given that activation of the KOR can cause

hallucinations, salvinorin A derivatives that repress this receptor class will likely be effective tools in the treatment of hallucination associated disorders.

1.14 Toxicology of salvinorin A

The toxic effects of salvinorin A must be quantified if the compound is to be safely used as a medicine. Classical hallucinogens (that bind the 5-hydroxytryptamine 2A receptor) are physiologically safe, and do not cause addiction or dependence [74]. These hallucinogens do not interact primarily with dopamine or dopamine receptors, and do not cause an addictive euphoric high as a result. Toxicology studies show that classical hallucinogens do not cause damage to the body or mind, even at extremely high doses. This trend is not necessarily true for salvinorin A, and studies involving rodents were performed in an effort to measure the toxicity of the drug. Rats and mice were treated with chronic (2 week duration) and acute (several times higher than what a human would experience) doses of salvinorin A, and no histologic changes were noted in rodent organs [75]. An increase in blood pressure was observed, but this increase was non-significant. No long term studies on the toxicological effects of salvinorin A have thus far been performed, but given these results it is likely that the compound is safe for human use.

1.15 The salvinorin A biosynthetic pathway

As indicated previously, IPP that is synthesized by either the MVA or DXP metabolic pathway is the key precursor of plant isoprenoids. Feeding studies

involving isotope labeled precursors have revealed that the carbons of salvinorin A are supplied by the plastidic DXP pathway [76]. Other than the carbon supply pathway, the biochemistry of salvinorin A biosynthesis is entirely unknown. No enzymes in the pathway have thus far been identified, including the diterpene synthases that catalyze the first committed steps in salvinorin A biosynthesis. As a result, several hypothetical intermediates in the pathway are proposed here.

Based on the well characterized gibberellin pathway (a diterpenoid plant hormone) [77], a type II and type I diterpene synthase were expected to catalyze the first two salvinorin A pathway enzymatic reactions. The general scheme of type II and type I diterpene synthase reactions (using the gibberellin pathway as an example) is shown in Figure 6. Type II diterpene synthases use protonation of a double-bond to create a carbocation, and then use this carbocation to drive the cyclization and rearrangement of the GGPP substrate [78]. The carbocation is stabilized by deprotonation and double-bond rearrangement (as in the gibberellin pathway), or by reaction with a water molecule, creating a hydroxyl group. Type I diterpene synthases then act on the resultant diphosphate compound, using the divalent cation (e.g., Mg^{2+} or Mn^{2+})-dependent ionization of prenyl diphosphate to create a second carbocation. This carbocation is either quenched immediately or used to drive secondary rearrangements of the compound. The carbocations that type I diterpene synthases create are also stabilized by double-bond rearrangement or by reaction with water. In the gibberellin pathway, this carbocation is used to rearrange the *ent*-copalyl diphosphate (CPP) substrate, and

is then quenched by double-bond rearrangement. In the salvinorin A pathway, the carbons of the type I diterpene synthase product are likely modified by several cytochrome P450s, an acetyl transferase, and a methyl transferase, generating the furan ring, lactone ring, ketone, acetyl group, and methyl ester of the metabolite. This generalized scheme of salvinorin A biosynthesis is shown in Figure 7. The diterpene synthase steps of the pathway are discussed in detail in the following paragraphs.

Figure 6. Diterpene synthase catalyzed reactions in the gibberellin biosynthetic pathway

1 = Reactions catalyzed by *ent*-CPP synthase (CPPS), a type II diterpene synthase. 2 = Reactions catalyzed by *ent*-kaurene synthase (KS), a type I diterpene synthase.

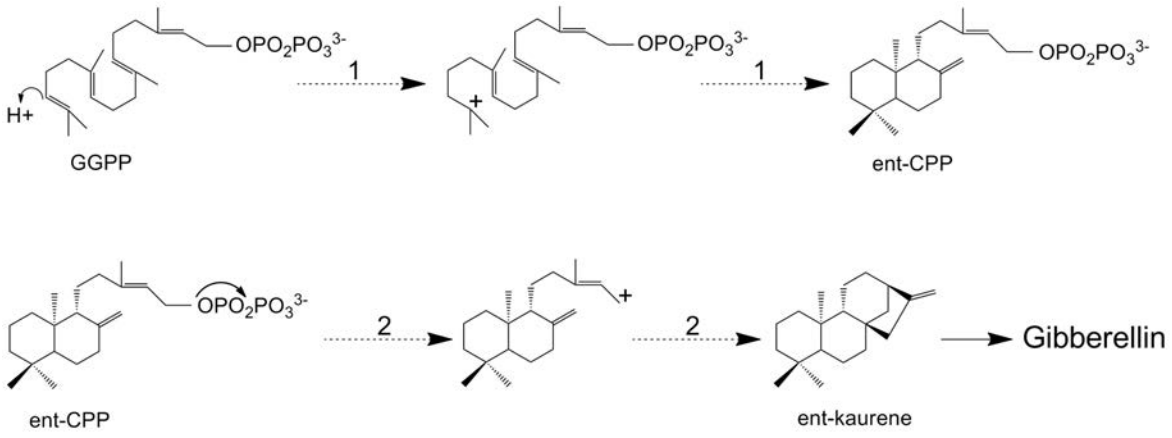
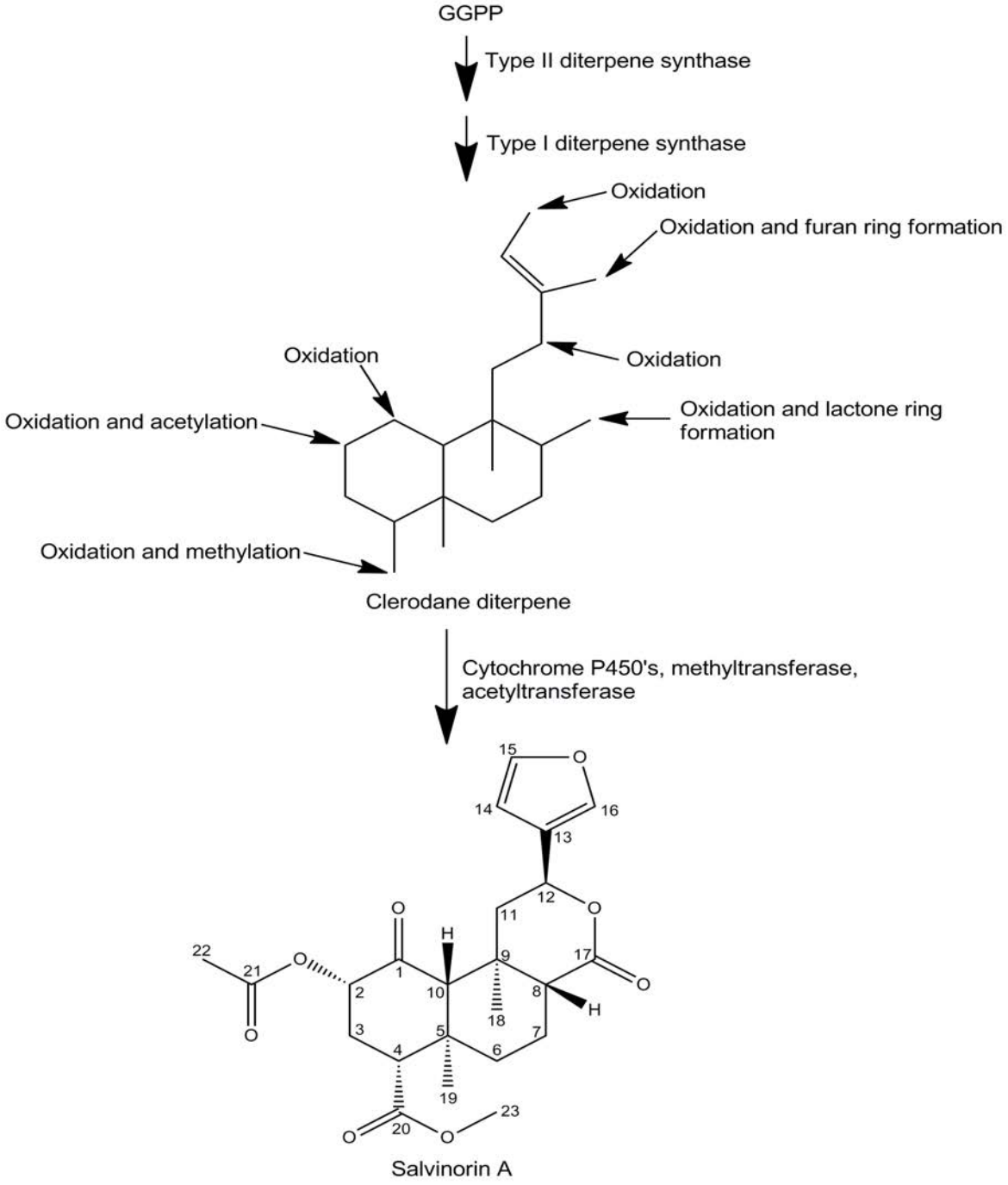


Figure 7. Hypothetical salvinorin A biosynthetic pathway (generalized)



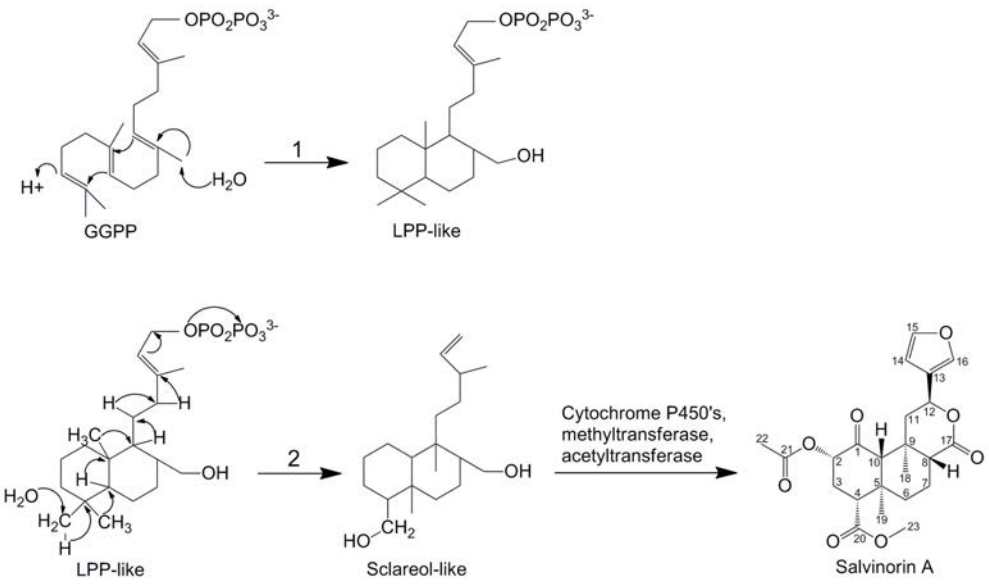
The plant *Salvia sclarea* also produces a clerodane diterpene, the compound sclareol. Sclareol is biosynthesized and secreted by the capitate glandular trichomes of *S. sclarea*, and can be used as a substrate in the chemical synthesis of the perfume fixative ambergris [79]. Given *S. sclarea's* close phylogenetic relation to *S. divinorum*, study of the sclareol pathway can help us better understand the salvinorin A pathway. In the sclareol pathway, GGPP is converted into the labdane diterpene labdiendol diphosphate (LPP) by LPP synthase (LPPS) [80]. Sclareol synthase then converts LPP into the clerodane diterpene sclareol. LPPS is a type II diterpene synthase, and sclareol synthase is a type I diterpene synthase. There are several similarities in the sclareol and hypothesized salvinorin A pathways; type II and type I diterpene synthases operate to convert GGPP into a clerodane diterpene in both cases. It is therefore possible that diterpene synthases that are similar to LPPS and sclareol synthase act in the salvinorin A pathway.

Figure 8 shows the possible reactions these LPPS-like and sclareol synthase-like enzymes would catalyze in the salvinorin A biosynthetic pathway. The LPPS-like enzyme would hypothetically drive the cyclization of GGPP by protonating the compounds C3 double-bond. The resulting carbocation would then be quenched by reaction with water, generating a hydroxyl group. The sclareol synthase-like enzyme would ionize this compound by cleaving the diphosphate, creating a carbocation that would drive a cascade of hydrogen- and methyl-shifts. This second carbocation would also be quenched by reaction with water. The

resulting clerodane diterpene would then be acted on by cytochrome P450's, an acetyltransferase, and a methyltransferase, as in Figure 7.

Figure 8. Hypothetical LPPS-like and sclareol synthase-like reactions in the salvinorin A biosynthetic pathway

1 = Reaction catalyzed by the LPPS-like enzyme. 2 = Reaction catalyzed by the sclareol synthase-like enzyme.

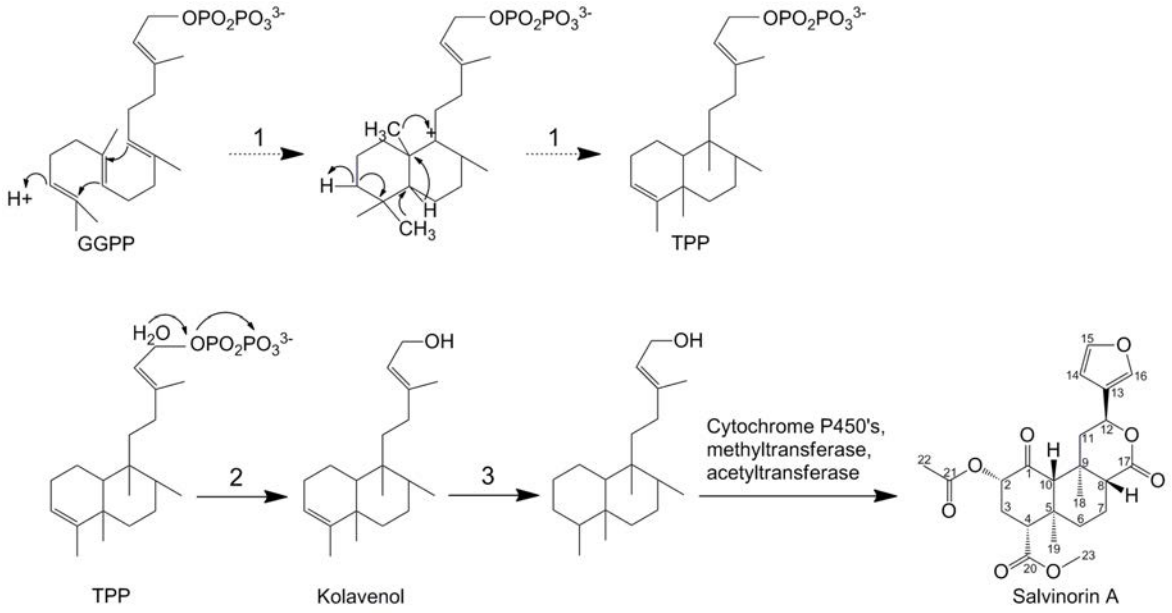


It is also possible that enzymes similar to those in the terpentecin pathway act in the salvinorin A pathway. Terpentecin is a diterpenoid antibiotic found in *Streptomyces griseolosporeus* [81]. In the first step of the terpentecin pathway, terpentedienyl diphosphate synthase (TPPS) (a type II diterpene synthase) catalyzes the conversion of GGPP into terpentedienyl diphosphate (TPP). Terpentriene synthase (a type I diterpene synthase) then converts TPP into

terpentriene. TPP and a compound that is similar to terpentriene are potential intermediates in the salvinorin A biosynthetic pathway, and TPPS and a terpentriene synthase-like enzyme may act in the salvinorin A pathway as a result (Figure 9). As in the previous hypothetical pathway, the TPPS enzyme would first protonate the GGPP C3 double-bond and use the resulting carbocation to drive hydrogen- and methyl-shifts. Deprotonation of C3, creating a double bond to C4, would be used to quench this carbocation. The terpentriene synthase-like enzyme would then ionize the compounds diphosphate, creating a second carbocation that would be immediately quenched by reaction with water. Given the structure of salvinorin A, it is likely that this second carbocation would not used to to drive any cyclizations or rearrangements. The terpentriene-like compound that this type I diterpene synthase would generate is also known as kolavenol [81]. Kolavenol is a secondary metabolite that accumulates in several different plant species [82, 83]. The enzymatic reactions that create kolavenol are thus far poorly understood. This model salvinorin A pathway will generate a double-bond at the C3-C4 position due to deprotonation for carbocation stabilization, and a double-bond reductase would therefore be required to remove the resulting double-bond at some point in the pathway.

Figure 9. Hypothetical TPPS and terpentriene synthase-like reactions in the salvinorin A biosynthetic pathway

1 = Reactions catalyzed by the TPPS enzyme. 2 = Reaction catalyzed by the terpentriene synthase-like enzyme. 3 = Reaction catalyzed by a double bond reductase.

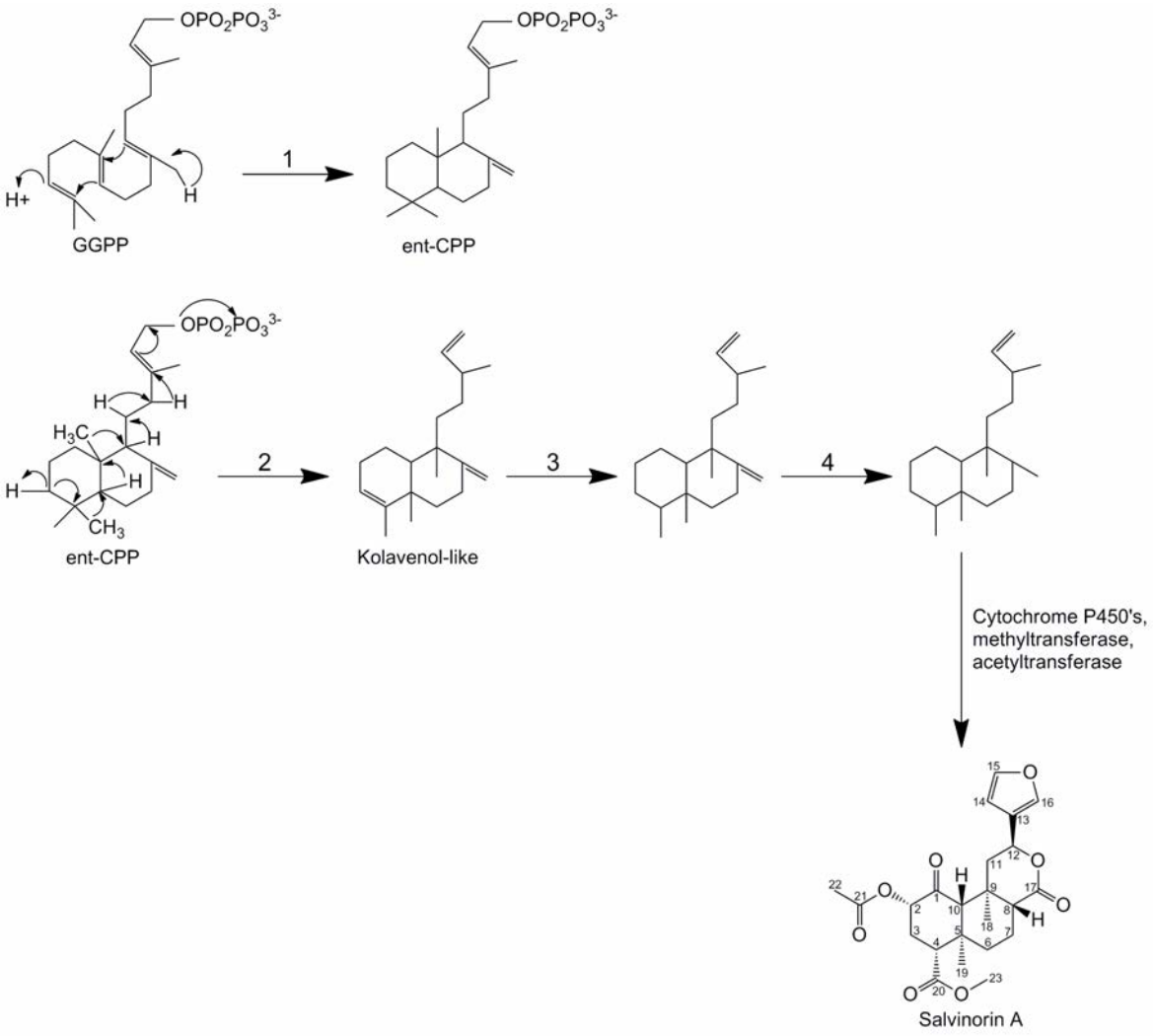


A third possibility is that *ent*-CPPS is the type II diterpene synthase that acts in the salvinorin A biosynthetic pathway. *Ent*-CPPS is a type II diterpene synthase that functions to cyclize GGPP into *ent*-CPP . *Ent*-KS, a type I diterpene synthase, then rearranges *ent*-CPP into *ent*-kaurene, and an array of enzymes convert *ent*-kaurene into the gibberellin phytohormones. In rice, *ent*-CPP is used as both a phytohormone and phytoalexin precursor [84]. The rice genome encodes two *ent*-CPPSs; one of these synthases is involved in phytoalexin biosynthesis while the other is involved in gibberellin biosynthesis. The two *ent*-CPPSs catalyze the

same cyclization of GGPP into *ent*-CPP, but are regulated differently. It is possible then, that the *S. divinorum* genome also encodes separate *ent*-CPPSs; one that functions in the gibberellin pathway, and one that functions in the salvinorin A pathway. Figure 10 details the reactions that these enzymes would hypothetically catalyze. The *ent*-CPPS enzyme would again drive the cyclization of GGPP by protonating the C3 double-bond. The resulting carbocation would then be quenched by deprotonation, creating a double-bond at the C8 position. A kolavenol synthase-like enzyme would then catalyze further rearrangements of the compound, and the carbocation the enzyme generates would be quenched by deprotonation of C3. Double-bond reductases would then act to remove the C8 double-bond, and the C3 double-bond.

Figure 10. Hypothetical *ent*-CPPS and kolavenol synthase-like reactions in the salvinorin A biosynthetic pathway

1 = Reaction catalyzed by the *ent*-CPPS enzyme. 2 = Reaction catalyzed by the kolavenol synthase-like enzyme. 3 and 4 = Reactions catalyzed by double bond reductases.



1.16 Research rationale, goals, and objectives

Salvinorin A has several medical uses, and derivatives of the compound are potentially useful in the treatment of mood and hallucination associated disorders. Given that the chemical synthesis of salvinorin A is an energy inefficient and complex process, reconstruction of whole or part of the salvinorin A biosynthetic pathway in fast growing organisms may be the only effective means to mass produce the compound. In addition, the biochemistry of neoclerodane diterpenoid biosynthesis remains unexplored in plants, and study of the first step in salvinorin A biosynthesis will add new pieces of information to the field of isoprenoid metabolism.

The primary goal of this project is to elucidate the first biochemical step of salvinorin A biosynthesis in *S. divinorum* using genomic, biochemical, chemical, and metabolic engineering tools. Specific objectives are as follows:

1. To generate genomic resources by next generation sequencing (NGS) and identify candidate type I and type II diterpene synthase transcripts from the database.
2. To characterize the expression pattern and evolutionary lineage of the identified diterpene synthases.
3. To functionally produce and purify type I and type II diterpene synthases from *Escherichia coli*.
4. To elucidate the chemical structure of the diterpenoid synthesized from the

purified recombinant enzyme or from metabolically engineered microbes.

CHAPTER TWO: MATERIALS AND METHODS

2.1 *S. divinorum* growth conditions

Our first *S. divinorum* plant was sent to us by Dr. John Page (National Research Council-Plant Biotechnology Institute, Saskatoon, Canada). An additional four plants were purchased from the Vancouver Seed Bank. Cuttings from these were used to propagate additional *S. divinorum* plants. These cuttings were always taken from large, mature plants. A branch with two or three nodes was selected, and sterile scissors were used to cut the branch an inch below the node closest to the plant's primary stem. The cutting was then immediately potted in soil.

All *S. divinorum* plants were initially cultivated in a plastic chamber in the University of Calgary Science B greenhouse. The temperature of the chamber was approximately 20°C, and the inside of the chamber was sprayed with water once a day to increase the humidity. The chamber was kept in a shaded area of the greenhouse in order to limit direct sunlight. *S. divinorum* grows under large trees in nature, and as a result direct sunlight can damage the plants. Large, mature plants were removed from the chamber and placed directly on greenhouse benches in an effort to acclimatize them to low humidity conditions. A protective layer of cheesecloth was used to further limit exposure of these plants to direct sunlight.

2.2 High performance liquid chromatography (HPLC) analyses

HPLC-diode array detector (DAD) analyses were conducted on a Nova-Pak C18 column (particle size 4 μm, length 150 mm, inner diameter 3.9 mm) at 30°C, and a Waters separation module. A 10 μl volume of each sample was injected into the column. A gradient profile starting at 45% methanol in 0.1% phosphoric acid and increasing linearly to 85% methanol in 0.1% phosphoric acid over 18 min was used as the mobile phase.

2.3 Gas Chromatography-Mass Spectrometry (GC-MS) analyses

GC-MS analyses were conducted on the following columns: an Agilent DB-5ms (5%-Phenyl-methylpolysiloxane, 30 m length, 250 μm inner diameter, 0.25 μm film) and an Agilent DB-1ms (100% dimethylpolysiloxane, 30 m length, 250 μm inner diameter, 0.25 μm film). 1 ml/min helium on an Agilent 6890N series autosampler and an Agilent 250 5975B inert XL EI/CI MS detector at 70 eV was used to perform the analyses. The temperatures of the MS transfer line and source were 230°C and 250°C, respectively. The following oven programs were used: start at 40°C, then 10°C/min to 300°C, and hold at 300°C for 1 min (27 min program); and start at 50°C, then 20°C/min to 300°C, and hold at 300°C for 3 min (15.5 min program). 1 μl of each sample was injected into the column using the splitless mode.

2.4 Liquid Chromatography-Mass Spectrometry (LC-MS) analyses

LC-MS analyses were conducted on an Agilent 1200 series HPLC and

Agilent 6410 Triple Quad LC-MS. Chromatography was completed on a Zorbax Eclipse Plus C18 column (2.1X50 mm size, 1.8 μM particle, 600 bar) with a gradient starting at 50% acetonitrile in 0.1% acetic acid and increasing to 100% acetonitrile in 0.1% acetic acid over 10 min at .4ml/min and 40°C. A 5 μl injection volume was used. MS conditions were as follows: fragmentor set to 80 V, 100 amu to 500 amu scan range for (+)-ion mode and 100 amu to 600 amu scan range for (-)-ion mode.

2.5 *S. divinorum* organic solvent extractions

1 min chloforom dip leaf extractions were completed by harvesting sets of leaves from mature *S. divinorum* plants, and dipping each set of leaves in 50 ml chloroform for 1 min. Harvested leaves ranged in length from 8 to 10 cm. These dipped leaves were flash frozen using LN_2 and ground into a flour-like powder. 1 hr leaf extractions were prepared by extracting 500 mg of each powdered leaf set in 50 ml chloroform for 1 hr. 100 μM coumarin was added to the chloroform used in the 1 min and 1 hr extractions as an internal standard. All extractions were completely dried using a rotary evaporator, re-suspended in 1 ml hexanes, and HPLC analyzed.

Tissue specific extractions were completed by harvesting 3 g sets of leaf and stem tissue from mature *S. divinorum* plants. Harvested leaves ranged in length from 5 to 15 cm. Abaxial leaf surface extractions were prepared by rubbing chloroform soaked cotton tips on the entire abaxial surface of each leaf for 5 sec.

Each set of cotton tips was then shaken in 50 ml chloroform for 30 sec. Stem, leaf and leaf (post-abaxial surface rubbing) extractions were completed by drying each respective set of plant tissue overnight at 45°C, and extracting each dried set in 50 ml chloroform for 2 hr at 20°C. 10 μM abietadiene was added to the chloroform used in all extractions as an internal standard. All extractions were filtered through a cheesecloth and dried completely over 2 days in a fume hood. The dried extractions were then re-suspended in 1 ml hexanes, and GC-MS analyzed.

A Pasteur pipette loaded with 200 mg of silica gel (Sigma) held in place with glass wool was used to purify salvinorin A present in *S. divinorum* 1 min chloroform dip leaf extractions. The column was washed with 10 ml hexanes, and combined leaf extractions were then applied to it. Ten 1 ml fractions of 50% ethyl acetate/50% hexanes were collected, and the salvinorin A fraction was LC-MS analyzed.

2.6 RNA isolation and cDNA library construction

Six sets of 5 g young leaf tissue were harvested from mature *S. divinorum* plants. Leaves were flash frozen by LN_2 and powdered, and total RNA was isolated by E.Z.N.A plant RNA maxi kit (Omega Bio-tek) according to the manufacturers instructions with some modifications. The E.Z.N.A. plant RNA maxi kit recommends using 10 g of plant tissue in each extraction, but 10 g of *S. divinorum* tissue was found to clog the kits homogenizer columns. For this reason, only 5 g of *S. divinorum* young leaf tissue was used in each extraction. Nuclease

free water was also heated to 70°C before being used to elute total RNA from the kits RNA Hi-Bind columns, as *S. divinorum* total RNA was not significantly eluted by room temperature water. RNA quantity and integrity was evaluated by Nanodrop spectrophotometer and RNA agarose gel size separation, respectively. mRNA in these total RNA samples was isolated by Dynabeads mRNA DIRECT kit (Invitrogen) according to the manufacturers instructions. The Superscript III reverse transcriptase kit (Invitrogen) was then used to generate double stranded cDNA. First strand cDNA synthesis was primed by an anchored oligo dT primer (5'-T(25)VNN-3'). 2 µg *S. divinorum* young leaf mRNA was used in each of eight cDNA synthesis reactions.

Double stranded cDNA was concentrated by phenol chloroform isoamyl alcohol precipitation, and the Trimmer Direct kit (Evrogen) was used to normalize four aliquots of 250 ng *S. divinorum* young leaf cDNA. The manufacturers protocol was followed with some modifications. 0.5 µl duplex specific nuclease (DSN) was added to each cDNA aliquot, as the addition of 1 µl DSN was found to result in excessive cDNA degradation, and the addition of 0.25 µl DSN was observed to result in incomplete normalization. Normalized cDNA was also only polymerase chain reaction (PCR) amplified for 11 cycles. The yield of PCR products was found to stop increasing at the 12^{th} cycle of PCR amplification, indicating that the reaction had plateaued. 5 µg of the standard and normalized *S. divinorum* young leaf cDNA preparations were 454 pyrosequenced (half plate each) to generate individual data sets of transcript sequencings. GS FLX titanium series reagents were used on a

Genome Sequencer FLX instrument to generate sequence data, and the FIESTA 2.0 system was used for transcript assembly (National Research Council-Plant Biotechnology Institute, Saskatoon, Canada).

2.7 Rapid amplification of cDNA ends (RACE) reactions

Oligonucleotides used in these RACE reactions are given in Table 1. The SMART RACE cDNA Amplification kit (Clontech) was used to perform 5' and 3' RACE reactions according to the manufacturers protocol. 5' RACE reactions were performed for *CPPSL1, CPPSL2, KSL1, KSL2,* and *KSL3*, and 3 'RACE reactions were performed for *CPPSL1, CPPSL2*, and *KSL3*. 5' or 3' RACE reactions were completed by primers (*1a* to *5b* in Table 1) and 5' or 3' RACE-ready *S. divinorum* young leaf cDNA template. Purified PCR products were re-amplified by primers (*6a* to *10b* in Table 1). Re-amplified PCR products were subcloned into the pGEM-Teasy vector (Promega), and the resultant plasmids were transformed into *E. coli* TOP10 competent cells. Subcloned RACE PCR products were sequenced by Sanger method.

Table 1. Oligonucelotides used in RACE amplifications

Unigene	No.	Orientation	Amplification	Primer sequence
CPPSL1	*1a*	Forward	Primary	5'-GACTCGATCATCTGCCCGCCGTAGC-3'

Unigene	No.	Orientation	Amplification	Primer sequence
CPPSL1	*1b*	Reverse	Primary	5'-CATGGGAGTGAGGCTTCTCAGAATG-3'
CPPSL2	*2a*	Forward	Primary	5'-GGCGTAGAATGGCACATCCAATGC-3'
	2b	Reverse	Primary	5'-GTGACGACTCCTCAATGGGATTCAGGCTG-3'
KSL1	*3*	Forward	Primary	5'-GGGGTTCGTAAACCAGGGATTTCACGTC-3'
KSL2	*4*	Forward	Primary	5'-GAGGCTCGTATATCAAGGATTTCACATCCG-3'
KSL3	*5a*	Forward	Primary	5'-CACGGATGTCATTCAGAAGGCGTCCGCAC-3'

Unigene	No.	Orientation	Amplification	Primer sequence
KSL3	*5b*	Reverse	Primary	5'-GCACTGAATGCTCTTCCAACAATGTCCAG-3'
CPPSL1	*6a*	Forward	Secondary	5'-CTGCCCGCCGTAGCAAGAGAATTTACCATC-3'
	6b	Reverse	Secondary	5'-CTACGGCGGGCAGATGATCGAGTCGC-3'
CPPSL2	*7a*	Forward	Secondary	5'-GCCACCTCGCCTGGAAGGTCTTTGG-3'
	7b	Reverse	Secondary	5'-CAGGCTGTTAAGGTTGTACGGTTACGATG-3'
KSL1	*8*	Forward	Secondary	5'-CGCCGCTTGCATACAAATAACACCCA-3'

Unigene	No.	Orientation	Amplification	Primer sequence
KSL2	*9*	Forward	Secondary	5'-CAGTTGCTGAGGAGAAGTGAATTCGTCGC-3'
KSL3	*10a*	Forward	Secondary	5'-CCCCACAAGATAAAGAGCTGGTAGGAC-3'
KSL3	*10b*	Reverse	Secondary	5'-GCACACAATCTGTGAGATTGGAGCAATGC-3'

2.8 Methyl jasmonate treatment

S. divinorum plants of approximately the same age and size were selected, and three young leaves ranging in length from 5 to 10 cm were harvested from each plant. A hole punch was used to remove three 6 mm diameter discs from each harvested leaf, and peltate glandular trichomes on the abaxial surface of each leaf disc were counted by dissecting microscope. 95% methyl jasmonate (Sigma-Aldrich) was used to prepare a 1 mM methyl jasmonate solution, and 0.1% v/v Tween 20 was added to the solution as a surfactant. Plastic spray bottles were used to spray each plant with 50 ml of the 1 mM methyl jasmonate solution, or with 50 ml of a control solution composed of sterile water and 0.1% v/v Tween 20.

Control plants and experimental plants were sprayed in opposite corners of the laboratory, and then covered with plastic autoclave bags to prevent dispersal of the methyl jasmonate. Plastic bags were removed after 90 min, and the methyl jasmonate was allowed to diffuse away from the experimental plants for 2 hr. The experimental and control plants were then grown in opposite corners of the greenhouse, and leaves were harvested from each plant 3, 7, 14, 21, and 28 days after the spraying. Three young leaves of approximately the same length as the leaves collected prior to spraying were harvested from each plant at each time point.

2.9 Tissue specific real-time quantitative PCR (qPCR)

Crystal clear adhesive tape (3M) was used to isolate *S. divinorum g*landular trichomes. Tape was rubbed onto the entire abaxial surface (except for the central vein) of leaves harvested from mature plants. The tape was then pulled off, with care being taken to avoid ripping leaf tissue. Each piece of tape was cut into small pieces, and the RNAqueous-Micro kit (Ambion) was used according to the manufacturers specifications to isolate RNA from glandular trichomes stuck to the tape. *S. divinorum* stem, leaf, and tape-rubbed leaf RNA was isolated by E.Z.N.A. Plant RNA kit (Omega Bio-Tek). The kit was used according to the manufacturers specifications. A chloroform:isoamyl acid alcohol (25:1) extraction was used to concentrate all RNA samples, and the Superscript III reverse transcriptase kit (Invitrogen) was used to generate cDNA. cDNA was quantified by NanoDrop spectrophotometer.

Oligonucleotides used in these qPCR reactions are given in Table 2. The SsoFast EvaGreen Supermix kit (Bio-Rad) was used according to the manufacturers protocol to complete the qPCR reactions. 11.25 ng of each *S. divinorum* cDNA preparation, and a 300 nM final concentration of each forward and reverse orientation primer was used in each reaction. Actin reference gene primers were designed by *S. miltiorrhiza* nucelotide sequence (HM231319.1). qPCR reactions for *CPPSL1, CPPSL2, KSL1, KSL2,* and *KSL3* were performed by primers (*1a* to *5b* in Table 2), and actin reference reactions were performed by primers (*6a* to *6b* in Table 2). The reactions were incubated in a real-time qPCR thermal cycler using a PCR program employing a 30 sec hot start at 95°C, followed by 45 cycles of 5 sec at 95°C and 5 sec at 57°C. A melt curve of 65°C to 90°C (0.5°C every 5 sec) was then applied to the reactions.

Table 2. Oligonucleotides used in qPCR amplifications

Sd = Salvia divinorum. Sm = Salvia miltiorrhiza. S. miltiorrhiza actin nucelotide sequence (HM231319.1) used to design primers (*6a* to *6b)*.

Unigene	No.	Orientation	Primer sequence
Sd CPPSL1	*1a*	Forward	5'-ACAGCAGAAACAGTGGCGAGAGAG-3'
	1b	Reverse	5'-TCGCTGTGCGACAGTGTCTCTT-3'
Sd CPPSL2	*2a*	Forward	5'-CGGATGTTATGAGGGGATTCACAG-3'
	2b	Reverse	5'-TGTTAGAAATCCACTCTAGGCTCGT-3'
Sd KSL1	*3a*	Forward	5'-AGAGTGAGAGAGGTTTTTCCGGT-3'
	3b	Reverse	5'-TTGTTCAAAAGCGGCAGCTGG-3'
Sd KSL2	*4a*	Forward	5'-TCTAAATCGTGGATGTCGACCG-3'
	4b	Reverse	5'-GGCACTTCTTCCCTCTGCAAT-3'
Sd KSL3	*5a*	Forward	5'-CGAGGAGACGAGAGGAAGGATA-3'
	5b	Reverse	5'-GAAAGGACATCTTTCTTTAGCAAAGGG-3'
Sm actin	*6a*	Forward	5'-AGGACCCACCGATCCAGACA-3'
	6b	Reverse	5'-GGTGCCCTGAGGTCCTGTT-3'

2.10 Subcloning into *E. coli* expression vectors

Pseudomature (FC predictions in Figure 20) *CPPSL1* and *CPPSL2* open reading frames (ORFs) were subcloned into pDEST17 plasmids for expression in

E. coli. Oligonucleotides used in these ORF amplifications are given in Table 3. Forward orientation primers were split into a attB site linker primer (5'-AAAAAGCAGGCTTCGAAAACCTGTATTTTCAGG-3') and gene-specific primers (*1a* to *2b* in Table 3). The ORFs of *CPPSL1* and *CPPSL2* were amplified by primers (*1a* to *2b* in Table 3) and *S. divinorum* young leaf cDNA template. Purified PCR products were re-amplified by attB site linker primer and gene-specific reverse orientation primers (*1b* and *2b* in Table 3) and subcloned into pDONR221 plasmid by gateway BP reaction (Invitrogen) according to the manufacturers specifications. Plasmids were transformed into *E. coli* TOP10 competent cells and ORFs were sequenced by Sanger method to assay proofreading errors. ORFs were then subcloned into pDEST17 plasmids by gateway LR reaction (Invitrogen) according to the manufacturers specifications to make translational fusions to the plasmids 6 X histidine tag.

Table 3. Oligonucelotides used in pDEST17 subclonings

Unigene	No.	Orientation	Primer sequence
CPPSL1	*1a*	Forward	5'-CCTGTATTTTCAGGGCGCGACAGTAGATGCTCCACAGG-3'
	1b	Reverse	5'-AGAAAGCTGGGTCTCAAACGGGTTCAAAAAGCACTTTG-3'
CPPSL2	*2a*	Forward	5'-CCTGTATTTTCAGGGCCTGAAATCTAAGGCTGGCAGCGGATG-3'
	2b	Reverse	5'-AGAAAGCTGGGTCTTAAACAATTTTTTCAAACAATACTTTG-3'

Pseudomature (numbered predictions in Figure 20) *KSL2* and *KSL3* ORFs were subcloned into pET28b plasmids for expression in *E. coli*. Oligonucleotides used in these ORF amplifications are given in Table 4. The ORFs of *KSL2* and *KSL3* were amplified by primers (*1a* to *2h* in Table 4*)* and *S. divinorum* young leaf cDNA template*.* Purified PCR products were subcloned into pGEM-Teasy plasmids, and the resultant plasmids were transformed into *E. coli* TOP10

competent cells and sequenced by Sanger method to assay proofreading errors. pGEM-Teasy plasmids were digested with SacI and NotI and the *KSL2* and *KSL3* ORF fragments were subcloned into pET28b plasmids to make translational fusions to the plasmids 6 X histidine tag.

Table 4. Oligonucelotides used in pET28b subclonings

Unigene	No.	Orientation	Transit peptide prediction	Primer sequence
KSL2	*1a*	Forward	-13	5'-CTGATGAGCTCCCGAGTTGGGTGTTTGAGCCAAAC-3'
	1b		-8	5'-CTGATGAGCTCCAGCCAAACTTTTCCAGTTGTC-3'
	1c		-3	5'-CTGATGAGCTCCGTTGTCAAAGCCATCGCCAAAAAG-3'
	1d		0	5'-CTGATGAGCTCCGCCATCGCCAAAAAGTCATATGC-3'

Unigene	No.	Orientation	Transit peptide prediction	Primer sequence
KSL2	*1e*	Forward	+3	5'-CTGATGAGCTCCAAAAAGTCATATGCCATCGGTAAATCC-3'
	1f		+8	5'-CTGATGAGCTCCATCGGTAAATCCAGCCTTACTACAAC-3'
	1g		+13	5'-CTGATGAGCTCCCTTACTACAACAGATTTGTTGG-3'
	1h	Reverse	-	5'-ATCACGCGGCCGCTTAACTCGAAAGTGTAAGAGGCTCG-3'
KSL3	*2a*	Forward	-13	5'-CTGATGAGCTCCGAATCACATTTTCGGAGATTC-3'
	2b		-8	5'-CTGATGAGCTCCAGATTCCGTTTCTCTACTG-3'
	2c		-3	5'-CTGATGAGCTCCACTGCTTCAGCTTCTTTG-3'

Unigene	No.	Orientation	Transit peptide prediction	Primer sequence
KSL3	*2d*	Forward	0	5'-CTGATGAGCTCCGCTTCTTTGAAAGCTG-3'
	2e		+3	5'-CTGATGAGCTCCAAAGCTGGGCTTCAAAC-3'
	2f		+8	5'-CTGATGAGCTCCACTACTTCACCAAAAAT-C-3'
	2g		+13	5'-CTGATGAGCTCCATCACCTCAATGCCAGCGTG-3'
	2h	Reverse	-	5'-ATCACGCGGCCGCTCACAACTTTTGCTCCTTGAGAAG-3'

2.11 *E. coli* cultivation and protein extraction

Plasmids of interest were heat shock transformed into *E. coli* BL21-AI

competent cells (Invitrogen), and single transformant colonies were inoculated in 5 ml sterile LB culture composed of 1% w/v Bacto-tryptone, 0.5% w/v yeast extract, 1% w/v sodium chloride, and 0.1 mg/ml of the appropriate antibiotic. The inocula were cultured overnight at 37ºC. The start culture was then diluted fifty fold in a 50 ml LB culture that was also cultured overnight at 37ºC. This culture was diluted fifty fold in TB media composed of 1.2% w/v tryptone, 2.4% w/v yeast extract, 0.4% v/v glycerol, 0.017 M KH_2PO_4, 0.072 M K_2HPO_4, and 0.1 mg/ml of the appropriate antibiotic. The *E. coli* was cultured at 37ºC until an optical density at 600 nm (OD_{600}) of 0.4 was reached. Plasmid expression was then induced by addition of 0.2% arabinose, and the *E. coli* was cultured for a desired time period and at a desired temperature. *E. coli* cells were pelleted by centrifugation and frozen at -20ºC.

Frozen *E. coli* cell pellets were re-suspended in 2 ml per g lysis buffer composed of 20 mM HEPES (pH 7.5), 0.5 M NaCl, 25 mM Imidazole, 5% v/v glycerol, 0.01% w/v lyzozyme, and 1 mM phenylmethanesulfonylfluoride. Re-suspended pellets were incubated for 30 min at 4°C and then sonicated (three 30 sec bursts, cells were cooled on ice for 1 min between bursts). Cell debris and insoluble proteins were collected in a pellet by centrifugation, and soluble recombinant protein in the supernatant was affinity chromatography purified. HisTrap columns (GE Healthcare) were used to complete small scale purifications, and a Ni^{2+}-Sepharose matrix (GE Healthcare) was used to perform large scale protein purifications.

HisTrap columns (GE Healthcare) were used according to the manufacturers specifications with some modifications. Binding buffer composed of 20 mM HEPES (pH 7.5), 0.5 M NaCl, 25 mM Imidazole, and 5% v/v glycerol was used to equilibrate and wash each column, and elution buffer composed of 20 mM HEPES (pH 7.5), 0.5 M NaCl, 350 mM Imidazole, and 5% v/v glycerol was used to elute bound proteins. The Ni^{2+}-sepharose matrix (GE Healthcare) was also used according to the manufacturers specifications with some modifications. Each 1 ml aliquot of Ni^{2+}-sepharose matrix was collected in a column and equilibrated by 30 ml of the binding buffer described above. 1 ml equilibrated matrix was added to each 20 ml aliquot of sonicated *E. coli* cell supernatant, and incubated with gentle mixing for 1 hr at 4°C. Each 1 ml matrix aliquot was then collected in a column and washed with 50 ml of binding buffer. Bound proteins were eluted by the elution buffer described above. Purified recombinant proteins generated by both purification methods were desalted by PD MiniTrap G-25 columns (GE Healthcare) used according to the manufacturers specifications. Desalted purified protein was flash frozen by LN_2 and stored at -80°C. Size separation by sodium dodecyl sulfate-poly-acrylamide gel electrophoresis (SDS-PAGE) and Coomassie blue staining was used to assay expression of subcloned ORFs.

2.12 *In vitro* activity assays

In vitro enzyme assays were carried out in 500 µl of 50 mM HEPES (pH 7.2), 10 µM MgCl2, 5% v/v glycerol, and 5 mM dithiothreitol buffer. 100 µg each purified desalted recombinant protein and 25 µM GGPP (Echelon) was used in

each assay. All assays were conducted at 30°C for 1 hr, and CPPSL1 and CPPSL2 assays were then de-phosphorylated by addition of 10 U calf intestinal alkaline phosphatase (Invitrogen) and incubation at 37°C overnight. Reaction products were extracted with 500 µl pentane that was pre-purifed by alumina-silica column, and analyzed by GC-MS. The pentane was replaced with 50% methanol for LC-MS analysis. Denatured proteins were used in negative control assays. Proteins were denatured by 10 min incubation at 99°C.

2.13 CPPSL2 product NMR analysis

Six large scale CPPSL2 *in vitro* assays were carried out in 10 ml pre-extracted buffer composed of 50 mM HEPES (pH 7.2), 10 µM MgCl2, 5% v/v glycerol, and 5 mM dithiothreitol. Each assay contained 1 mg purified desalted pseudomature (FC prediction in Figure 20) CPPSL2 protein. 1.25 mg GGPP (Echelon) was added to each reaction over 60 min with gentle stirring at 30°C, and the assays were then continued for 90 min. Reaction products were de-phosphorylated by addition of 100 U calf intestinal alkaline phosphatase (Invitrogen) and incubation at 37°C overnight. Two volumes 10 ml pentane were used to extract each assay, and purity was assayed by GC-MS analysis. Combined extracts were gently evaporated to 100 µl with a N_2 gas stream, diluted with Benzene-d_6, and further evaporated to 300 µl. The solution was transferred to a 3 mm NMR tube and analyzed by 1H NMR, heteronuclear multiple-bond coherence, heteronuclear single quantum coherence, correlation spectroscopy and 13C J-modulated spin-echo using a Bruker AVANCE 600-MHz NMR spectrometer.

2.14 Subcloning into yeast expression vectors

Pseudomature (FC predictions in Figure 20) *CPPSL2, KSL1, KSL2,* and *KSL3* ORFs were subcloned into pESC-Leu plasmids for expression in *Saccharomyces cerevisiae*. Oligonucleotides used in these ORF amplifications are given in Table 5. Reverse orientation primers were split into FLAG (*CPPSL1*) or c-MYC (*KSL1, KSL2,* and *KSL3*) epitope tag linker primers (*1c, 2c, 3c,* and *4c* in Table 5) and gene-specific primers (*1b, 2b, 3b,* and *4b* in Table 5). Figure 35 shows the epitope tags that are fused to each ORF. *CPPSL2, KSL1, KSL2* and *KSL3* ORFs were amplified by forward (*1a, 2a, 3a,* and *4a* in Table 5*)* and reverse (*1b, 2b, 3b,* and *4b* in Table 5) orientation gene-specific primers and *S. divinorum* young leaf cDNA template. Purified PCR products were re-amplified by forward orientation gene-specific primers (*1a, 2a, 3a,* and *4a* in Table 5) and epitope tag linker primers (*1c, 2c, 3c, and 4c in* Table 5*).* Re-amplified PCR products were subcloned into pGEM-Teasy (Promega) plasmids, and the resultant plasmids were transformed into *E. coli* TOP10 competent cells and sequenced by Sanger method to assay proofreading errors. pGEM-Teasy (*CPPSL2* ORF) plasmids were digested with BglII and PacI, and the *CPPSL2* ORF fragment was subcloned into the pESC-Leu plasmid to make an in frame fusion to the GAL10 promoter. pGEM-Teasy (*KSL1, KSL2* and *KSL3* ORF) plasmids were then digested with BamHI or ApaI and the resultant ORF fragments were subcloned into pESC-Leu (*CPPSL2* ORF) plasmids to make in frame fusions to the GAL1 promoter.

The “GGPP helper” plasmid was prepared by subcloning *S. cerevisiae*

BTS1 (ENA U31632.1) into a pESC-His backbone encoding a truncated *S. cerevisiae* HMG-CoA reductase (strain EYP208 [85]). The *BTS1* ORF was amplified by forward (5'-GCACTAGTAATGGCTGCAGACCAATTGGTGAAA-3') and reverse (5'-CCACTAGTTTAGGATTTAATGCAGGTGACGGA-3') orientation gene-specific primers and *S. cerevisiae* cDNA template. Purified PCR products were digested with SpeI and *BTS1* ORF fragments were subcloned into the pESC-His backbone to make a translational fusion to the C-terminus c-MYC epitope tag.

Table 5. Oligonucelotides used in pESC-Leu subclonings

Unigene	No.	Orientation	Amplification	Primer sequence
CPPSL2	*1a*	Forward	Primary and secondary	5'-CTCAAGATCTAACATGGAGGGTATTCATAGAAGTCAATTAGCAG-3'
	1b	Reverse	Primary	5'-TCGTCGTCATCCTTGTAATCAACAATTTTTTCAAACAATAC-3'
	1c		Secondary	5'-CATCTTAATTAATCAGATCTTATCGTCGTCATCCTTGTAATC-3'

Unigene	No.	Orientation	Amplification	Primer sequence
KSL1	*2a*	Forward	Primary and secondary	5'-CTCAGGATCCAACATGGATTTGATGGGTAAGATAGCAGAGAAG-3'
	2b	Reverse	Primary	5'-CTTCGGAAATCAACTTCTGTTCTTTGCCACTCACATTGTTAGC-3'
	2c		Secondary	5'-CATCGGGCCCTTACTCGAGGTCTTCTTCGGAAATCAACTTCTG-3'
KSL2	*3a*	Forward	Primary and secondary	5'-CATCGGGCCCAACATGGATTTGTTGGGTGAATTAAAGGAGAAG-3'
	3b	Reverse	Primary	5'-CTTCTTCGGAAATCAACTTCTGTTCACTCGAAAGTGTAAGAGGCTC-3'
	3c		Secondary	5'-CATCGGGCCCTTACTCGAGGTCTTCTTCGGAAATCAACTTCTG-3'

Unigene	No.	Orientation	Amplification	Primer sequence
KSL3	*4a*	Forward	Primary and secondary	5'-CATCGGATCCAACATGTGTTTCGAGGAGACTAGAGGAAGG-3'
	4b	Reverse	Primary	5'-CTTCTTCGGAAATCAACTTCTGTTCCAACTTTTGCTCCTTGAGAAG-3'
	4c		Secondary	5'-CATCGGGCCCTTACTCGAGGTCTTCTTCGGAAATCAACTTC-3'

2.15 *S. cerevisiae* cultivation and protein extraction

Plasmids of interest were lithium acetate transformed into *S. cerevisiae* cells, and single transformant colonies were inoculated in 5 ml yeast synthetic sucrose minimal media composed of 0.67% w/v yeast nitrogen base, 2% w/v sucrose, and 0.13% w/v amino acid dropout powder omitting the appropriate amino acids. The inocula were cultured overnight at 30ºC, and the start culture was then diluted fifty fold in yeast synthetic galactose minimal media culture composed of 0.67% w/v yeast nitrogen base, 1.8% w/v galactose, 0.2% w/v sucrose, and 0.13% w/v amino acid dropout powder omitting the appropriate amino acids. The yeast was cultured for 72 hr at 30ºC, and used for metabolite or protein expression

analysis.

Yeast culture metabolite analysis was completed by ethyl acetate extraction. The ethyl acetate was evaporated by rotary evaporator, and metabolites were re-suspended in hexanes and analyzed by GC-MS. Yeast culture protein isolates were prepared by re-suspending pelleted yeast cells in a lysis buffer composed of 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM dithiothreitol, 5% v/v glycerol, 1 mM ethylenediaminetetraacetic acid, and 1 mM phenylmethanesulfonylfluoride. 0.5 mm diameter glass beads and a bead beater apparatus was then used to rupture yeast cells. Cellular debris was collected in a pellet by centrifugation, and the supernatant was stored at -80°C. Expression of subcloned ORFs was assessed by SDS-PAGE size separation and anti-cMYC or anti-FLAG antibody (Sigma) immunoblot. Secondary antibodies used in the immunoblots were linked to horseradish peroxidase, and the ECL plus acridan kit (GE Healthcare) was used according to the manufacturers specifications to visualize bound secondary antibodies.

CHAPTER THREE: RESULTS

3.1 Evidence of salvinorin A production by *S. divinorum*

Phytochemicals of interest have to be analyzed before biochemical and genomic studies can begin, as the synthesis of secondary metabolites is significantly influenced by plant genotypes and cultivation conditions. Therefore, the presence of salvinorin A in *S. divinorum* plants grown in the University of Calgary greenhouse was analyzed by HPLC-DAD and LC-MS prior to NGS. HPLC-DAD analysis of the salvinorin A standard on a C18 reverse phase column revealed a single peak of λ_{max} at 207.6 nm (Figure 11B and D). Simple chloroform extraction of *S. divinorum* leaves identified one metabolite that had the same elution time (8.39 min) and UV spectra (λ_{max} at 207.6 nm) as the salvinorin A standard (Figure 11A and C). This *S. divinorum* compound was collected and further analyzed by LC-MS to ensure its mass. In the (-)-ion mode, a dominant $[M-H]^-$ ion with *m/z* 431.2 was detected alongside an adduct of acetic acid $[M+CH_3COOH-H]^-$ at *m/z* 491.2 (Figure 12A). In the (+)-ion mode, a dominant $[(M+H)^+$ - CH3COOH] ion with *m/z* 373.3 was detected alongside adducts of sodium (*m/z* 455.3), potassium (*m/z* 471.3), and hydrogen proton (*m/z* 433.2) (Figure 12B). The molecular mass of the neutral compound was therefore inferred to be 432.2, consistent with the molecular mass of salvinorin A. These HPLC-DAD and LC-MS analysis results confirmed that salvinorin A is synthesized in *S. divinorum.*

Figure 11. **HPLC-DAD analysis of *S. divinorum* leaf extraction**

Chromatogram of *S. divinorum* leaf extraction (A) and Salvinorin A standard purchased from Chromadex (B). UV spectra of the peak at 8.39 min from the leaf extraction and salvinorin A standard chromatograms are shown in C and D, respectively. Chromatograms were extracted at 208 nm to generate UV spectra. Leaves were dipped in 50 ml chloroform for 1 min, and extractions were then dried and resuspended in 1 ml chloroform. Analysis was performed on a Nova-Pak C18 reverse phase column.

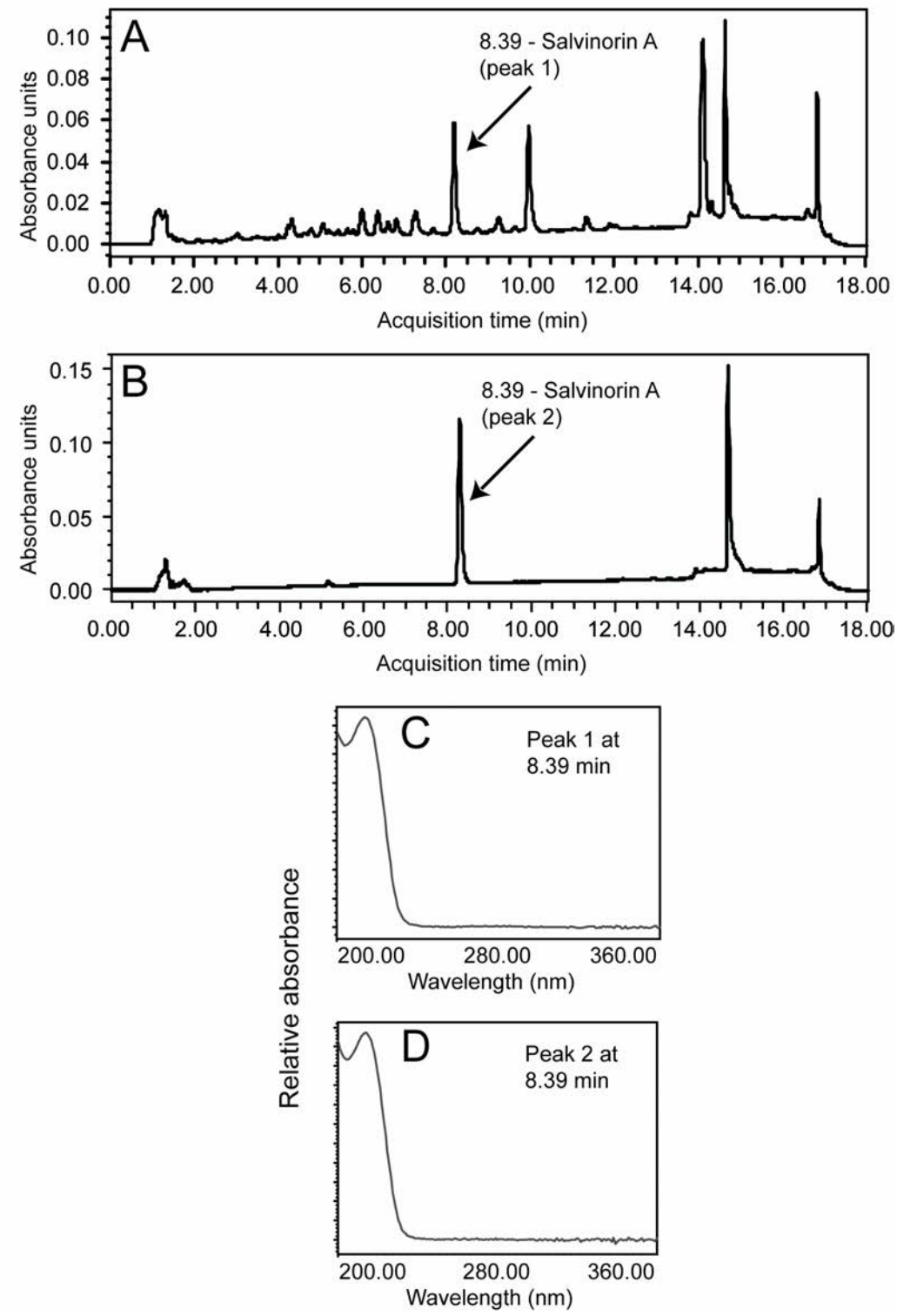
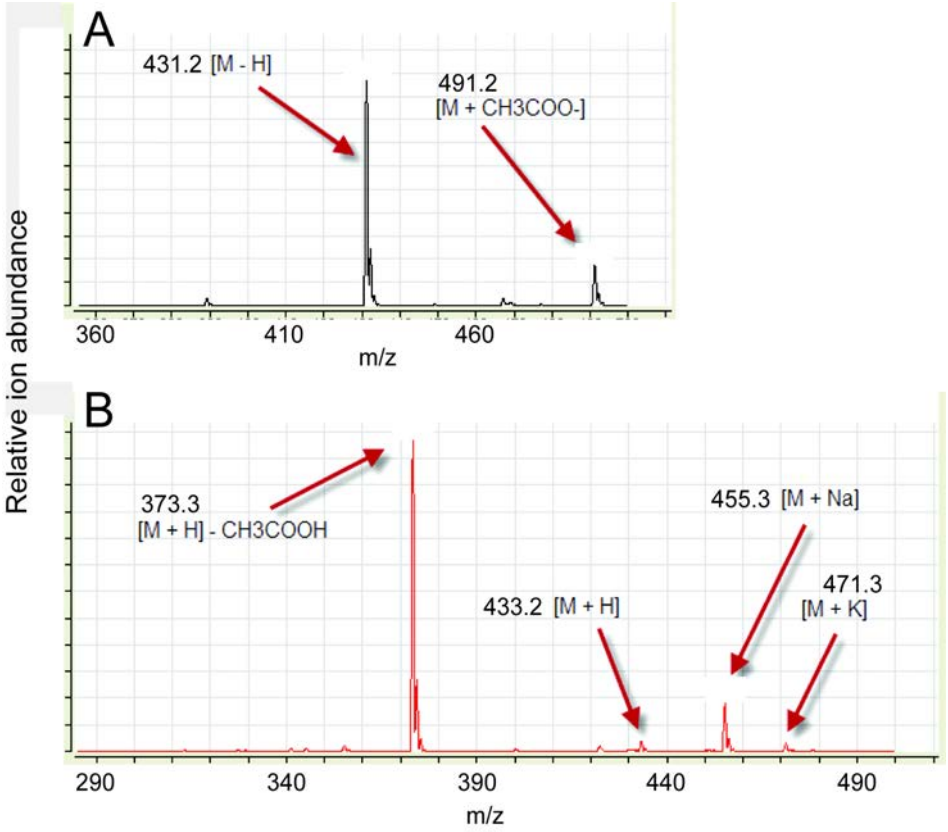


Figure 11. HPLC-DAD analysis of *S. divinorum* leaf extraction

Figure 12. LC-MS analysis of *S. divinorum* leaf extraction

Mass spectra in (-)-ion mode (A) and (+)-ion mode (B). Leaf sets were dipped in 50 ml chloroform for 1 min, and extractions were then dried and resuspended in 1 ml chloroform. Extractions were pooled, and the compound with HPLC-DAD retention time 8.39 min was purified by silica gel column (50% ethyl acetate/50% hexanes mobile phase). Analysis of the purified metabolite was performed on a Zorbax Eclipse Plus C18 column.



After confirming the presence of salvinorin A in *S. divinorum*, leaf salvinorin A content in the five initial *S. divinorum* plants was quantified. Of these five plants, plant A was provided by Dr. Jon Page at the NRC-Plant Biotechnology Institute,

while plants B, C, D, and E were purchased from the Vancouver Seed Bank. Average leaf salvinorin A content from the five plants was determined to be 903±354 μg per g dry weight. This concentration of salvinorin A falls in the lower end of the range previously reported (890-3,700 μg salvinorin A per g leaf dry weight) [57]. Therefore, we confirmed that salvinorin A is reliably synthesized in *S. divinorum* leaves.

3.2 Evidence of salvinorin A accumulation in peltate glandular trichomes

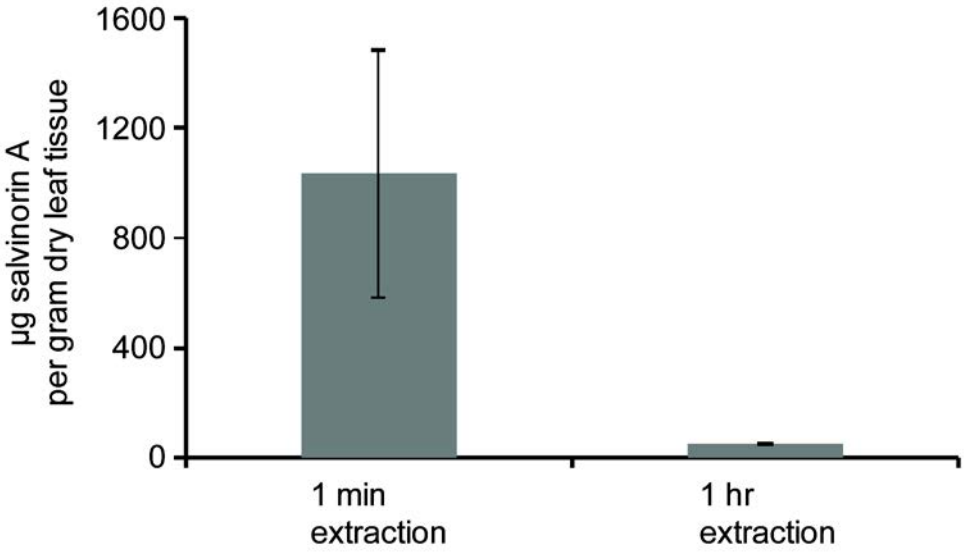
As with other plant isoprenoids, salvinorin A was expected to accumulate in the glandular trichome secretory cells that are displayed on *S. divinorum* leaves [53]. Therefore, organ-specific accumulation of salvinorin A was examined by differential organic extraction. In this experiment, *S. divinorum* leaves were exposed to chloroform for 1 min, and the chloroform-dipped leaves were then powdered and extracted with chloroform for an additional 1 hr. Subsequent HPLC-DAD analysis showed that the initial 1 min chloroform dip extracted more than 95% of *S. divinorum* leaf salvinorin A content (Figure 13). The amount of salvinorin A in the 1 min dip was estimated to be 1035±451 μg per g leaf dry weight, whereas only 50.7±0.8 μg/g salvinorin A was detected in the leaf tissue that was extracted for 1 hr after the chloroform dip.

Leaf browning was used to qualitatively gauge chloroform penetration, as browning indicates extraction of chlorophylls by rupture of leaf epidermal cells. Very little penetration of chloroform was observed in the 1 min dip. As a result,

contamination from leaf cell metabolites (e.g., chlorophylls) was markedly reduced in the 1 min dip extraction. The low standard deviation (0.77 µg/g salvinorin A) of the 1 hr chloroform extraction demonstrated that the 1 min chloroform dip reliably extracted a majority of salvinorin A from the leaves. These results suggested that salvinorin A accumulation occurs on the *S. divinorum* leaf surface, possibly in the subcuticular sacs of the peltate glandular trichomes, or on capitate glandular trichomes.

Figure 13. Differential extraction of *S. divinorum* leaf tissue

1 min extractions were prepared by dipping leaves in 50 ml chloroform for 1 min. 1 hr extractions were prepared by freezing and powdering the dipped leaves, and exposing 500 mg of leaves to 50 ml chloroform for 1 hr. 100 µM coumarin was added to all extractions as an internal standard, and the organic solvents were then dried and resuspended in 1 ml chloroform. Analysis was performed on a Nova-Pak C18 reverse phase column. A salvinorin A standard curve was used to quantify salvinorin A content. Values are averages ± SD (n=4).



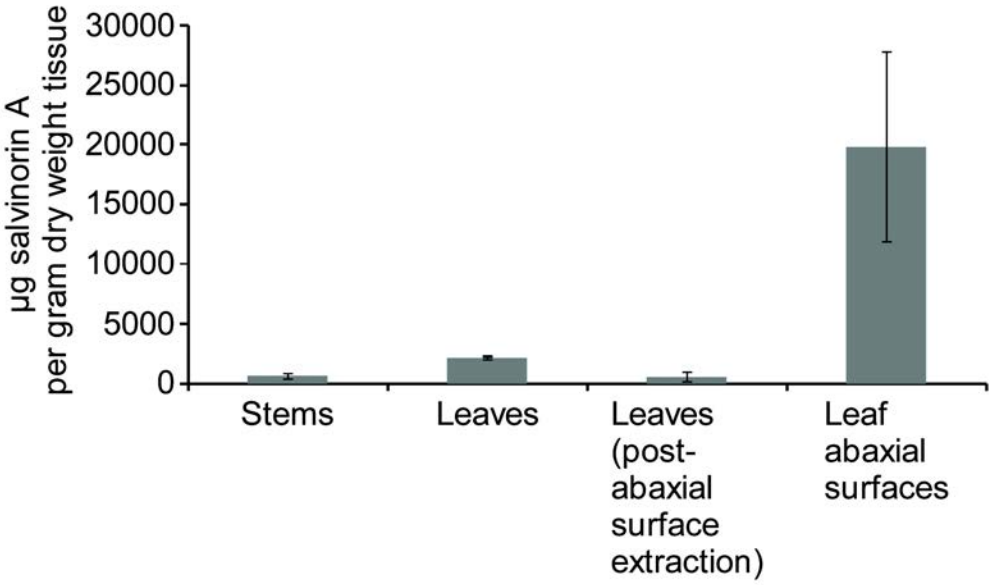
Peltate glandular trichomes are present only on the abaxial surface of *S. divinorum* leaves, whereas capitate glandular trichomes are found on both *S. divinorum* abaxial and adaxial leaf surfaces [48]. In order to obtain more cellular specificity of salvinorin A accumulation, chloroform soaked cotton tips were applied solely to the abaxial surface of the leaves. The cotton tip rubbed leaves were then

further extracted by chloroform dip to measure residual salvinorin A. GC-MS analysis of these extractions revealed that salvinorin A content in the cotton tip extractions was more than ten fold higher than that observed in the leaves after cotton-tip rubbing (Figure 14). This result demonstrated that peltate glandular trichomes are likely the sole cellular sites for salvinorin A biosynthesis and storage in *S. divinorum.* Capitate glandular trichomes are present on both sides of *S. divinorum* leaves, and only the abaxial surface of each leaf was rubbed with chloroform soaked cotton tips. A significantly higher salvinorin A content would therefore be observed in the residual leaves that were extracted after cotton tip rubbing if capitate glandular trichomes were also able to synthesize salvinorin A.

Salvinorin A content was determined to be highest (19,837±7,968 µg/g) in the organic extraction of the abaxial leaf surface, and lowest (519±410 µg/g) in the leaves that were cotton tip rubbed prior to extraction (Figure 14). Relatively low concentrations of salvinorin A were also detected in *S. divinorum* stems (609±241 µg/g) and leaves (2,114±135 µg/g) (Figure 14). The presence of salvinorin A in stem tissue was expected, as peltate glandular trichomes are present on *S. divinorum* leaves and stems at roughly equal densities [48].

Figure 14. Salvinorin A content of *S. divinorum* tissues

Abaxial leaf surface extractions were prepared by rubbing the abaxial surface of *S. divinorum* leaves with chloroform soaked cotton tips for 5 sec. Each set of cotton tips was shaken in 50 ml chloroform for 30 sec. Cotton tip rubbed leaves, leaf tissue, and stem tissue was then dried overnight at 45°C, and sets of dried tissue were extracted in 50 ml chloroform for 2 hr at 20°C. 10 μM abietadiene was added to all extractions as an internal standard, and extractions were filtered, dried, and resuspended in 1 ml hexanes. Analysis was performed on an Agilent DB-5ms column using a 15.5 min oven program. Values are averages ± SD (n = 4).



In conclusion, HPLC-DAD, LC-MS, and differential chemical extractions involving dipping and cotton tip rubbing clearly demonstrated that salvinorin A is specifically synthesized in the peltate glandular trichomes present on *S. divinorum* leaves and stems. The amount of salvinorin A synthesized was comparable to the

reported values, and it was therefore concluded that the *S. divinorum* used in this study was suitable for further genomics investigation by NGS.

3.3 Preparation of *S. divinorum* cDNA

As salvinorin A accumulation in peltate glandular trichomes was evident, the isolation of trichomes through use of bead beater and sticky tape methods was attempted. These techniques were however not successful on a level that would provide the 5 µg cDNA necessary for NGS by 454 technology. As an alternative, peltate glandular trichome-rich tissues were selected for total RNA isolation and cDNA synthesis efforts. Analysis by using a dissecting microscope revealed that *S. divinorum* young leaves have a significantly higher peltate glandular trichome density than old leaves and stems (data not shown), and young leaves were therefore used to prepare total RNA samples. These total RNA samples were size separated on an RNA agarose gel to verify RNA integrity (Figure 15). cDNAs were synthesized using mRNA purified from *S. divinorum* young leaf total RNA, and a fraction of this cDNA was normalized to remove abundant transcripts. Qualitative differences between the standard (non-normalized) and normalized cDNAs were visualized by a size-separation of the samples on agarose gels (Figure 16). Several discrete bands that were observed in the standard cDNAs were not present in the normalized cDNAs, indicating that abundant cDNA populations were preferentially degraded by the normalization procedure. The standard and normalized cDNA samples were subjected to 454 pyrosequencing (half plate each) to generate individual data sets of transcript sequencings.

Figure 15. Integrity of *S. divinorum* young leaf total RNA

The E.Z.N.A kit (Omega Bio-tek) was used to extract *S. divinorum* young leaf RNA. 5 g tissue was used in each extraction.

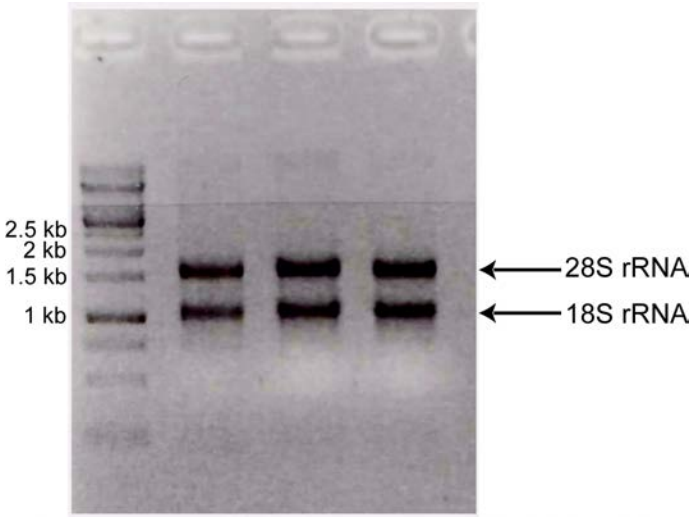
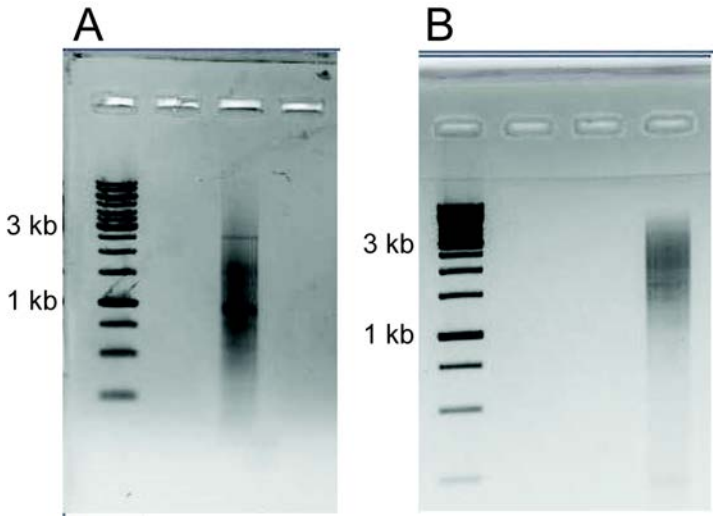


Figure 16. Qualitative comparison of *S. divinorum* young leaf standard and normalized cDNA

Standard cDNA (A) was prepared using purified *S. divinorum* young leaf mRNA as a template. A fraction of the standard cDNA preparation was normalized (B) using the Trimmer Direct kit (Evrogen).



3.4 Characterization of *S. divinorum* cDNA libraries

cDNA library transcript assembly was performed by FIESTA 2 system (National Research Council-Plant Biotechnology Institute, Saskatoon, Canada). The body of work on *S. divinorum* transcripts is relatively small, and there are no entries in the National Centre for Biotechnology Information (NCBI) nucleotide database thus far. These libraries represent the first *S. divinorum* transcript set. The coverage and quality of the standard and normalized *S. divinorum* cDNA libraries was therefore analyzed in several different ways. An initial comparison of the two data sets is described in Table 6. The libraries have a similar number of

reads, but the normalized library was revealed to have significantly more unigenes than the standard library. The percentage of sequences with NCBI Basic Local Alignment Search Tool (BLAST) hits was also significantly higher in the normalized library, possibly indicating higher coverage of *S. divinorum* transcripts. Several characterized transcripts for diterpenoid metabolism were identified in each library by NCBI BLAST hit.

Table 6. Comparison of *S. divinorum* cDNA libraries

Standard cDNA was prepared using purified *S. divinorum* young leaf mRNA as a template. A fraction of this standard cDNA preparation was normalized using the Trimmer Direct kit (Evrogen). cDNAs were 454 pyrosequenced (half plate each) to generate individual data sets, and assembled by FIESTA 2 system.

Characteristic	Standard library	Normalized library
Reads	110,862	118,602
Total unigenes	33,511	47,838
Contigs	13,636	17,905
Singletons	19,875	29,933
Sequences with BLAST hits	31,147 (92.9%)	38,021 (79.5%)
Average sequence length	461	382
Pathways identified	343	304
Transcripts related to secondary metabolism	62	53
Transcripts related to isoprenoid metabolism	22	18
Transcripts related to diterpenoid metabolism	7	5

MVA and DXP pathway coverage was also assessed in each library. All the enzymatic steps in the MVA pathway were present in the standard library, whereas two of these steps were not observed in the normalized library (Table 7). All DXP pathway enzymes were highly represented in the standard library except for the CDP-ME kinase (Table 8). In the normalized library, two DXP pathway enzymes were absent, and the DXP reductase was represented by only one singleton. Overall, the standard library was revealed to display superior coverage of both metabolic pathways. The CDP-ME kinase step that was not present in the standard library was however highly represented in the normalized library, suggesting that the libraries complemented each other in certain ways.

Table 7. MVA pathway representation in *S. divinorum* cDNA libraries

cDNA was normalized using the Trimmer Direct kit (Evrogen). cDNAs were 454 pyrosequenced (half plate each) to generate individual data sets and assembled using FIESTA 2.0.

Unigene	Standard library (bp)	Normalized library (bp)
HMG-CoA synthase	1903	1256
HMG-CoA reductase	2675	741
MVA kinase	942	No contigs or singletons
Phosphomevalonate kinase	1410	No contigs or singletons
Mevalonate pyrophosphate decarboxylase	1487	1565
IPP isomerase	1778	1582

Table 8. DXP pathway representation in *S. divinorum* cDNA libraries

cDNA was normalized using the Trimmer Direct kit (Evrogen). cDNAs were 454 pyrosequenced (half plate each) to generate individual data sets, and assembled using FIESTA 2.0.

Unigene	Standard library (bp)	Normalized library (bp)
DXP synthase	4071	860
DXP reductase	2070	282
CDP-ME synthase	603	No contigs or singletons
CDP-ME kinase	No contigs or singletons	405
MEcPP synthase	594	No contigs or singletons
HMB-PP synthase	3787	1414

Tables 9 and 10 detail the influence of normalization on individual transcripts. The most abundant transcripts in the standard library were selected, and their abundance in the standard and normalized libraries was compared in Table 9. Highly represented transcripts in the standard library were markedly reduced in the normalized library. For example, the read number of catalase and chlorophyll a/b-binding protein was reduced from 436 to 3 and 268 to 4, respectively. The normalization procedure therefore resulted in a significant decrease in repeated sequencings of highly abundant transcripts.

Conversely, the five most highly abundant transcripts in the normalized library were found to have a significantly lower representation in the standard library (Table 10). Normalization also increased sequence length for three of these five unigenes. This improvement in representation suggested that the relative read number of each transcript was increased by the normalization procedure. These cDNA library analyses demonstrated that the standard and normalized libraries provide coverage of the *S. divinorum* secondary metabolic transcriptome, and of diterpene-specific metabolism.

Table 9. Representation of the five most highly abundant unigenes in the *S. divinorum* standard cDNA library

cDNA was normalized using the Trimmer Direct kit (Evrogen). cDNAs were 454 pyrosequenced (half plate each) to generate individual data sets, and assembled using FIESTA 2.0.

Unigene	Standard library - unigene members	Standard library - length of unigene (bp)	Normalized library - unigene members	Normalized library - length of unigene (bp)
Catalase	436	1779	3	501
Chlorophyll a-b binding protein	268	1019	4	424
Photosystem 1 reaction center subunit III	267	715	4	592
Light harvesting complex II protein	262	1005	10	547
Photosystem I subunit VI	259	666	4	579

Table 10. Representation of the five most highly abundant unigenes in the *S. divinorum* normalized cDNA library

cDNA was normalized using the Trimmer Direct kit (Evrogen). cDNAs were 454 pyrosequenced (half plate each) to generate individual data sets, and assembled using FIESTA 2.0.

Unigene	Standard library - unigene members	Standard library - length of unigene (bp)	Normalized library - unigene members	Normalized library - length of unigene (bp)
Transaldolase	3	621	129	747
Ycf2 – ATP binding	2	887	92	671
ATP synthase subunit 9	3	575	83	355
Diacyl glycerol protein	6	649	82	1579
Lipid N-terminal conserved region protein	1	358	74	1386

3.5 Identification of candidate diterpene synthases

The major goal of this research was to identify and characterize the diterpene synthases that act in the salvinorin A biosynthetic pathway. Candidate diterpene synthases in the two *S. divinorum* libraries were thus identified by BLAST homology search to previously characterized diterpene synthases in *S. sclarea*, *A. thaliana,* and several other plants. Five of these candidates were found, and the full-length sequence of each transcript was obtained by RACE reaction. Possible functions of the diterpene synthase candidates were inferred by phylogenetic analysis. Two of the candidates are highly homologous to a *Salvia miltiorrhiza* type II diterpene synthase (Figure 17), and were named as copalyl diphosphate synthase-like (*CPPSL*) *1* and *CPPSL2.* The other three candidates were named kaurene synthase-like *(KSL) 1, KSL2 and KSL3*, as analysis revealed that they are highly homologous to a *S. miltiorrhiza* type I diterpene synthase (Figure 17). All five candidates were found to contain conserved Mg^{2+}-binding motifs that are present in all diterpene synthases. These are the "DXDD" type II diterpene synthase motif in *CPPSL1* and *CPPSL2*, and the "DDXXD" type I diterpene synthase motif in *KSL1*, *KSL2* and *KSL3*. *KSL3* also has a large internal sequence element known as a "gamma domain" that is not observed in *KSL1* and *KSL2*. Gamma domain loss in terpene synthases is common, and this sequence element is not strictly necessary for enzyme activity [86].

All the candidate diterpene synthases were represented in both the standard and normalized cDNA libraries (Table 11). Normalization increased the relative

representation of *CPPSL1, CPPSL2, KSL2,* and *KSL3,* and decreased the relative representation of *KSL1*. This result further showed the effectiveness of combinatorial use of standard and normalized libraries to study secondary metabolism.

Table 11. Representation of candidate diterpene synthases in *S. divinorum* cDNA libraries

cDNA was normalized using the Trimmer Direct kit (Evrogen). cDNAs were 454 pyrosequenced (half plate each) to generate individual data sets, and assembled using FIESTA 2.0.

Unigene	Representation in standard cDNA library - Length (read number)	Representation in normalized cDNA library - Length (read number)
CPPSL1	147 bp (1)	1800 bp (40)
CPPSL2	475 bp (1)	821 bp (5)
KSL1	1774 bp (10)	894 bp (3)
KSL2	435 bp (1)	826 bp (3)
KSL3	469 bp (1)	536 bp (2)

Phylogenetic analysis of the five candidate diterpene synthases was completed, and the results are shown in Figure 17. Shared homology for the five

genes supported their hypothesized roles as diterpene synthases. The ancestry tree also demonstrated that the *CPPSLs* and *KSLs* are likely different types of diterpene synthase, as was indicated by their conserved “DXDD” and “DDXXD” Mg^{2+}-binding motifs. *KSL2's* close relation to *S. sclarea* sclareol synthase additionally suggested possible involvement of the enzyme in a secondary metabolic pathway. Sclareol synthase is a type I diterpene synthase that acts to produce sclareol, a secondary metabolite in *S. sclarea*. Surprisingly, *KSL3* was found to be distantly related to all of the other *S. divinorum* candidate diterpene synthases.

Figure 17. Diterpene synthase phylogenetic tree

Dialign 2.2.1 [87] software and further manual inspection was used to generate protein sequence alignments. The maximum likelihood algorithm was used to perform phylogenetic analyses, using PhyML 3.0 software [88] with four rate substitution categories, LG substitution model, BIONJ starting tree and 100 bootstrap repetitions. Protein sequences used to generate the tree are as follows.

Abbreviation = *Protein*, *Species of origin*. *Accession No.*

AtECPS = *Ent*-copalyl diphosphate synthase, *Arabidopsis thaliana*. NP_192187.1

AtEKS = *Ent*-kaurene synthase, *Arabidopsis thaliana.* NP_178064.1

CcCLS = Copal-8-ol synthase, *Cistus creticus.* ADJ93862.1

CmCPS1 = *Ent*-copalyl diphosphate synthase 1, *Cucurbita maxima.* AF049905

CmEKS = *Ent*-kaurene synthase, *Cucurbita maxima.* Q39548.1

LsECPS = *Ent*-copalyl diphosphate synthase, *Lactuca sativa.* AB031205

LsEKS = *Ent*-kaurene synthase, *Lactuca sativa.* AB031204

NtABS = *Cis*-abienol synthase, *Nicotiana tabacum.* CCD33019.1

NtLPPS = 8-hydroxy-copalyl diphosphate synthase, *Nicotiana tabacum.* CCD33018.1

OsCPS4 = *Syn*-copalyl diphosphate synthase, *Oryza sativa.* AY530101

OsCPS1 = *Ent*-copalyl diphosphate synthase, *Oryza sativa.* AB126932

OsEKS1 = *Ent*-kaurene synthase, *Oryza sativa.* NP_001053841.1

OsKSL11 = Stemod-13(17)-ene synthase, *Oryza sativa.* Q1AHB2.1

OsKSL4 = Syn-pimara-7,15-diene synthase, *Oryza sativa.* Q0JEZ8.1

OsKSL7 = *Ent*-cassadiene synthase, *Oryza sativa.* Q00G37.2

PgECPS = *Ent*-copalyl diphosphate synthase, *Picea glauca.* GU045755

PgEKS = *Ent*-kaurene synthase, *Picea glauca.* ADB55708

PpCPS/KS = *Ent*-copalyl diphosphate/*ent*-kaurene synthase, *Physcomitrella patens.* BAF61135.1

ShSBS = Santalene/bergamotene synthase, *Solanum habrochaites.* B8XA41.1

SlECPS = *Ent*-copalyl diphosphate synthase, *Solanum lycopersicum.* JN412074

SlEKS = *Ent*-kaurene synthase, *Solanum lycopersicum.* JN412086

SlPHS = Phellandrene synthase, *Solanum lycopersicum.* FJ797957

SmCPS = (+)-copalyl diphosphate synthase, *Salvia miltiorrhiza.* ABV57835.1

SmKSL = Miltradiene synthase, *Salvia miltiorrhiza.* ABV08817.1

SrKS1 = *Ent*-kaurene synthase 1, *Stevia rebaudiana.* AF097310.1

SrECPS = *Ent*-copalyl diphosphate synthase, *Stevia rebaudiana.* AAB87091.1

SsLPPS = 13-labden-8,15-diol diphosphate synthase, *Salvia sclarea*. AET21247.1

SsSS = Sclareol synthase, *Salvia sclarea.* AET21246.1

TaCPS1 = *Ent*-copalyl diphosphate synthase, *Triticum aestivum.* AB439588

TaCPS2 = (+)-copalyl diphosphate synthase, *Triticum aestivum.* AB439589

TaEKS = *Ent*-kaurene synthase, *Triticum aestivum.* GU980889

TaKSL5 = *Ent*-kaurene/*ent*-beyerane synthase, *Triticum aestivum.* AB597961.1

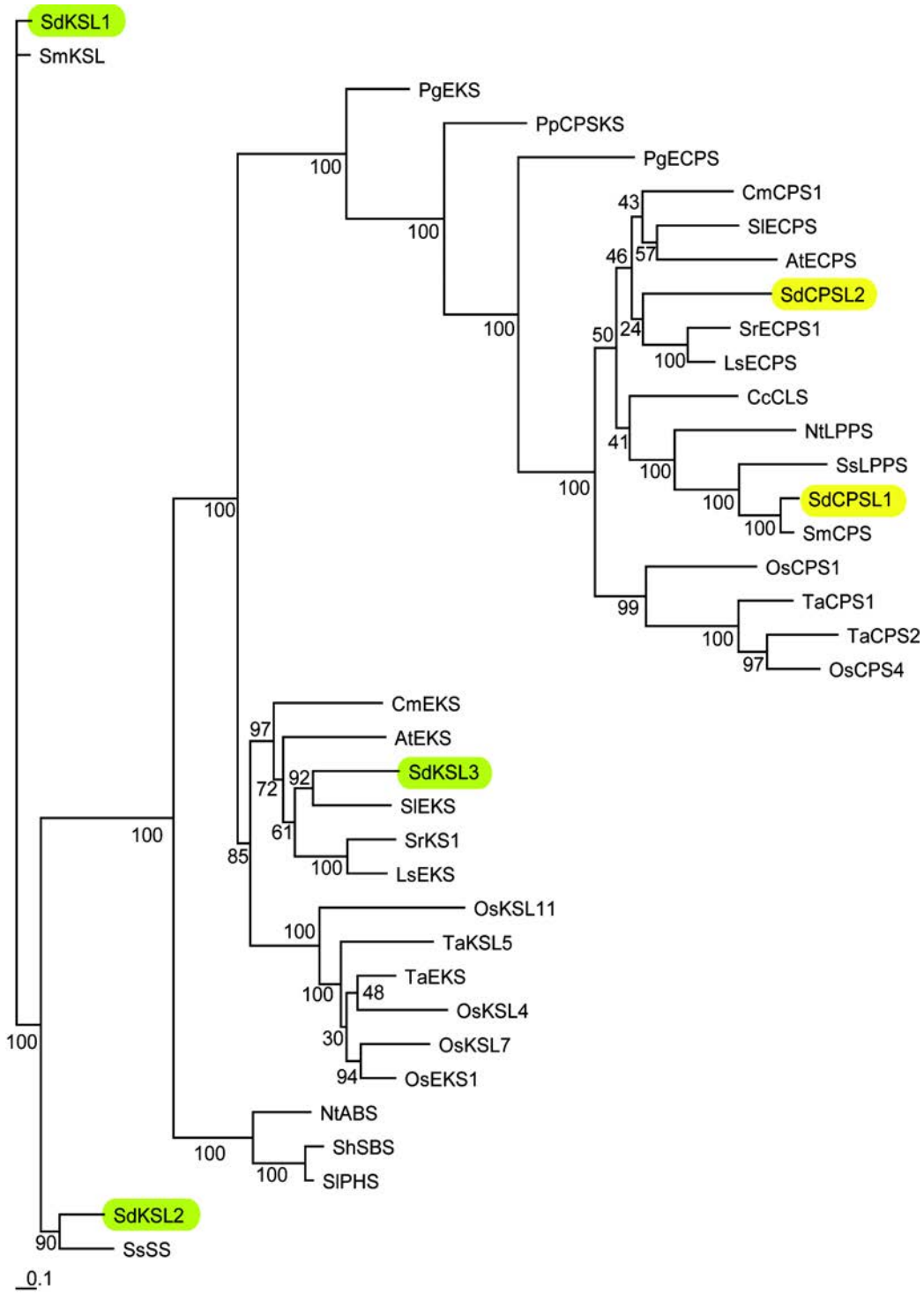


Figure 17. Diterpene synthase phylogenetic tree

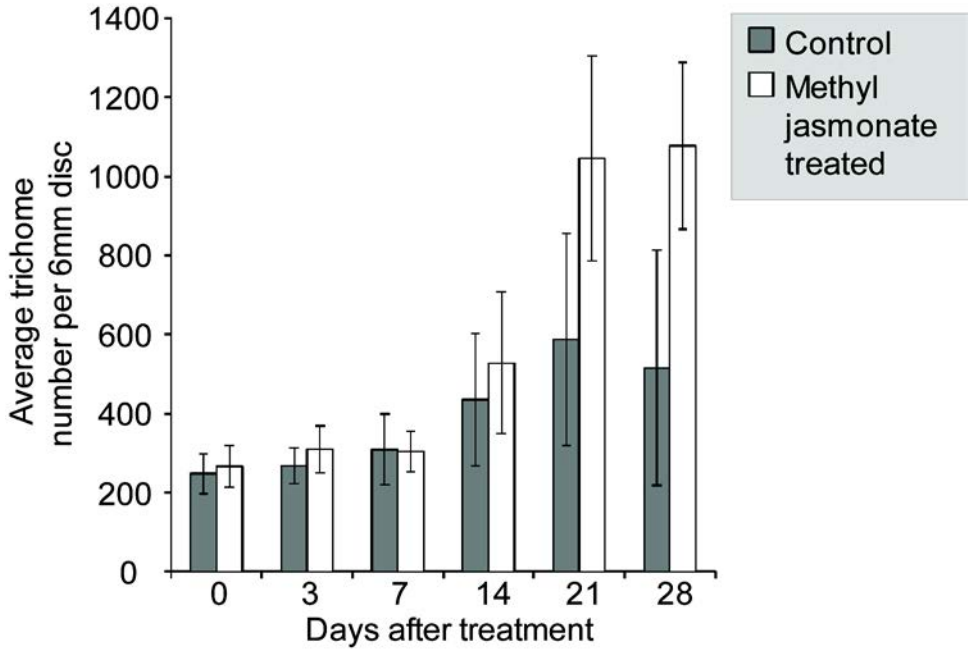
3.6 Methyl jasmonate induction of *S. divinorum*

Interestingly, *KSL2* and *KSL3* RACE reactions yielded PCR products only when methyl jasmonate induced leaves were used to prepare RACE ready cDNA. This suggested that methyl jasmonate treatment increased the abundance of *KSL2* and *KSL3* transcripts in *S. divinorum* leaves. Methyl jasmonate is a volatile signalling molecule that mediates developmental and defense processes in many plants [89]. Exposure to the chemical can also induce trichome formation [90].

As salvinorin A is stored in the *S. divinorum* peltate glandular trichomes, a method to increase the density of these structures would facilitate future studies into the salvinorin A biosynthetic pathway. *S. divinorum* plants were exposed to methyl jasmonate in an effort to assay the efficacy of the technique in this regard. Leaves harvested 3 and 4 weeks after methyl jasmonate treatment were found to have approximately twice the peltate glandular trichome density observed in the control plants at those time points (Figure 18). Leaves harvested at the 3 and 4 week time points were likely very young and still initiating new trichomes when the plants were exposed to methyl jasmonate. Trichome density in the experimental plants was also higher than trichome density in the control plants at the 2 week time point, but not to a significant extent (an average of 538±179 trichomes per 6 mm disc versus an average of 435±167).

Figure 18. *S. divinorum* peltate glandular trichome induction by methyl jasmonate

N = Four plants per treatment, three leaves per plant per time point. Densities calculated from counts of trichomes on three 6 mm leaf discs per leaf. All values represent an average of density counts for each treatment at each time point. Methyl jasmonate treated plants were sprayed with 50 ml 1 mM methyl jasmonate and 0.1% tween 20 solution, and control treated plants were sprayed with 50 ml 0.1% tween 20 solution. Plants were covered with a plastic bag for 90 min after spraying, and then uncovered for 2 hr. Control plants and methyl jasmonate treated plants were grown in opposite ends of the University of Calgary greenhouse.



3.7 Quantitative real-time PCR (qPCR) for candidate diterpene synthases

As the peltate glandular trichomes were found to secrete salvinorin A, it was expected that the genes responsible for salvinorin A biosynthesis are expressed in *S. divinorum* trichome secretory cells. qPCR were performed for each candidate diterpene synthase, in order to assay their tissue specific expression in *S. divinorum.* This work was performed in collaboration with Dr. Philipp Zerbe (Bohlmann laboratory, University of British Columbia).

CPPSL2, *KSL2,* and *KSL3* were found to be specifically expressed in *S. divinorum* trichomes (Figure 19). This result suggested that these three genes could be involved in trichome-specific salvinorin A biosynthetic pathways. *CPPSL2*, *KSL2,* and *KSL3* were therefore excellent candidates for further biochemical characterization to assess their activities. Conversely, predominant expression of *CPPSL1* was observed in the roots, and transcripts of *KSL1* were detected in leaves, roots, and stem as well as trichomes (Figure 19). Therefore, these two genes are not likely to be involved in salvinorin A biosynthesis in *S. divinorum. CPPSL1* and *KSL1* are likely *S. divinorum's* endogenous *ent-CPPS* and *ent-KS* involved in the biosynthesis of the plant hormone gibberellin.

Figure 19. Tissue specific abundance of candidate diterpene synthase transcripts in *S. divinorum*

The expression levels of *CPPSL1, CPPSL2, KSL1, KSL2,* and *KSL3* in *S. divinorum* root, stem, leaf, and trichome tissue was measured using qPCR, and graphed as fold changes relative to actin controls. All values represent an average of two independent experiments. Crystal clear adhesive tape (3M) was used to remove trichomes from the abaxial surface of *S. divinorum* leaves.

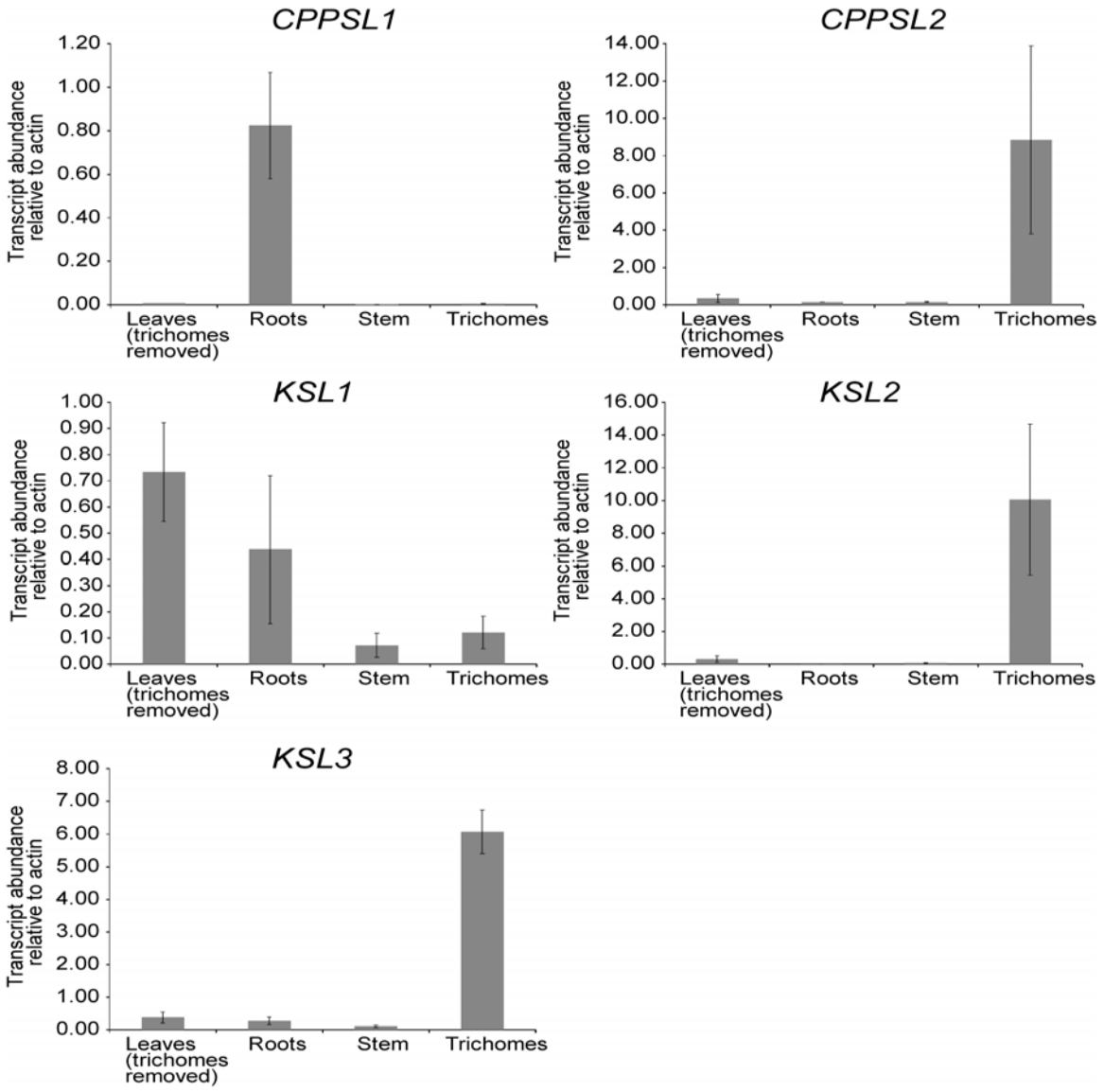


Figure 19. Tissue specific abundance of candidate diterpene synthase transcripts in *S. divinorum*

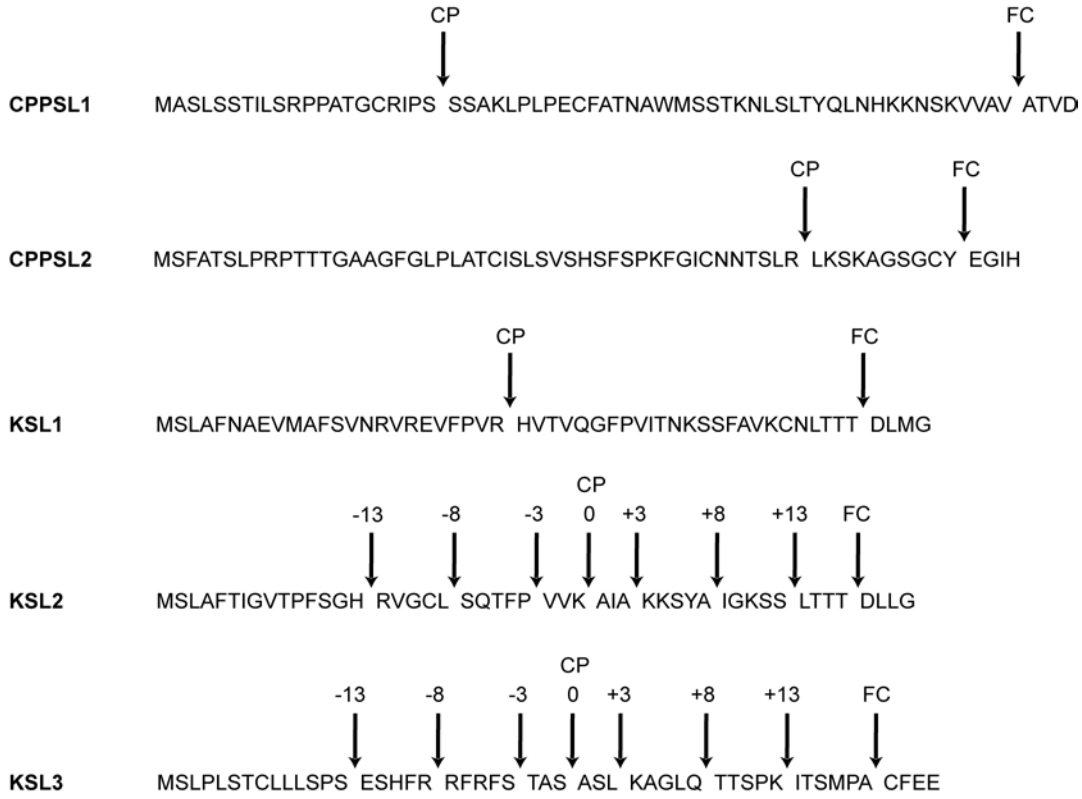
3.8 Production and characterization of recombinant CPPSL1/2

Diterpene synthases are localized to the chloroplasts in plants by N-terminus transit peptides that are cleaved as the enzymes enter the organelle. The

heterologous expression of diterpene synthases in *E. coli* or yeast often requires the removal of these transit peptides, as they can interfere with proper folding of the protein [91]. Chloroplast transit peptides are difficult to conclusively identify, and two different methods were therefore used to predict their length in the *S. divinorum* candidate diterpene synthases. These predictions are shown in Figure 20. In the first method, features common to chloroplast transit peptides, such as the absence of acidic amino acids [91], were used to roughly estimate their length (FC prediction in Figure 20). The second method used the ChloroP neural network [92] to predict the length of each chloroplast transit peptide (CP prediction in Figure 20). Several additional predictions were also prepared for KSL2 and KSL3 by using the ChloroP prediction as a baseline (numbered predictions in Figure 20). The +3 prediction for chloroplast transit peptide length, for example, is three amino acids longer than the ChloroP prediction, and the -3 prediction is three amino acids shorter. The probability of creating a highly active pseudomature enzyme is increased when several different chloroplast transit peptide predictions are used.

Figure 20. Chloroplast transit peptide predictions for candidate diterpene synthases

FC = Features common to chloroplast transit peptides were used to generate this prediction [91]. CP = ChloroP neural network prediction for chloroplast transit peptide length [92]. Numbered predictions were created using the ChloroP prediction as a baseline.

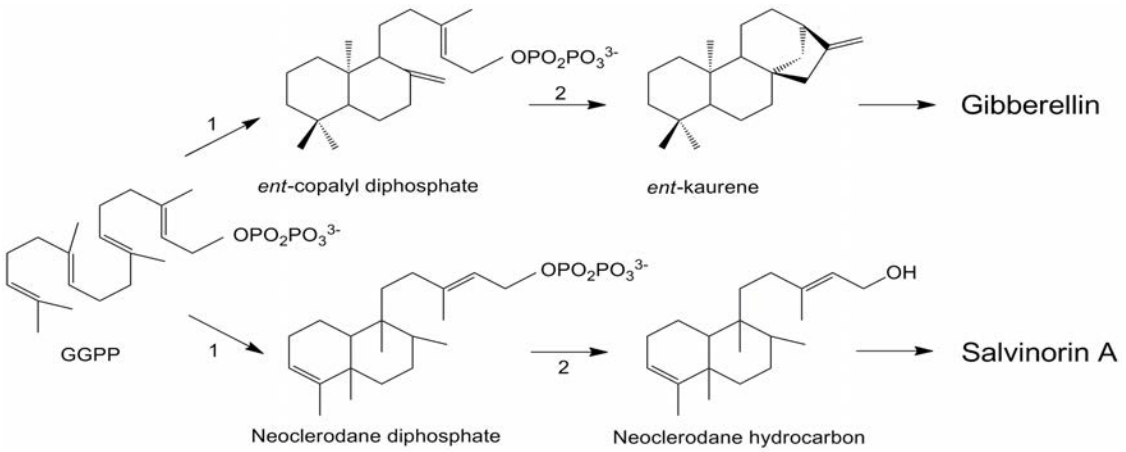


The type II diterpene synthase candidates (CPPSL1 and CPPSL2) were expected to catalyze a protonation initiated reaction that yields diterpene diphosphates. The type I diterpene synthase candidates (KSL1, KSL2, and KSL3) were then expected to convert these diterpene diphosphates into diterpene

hydrocarbons through an ionization (i.e., de-phosphorylation) initiated reaction. The general scheme of these type II and type I diterpene synthase reactions (in comparison to the well characterized gibberellin pathway) is shown in Figure 21. Each KSL enzyme uses the product of a CPPSL enzyme as a substrate, and KSL biochemical activity therefore cannot be evaluated until the relevant CPPSL is characterized. The substrate for CPPSL1/2 (i.e., GGPP) is commercially available, and initial biochemical efforts were focused on the characterization of these two enzymes.

Figure 21. Diterpene synthase catalyzed reactions in the gibberellin and hypothesized salvinorin A biosynthetic pathways

Reactions catalyzed by type II diterpene synthases (1), and type I diterpene synthases (2) in the gibberellin and hypothesized salvinorin A biosynthetic pathways.



Six histidine residues were added to the N-terminus of CPPSL1/2 after removal of their putative transit peptides (FC prediction in Figure 20), and an *E. coli* expression system was used to produce these pseudomature recombinant enzymes. When recombinant proteins are biosynthesized in *E. coli,* there is a tendency for insoluble aggregates called “inclusion bodies” to form. To identify the optimal culture conditions for minimal inclusion body formation, *E. coli* expressing *CPPSL1* and *CPPSL2* was cultivated at 37°C, 16°C, or 4°C for 3 to 24 hr. Cultivation at 37°C and 4°C was not suitable for recombinant enzyme production, as no CPPSL1/2 protein was observed in all 4°C time points (Figure 22A), and insoluble CPPSL1/2 inclusion bodies were produced in all 37°C time points (Figure 22B). Cultivation for 4 hr at 16°C however resulted in a small amount of soluble CPPSL1/2 protein (*1 and *2 in Figure 22D, respectively). These proteins were not detectible in the 3 hr incubation (Figure 24C) and were forming inclusion bodies in the 5 hr incubation (Figure 24E). A 4 hr at 16°C cultivation condition was therefore used to produce the first batch of recombinant CPPLS1/2. Enzymes were purified by Ni^{2+} affinity column.

Figure 22. Effect of induction time and temperature on production of soluble CPPSL1 and CPPSL2 protein in *E. coli*

*1 = Soluble CPPSL1 protein. *2 = Soluble CPPSL2 protein. 0.2% w/v arabinose was used to induce the expression of pseudomature (FC prediction in Figure 20) recombinant CPPSL1/2 protein in BL21-AI (Invitrogen) *E. coli* for 24 hr at 4°C (A), 3 hr at 37°C (B), 3 hr at 16°C (C), 4 hr at 16°C (D), and 5 hr at 16°C (E). Cultures were grown to an OD_{600} of 0.4 in TB media before induction of expression.

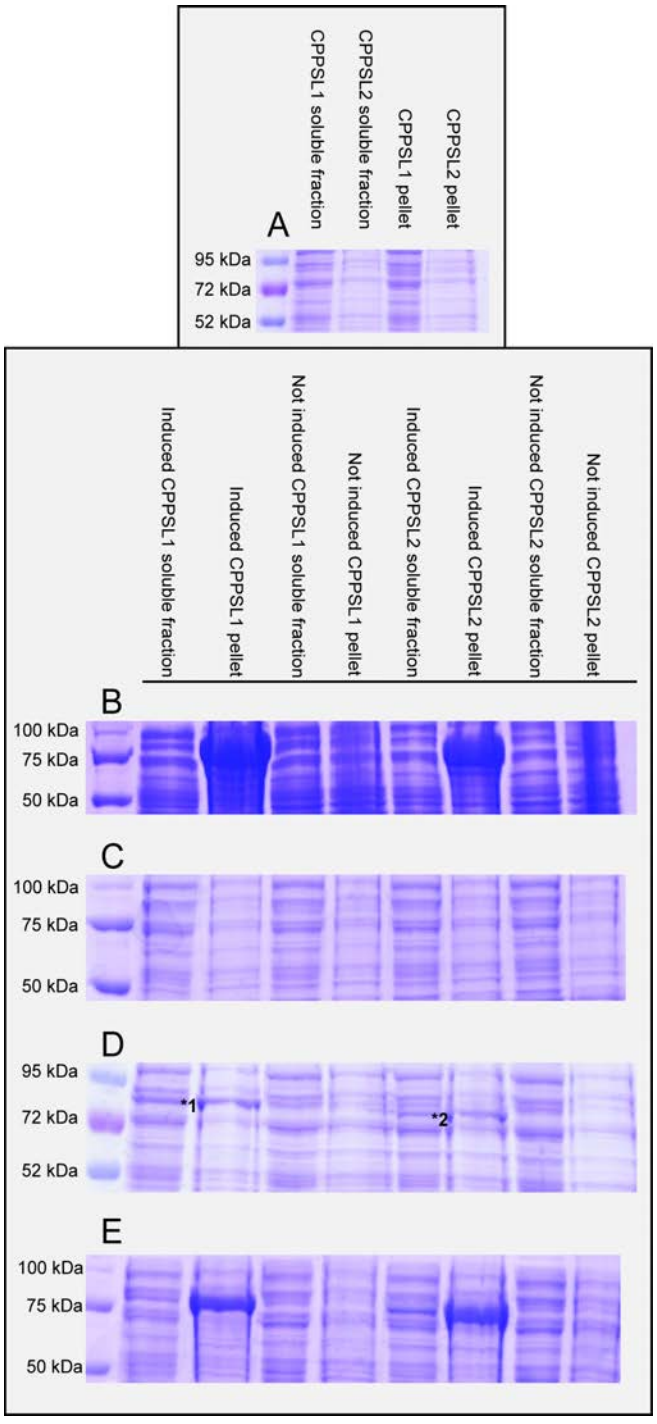
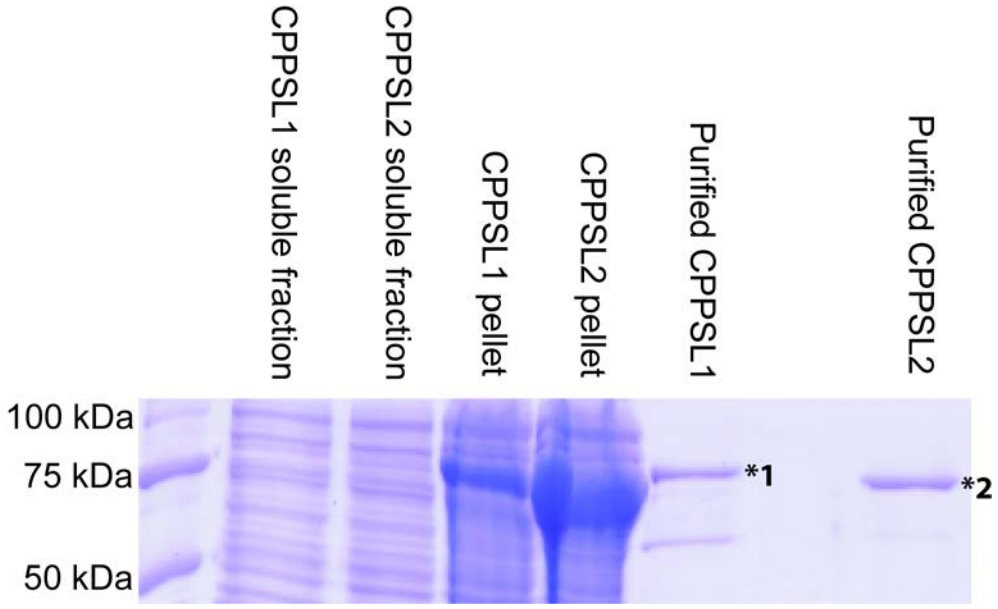


Figure 22. Effect of induction time and temperature on production of soluble CPPSL1 and CPPSL2 protein in *E. coli*

The yield of CPPSL1/2 protein using this induction method was minimal, less than 200 μg per L culture. To increase the yield, induction time was lengthened to 24 hr at 16°C. CPPSL1/2 inclusion body formation was significantly increased by this extended incubation period, and soluble CPPSL1/2 protein was not visible on representative SDS-PAGE gels (Figure 23). Nonetheless, when these protein extracts were used for affinity purification, more than 1 mg soluble recombinant CPPLS1/2 was purified per L culture (Figure 23). The CPPSL1 enzymes were purified to ~90% purity with one major contaminant, and the CPPSL2 enzymes were purified to near 99% purity. Therefore, the condition established (24 hr induction at 16°C) was routinely used to produce CPPSL1/2 recombinant enzymes in subsequent purifications.

Figure 23. Production of CPPSL1 and CPPSL2 protein in *E. coli* cultures induced for 24 hr at 16°C

*1 = Soluble CPPSL1 protein. *2 = Soluble CPPSL2 protein. Cultures were grown to an OD_{600} of 0.4 in 1 L TB media, and 0.2% w/v arabinose was then added to induce expression of pseudomature (FC prediction in Figure 20) recombinant CPPSL1/2 protein. Proteins were purified by Ni^{2+} affinity chromatography.

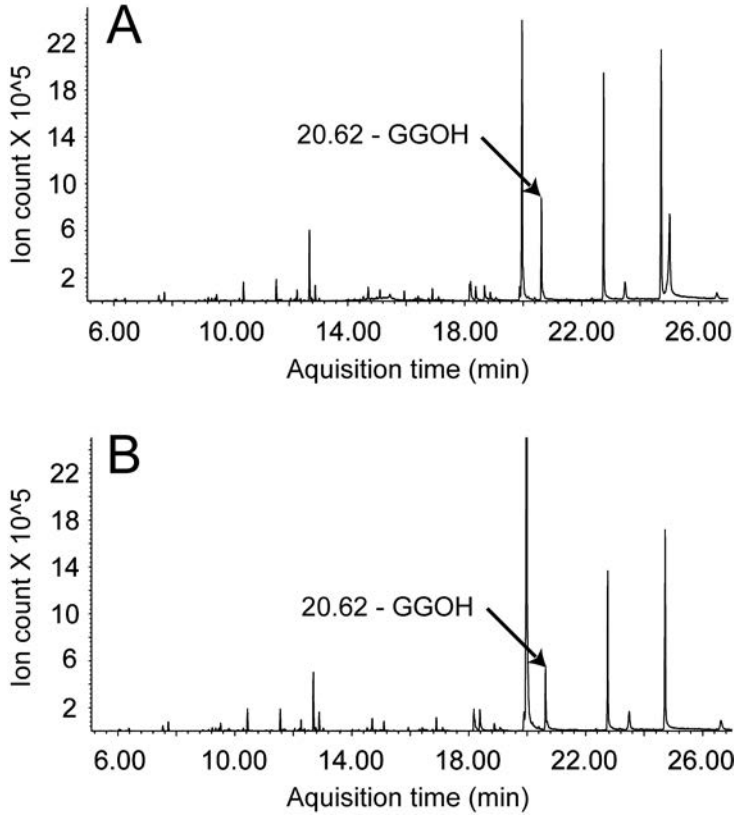


Purified CPPSL1 and CPPSL2 enzymes from *E. coli* cultures were utilized in *in vitro* assays using GGPP as a substrate. As type II diterpene synthases (e.g., CPPS) convert GGPP into a diphosphate-bearing product, completed enzyme reaction mixtures were treated with phosphatase to release de-phosphorylated hydrocarbon compounds (C20 diterpenes). This de-phosphorylation reaction converts diterpene diphosphates into diterpene alcohols that are detectable by GC-MS analysis. In the CPPSL1 assay, the GC-MS product profiles from boiled

CPPSL1 control and native CPPSL1 were qualitatively identical (Figure 24). Similar amounts of geranyl geranyl alcohol (GGOH), the de-phosphorylated GGPP substrate, were also observed in both assays. Therefore, it was concluded that no activity was detected from the pseudomature recombinant CPPSL1 enzyme.

Figure 24. GC-MS analysis of CPPSL1 *in vitro* assays

Chromatograms of native CPPSL1 *in vitro* assay (A) and boiled CPPSL1 control (B). 100 µg CPPSL1 enzyme, 25 µM GGPP and a 1 hr incubation at 30°C was used to perform each *in vitro* assay. Assays were de-phosphorylated at 37°C for 12 hr, and semi-purified by pentane extract. Analysis of the extractions was performed on an Agilent DB-1ms column using a 27 min oven program.



When the same enzyme assays were completed using purified CPPSL2, the CPPSL2 was able to catalyze the conversion of GGPP to a novel diterpenoid which was not present in the boiled control (Figure 25A and B). An additional compound eluting at 8.62 min was also observed in the CPPSL2 *in vitro* assay.

This early-eluting compound was found to have a mass spectrum very similar to that of the CPPSL2 product (Figure 25C and D), and is believed to be a thermal degradation of the compound. In support of this, LC-MS analysis of the CPPSL2 *in vitro* assay showed only a single product peak (see below). The mass spectra of these two new peaks were an excellent match (>95%) to the kolavelool in the mass spectrometry library (Figure 25C, D, and E), indicating that the skeleton of the CPPSL2 product is very closely related to the kolavelool backbone. Kolavelool and its relative natural products are shown in Figure 26. The kolavenol derivative is of particular interest, as its structure has been studied by two independent groups [93] [94]. Kolavenol was proposed as a possible intermediate in the salvinorin A pathway in Figure 9.

The mass spectrum of the CPPSL2 product revealed that this novel metabolite has a parental ion of *m/z* 290 (Figure 25C). Given that the diterpene hydrocarbon of a cyclized GGPP compound has a mass of 273, this parental ion suggests that the CPPSL2 product has a single hydroxyl group (mass 17). The CPPSL2 *in vitro* assays were de-phosphorylated by phosphatase prior to GC-MS analysis, converting diterpene diphosphates present in the assays into diterpene alcohols. The CPPSL2 products single hydroxyl group was therefore generated by the de-phosporylation reaction. As CPPSL2 is a type II diterpene synthase, it drives rearrangement of GGPP by creating a carbocation (Figure 6). Terpene synthases commonly stabilize this reactive carbocation intermediate by water quench, creating a hydroxyl, or by deprotonation and double-bond

rearrangmement. The CPPSL2 product has only a single hydroxyl group, and the CPPSL2 product carbocation must be quenched by deprotonation and double-bond rearrangment as a result. The CPPSL2 product was not detected when *in vitro* assays were not de-phosphorylated, demonstrating that conversion of diterpene diphosphates into diterpene alcohols is a necessary step for GC-MS detection of the compounds.

Figure 25. GC-MS analysis of CPPSL2 *in vitro* assays

Chromatograms of native CPPSL2 *in vitro* assay (A) and boiled CPPSL2 control (B). Mass spectra of the peaks at 11.28 min and 8.62 min in the native CPPSL2 *in vitro* assay chromatogram are shown in C and D, respectively. The mass spectrum of kolavelool (obtained from the NIST2/Wiley 5 mass spectra library [95]) is shown in E. 10 µM 1-eicosane (Sigma) was added to each assay as an internal control. 100 µg CPPSL2 enzyme, 25 µM GGPP and a 1 hr incubation at 30°C was used to perform each *in vitro* assay. Assays were de-phosphorylated at 37°C for 12 hr, and semi-purified by pentane extract. Analysis of the extractions was performed on an Agilent DB-5ms column using a 15.5 min oven program.

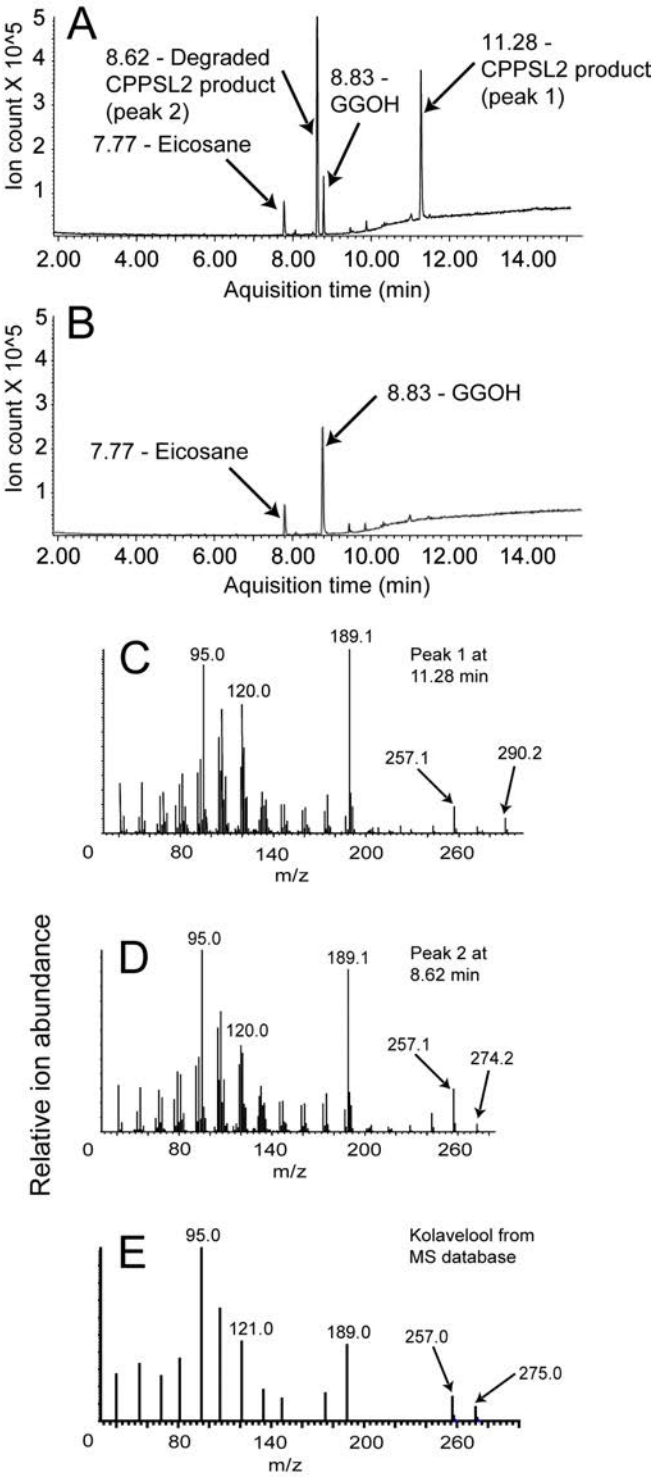


Figure 25. GC-MS analysis of CPPSL2 *in vitro* assays

The CPPSL2 enzyme product was further analyzed by LC-MS to examine its mass by a soft ionization method. LC-MS analysis of the CPPSL2 *in vitro* assay identified two compounds with *m/z* of 273.2 in (+)-ion mode (Figure 26C and D). Parallel analysis of a GGOH standard confirmed that the compound eluting at 8.82 min was GGOH. The compound eluting at 8.44 min was determined to be the CPPSL2 product, as it was not present in the boiled CPPSL2 control (Figure 26A and B). Significant differences in mass spectra were observed for the CPPSL2 product and GGOH, demonstrating that GGPP was cyclized and rearranged by the CPPSL2 enzyme (Figure 26C and D). In (+)-LC-MS, water loss (-18 mass) commonly occurs if a hydroxyl group is present in the analyte. Therefore, both de-phosphorylated CPPSL2 product and GGOH underwent dehydration reactions (-18 mass) in the (+)-ion mode analysis, and their $[M\text{-}H_2O+H]^+$ ions were detected as *m/z* 273 (Figure 26C and D). This result demonstrated that the CPPSL2 product likely contains a single hydroxyl group. Loss of the C_6H_{10} side chain (mass 82) in the CPPSL2 product (Figure 27) also occurred, resulting in the detection of a $[M\text{-}H_2O\text{-}C_6H_{10}+H]^+$ ion of *m/z* 191 (Figure 26C). The carbons of linear and branched hydrocarbons are commonly lost before the carbons of hydrocarbon ring structures, as the former are more susceptible to ionization. These ionization results indicated that the structure of the CPPSL2 product is more similar to kolavenol and kolavelool than to other kolavelool derivatives such as kolavenoic acid (Figure 27). No signal was detected in (-)-ion mode, probably due to the compound's inability to form a negative ion in the given ionization conditions.

Figure 26. LC-MS analysis of CPPSL2 *in vitro* assays

Chromatograms of native CPPSL2 *in vitro* assay (A) and boiled CPPSL2 control (B). Chromatograms were extracted at a *m/z* of 273. Mass spectra of the peaks at 8.44 min and 8.82 min in the native CPPSL2 *in vitro* assay chromatogram are shown in C and D, respectively. 100 µg CPPSL2 enzyme, 25 µM GGPP and a 1 hr incubation at 30°C was used to perform each *in vitro* assay. Assays were de-phosphorylated at 37°C for 12 hr, semi-purified by pentane extract, and resuspended in 50% methanol. Analysis was performed on a Zorbax Eclipse Plus C18 column.

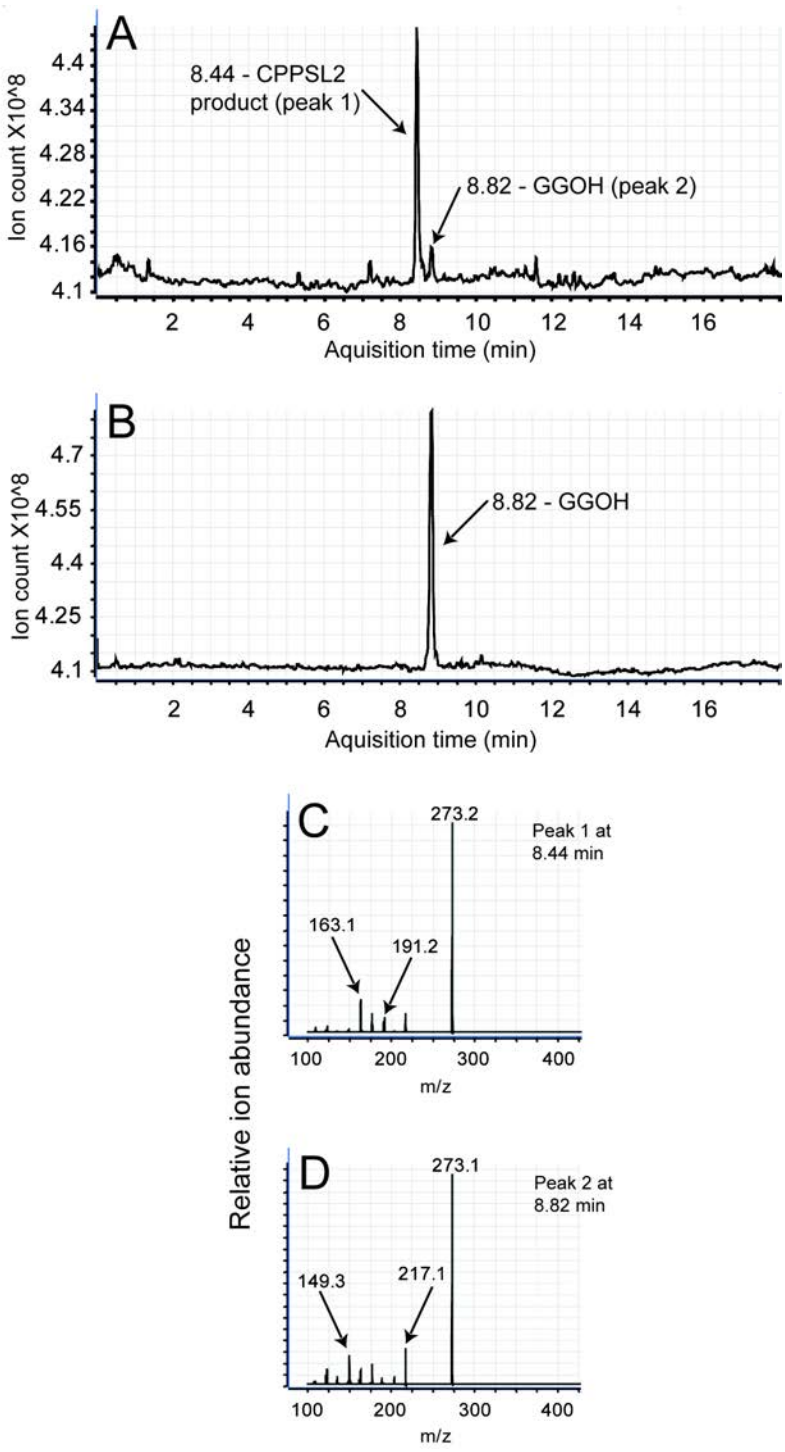
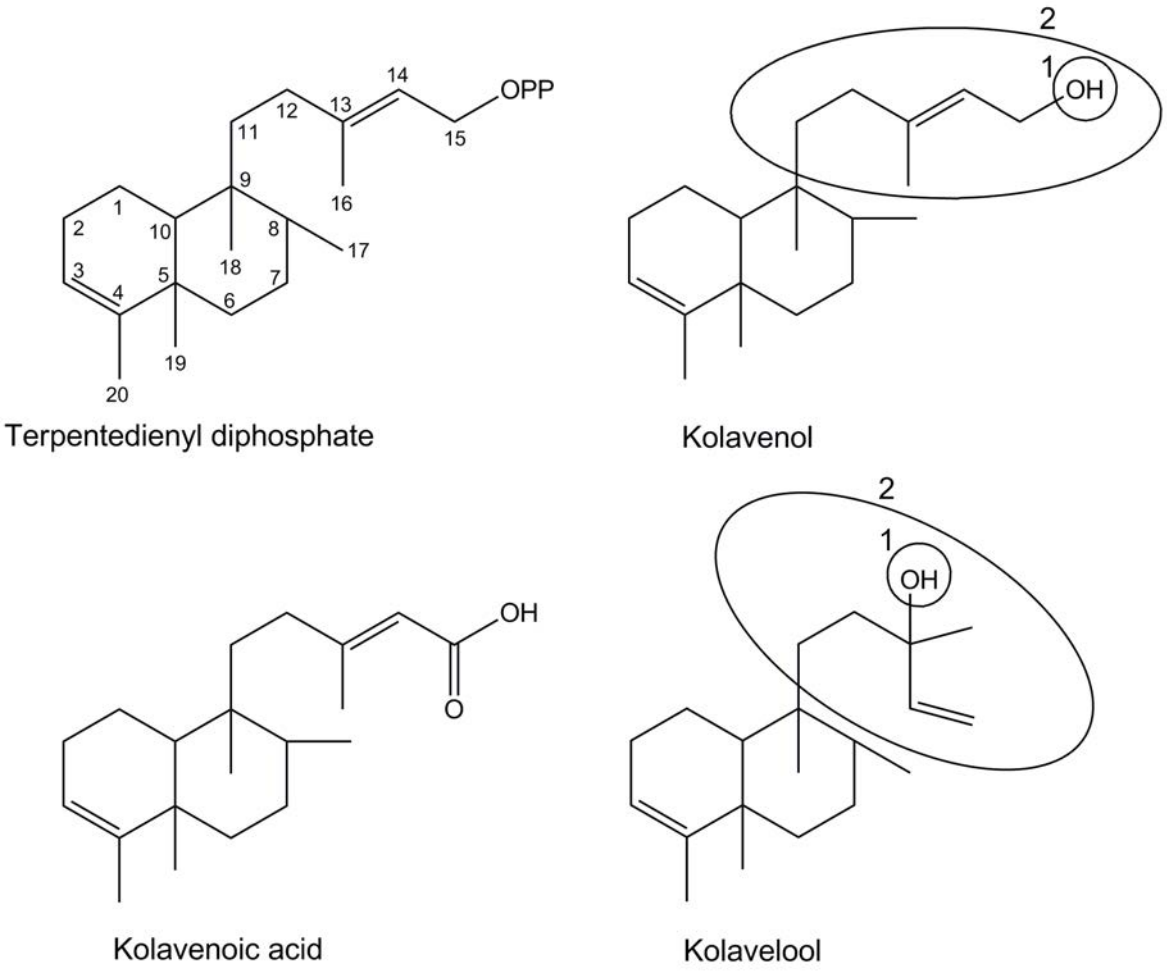


Figure 26. LC-MS analysis of CPPSL2 *in vitro* assays

Figure 27. Chemical structures of kolavelool derivatives

CPPSL2 product water loss (1) generated the $[M\text{-}H_2O+H]^+$ ion (*m/z* 273) observed in Figure 26C. Additional loss of the C_6H_{10} side chain (2) resulted in the $[M\text{-}H_2O\text{-}C_6H_{10}+H]^+$ ion (*m/z* 191). Analysis was performed on a Zorbax Eclipse Plus C18 column.



In summary, both GC- and LC-MS analyses showed that CPPSL2 enzyme can catalyze the rearrangement of GGPP to a novel hydroxylated diterpene compound which displays a mass spectral match to kolavelool and has a molecular

mass of 290. De-phosphorylation was necessary to analyze the CPPSL2 product by GC- and LC-MS, further suggesting that CPPSL2 can evoke the protonation driven carbocation reaction as predicted from its primary sequence data.

3.9 Production and characterization of recombinant KSL2/3

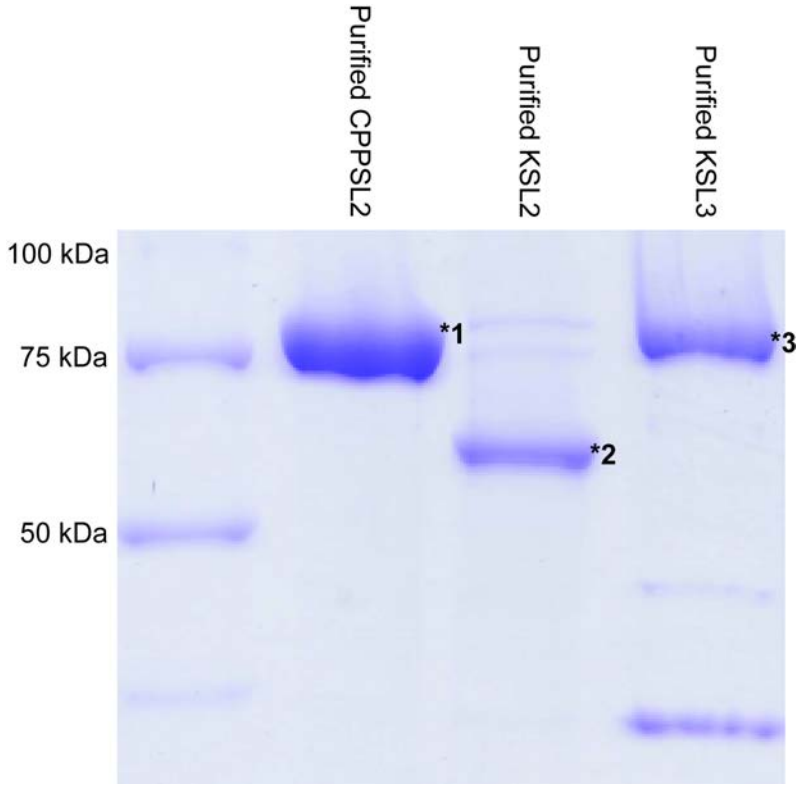
As expression and preliminary characterization of CPPSL2 was successful, the enzyme was used in co-incubation *in vitro* assays to generate substrates for type I diterpene synthases. Type I diterpene synthases, such as the KSLs, de-phosphorylate the diterpene diphosphates that type II diterpene synthases produce (Figure 21). The ability of the KSLs to use the CPPSL2 product as a substrate was therefore biochemically evaluated by *in vitro* assay.

The ChloroP neural network was used to prepare seven chloroplast transit peptide predictions for KSL2 and KSL3 as shown in Figure 20 (numbered predictions). Protein expression for all fourteen individual constructs was induced for 24 hr at 16°C by addition of 0.2% w/v arabinose. All seven constructs of pseudomature KSL2 were successfully Ni^{2+} affinity purified. However, for unknown reasons, only the KSL3 construct which had a thirty five residue deletion in its N-terminus (+8 prediction in Figure 20) yielded significant amounts of purified protein. Size fractionation of the other six cultures by SDS-PAGE showed that production of these other pseudomature KSL3 proteins in *E. coli* was very poor. The KSL2 enzymes were all purified to a near 99% purity, and the KSL3 enzyme was purified to ~90% purity with one major contaminant. An example of SDS-PAGE

fractionation of pseudomature KSL2 (0 prediction) and KSL3 (+8 prediction) is shown in Figure 28.

Figure 28. Purification of CPPSL2, KSL2 and KSL3 protein in *E. coli* cultures induced for 24 hr at 16°C

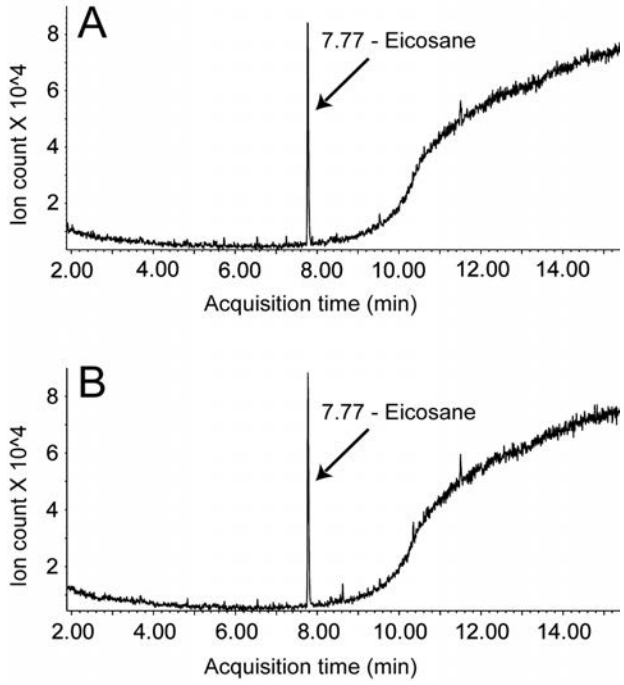
*1 = Soluble CPPSL2 (FC prediction) protein. *2 = Soluble KSL2 (0 prediction) protein. *3 = Soluble KSL3 (+8 prediction) protein. 0.2% w/v arabinose was used to induce the expression of pseudomature recombinant protein in BL21-AI (Invitrogen) *E. coli* for 24 hr at 16°C. Cultures were grown to an OD_{600} of 0.4 in 1 L TB media before induction of expression. Proteins were purified by Ni^{2+} affinity chromatography.



Co-incubation *in vitro* assays including purified KSL2 or KSL3 enzyme with recombinant CPPSL2 and GGPP substrate were performed with the assumption that CPPSL2 product can serve as a substrate for either KSL2 or KSL3. When the reaction products were analyzed by GC-MS, it was found that none of the seven pseudomature KSL2 enzymes were able to catalyze the conversion of the CPPSL2 product to a unique diterpene. One example of the negative data is shown in Figure 29.

Figure 29. GC-MS analysis of CPPSL2 and KSL2 co-incubation *in vitro* assays

Chromatograms of native CPPSL2 and KSL2 co-incuation *in vitro* assay (A) and boiled KSL2 control (B). 10 µM 1-eicosane (Sigma) was added to each assay as an internal control. 100 µg CPPSL2 and KSL2 enzyme, 25 µM GGPP and a 1 hr incubation at 30°C was used to perform each *in vitro* assay. Assays were semi-purified by pentane extract and analyzed on an Agilent DB-5ms column using a 15.5 min oven program.



Similarly, the purified KSL3 (+8) enzyme was co-incubated with CPPSL2 protein and GGPP, and the reaction products were analyzed by GC-MS. In the GC-MS chromatogram, two small peaks unique to the assay sample were identified (Figure 30A). Interestingly, the mass spectra of these two peaks were

determined to be essentially identical to that of the de-phosphorylated CPPSL2 product and its degraded derivative described in previous section (3.7) (Figure 30E and F). This result implied that KSL3 can cleave the diphosphate moiety of the CPPSL2 product, generating a C20 diterpene carbocation that seems to be immediately quenched by water molecule. Therefore, KSL3 de-phosphorylates the CPPSL2 product but does not initiate any further rearrangements of the diterpene backbone.

Figure 30. GC-MS analysis of CPPSL2 and KSL3 co-incubation *in vitro* assays

Chromatograms of native CPPSL2 and KSL3 *in vitro* assay (A and C) and boiled KSL3 control (B and D). A and B chromatograms were extracted at a *m/z* of 257 to show the KSL3 enzymatic products, and C and D chromatograms were extracted at a *m/z* of 280 (parental ion of 1-eicosane) to show the 10 μM 1-eicosane (Sigma) internal control. Mass spectra of the peaks at 11.21 min and 8.53 min in the native CPPSL2 and KSL3 *in vitro* assay are shown in E and F, respectively. 100 μg CPPSL2 and KSL3 enzyme, 25 μM GGPP and a 1 hr incubation at 30°C was used to perform each *in vitro* assay. Assays were semi-purified by pentane extract and analyzed on an Agilent DB-5ms column using a 15.5 min oven program.

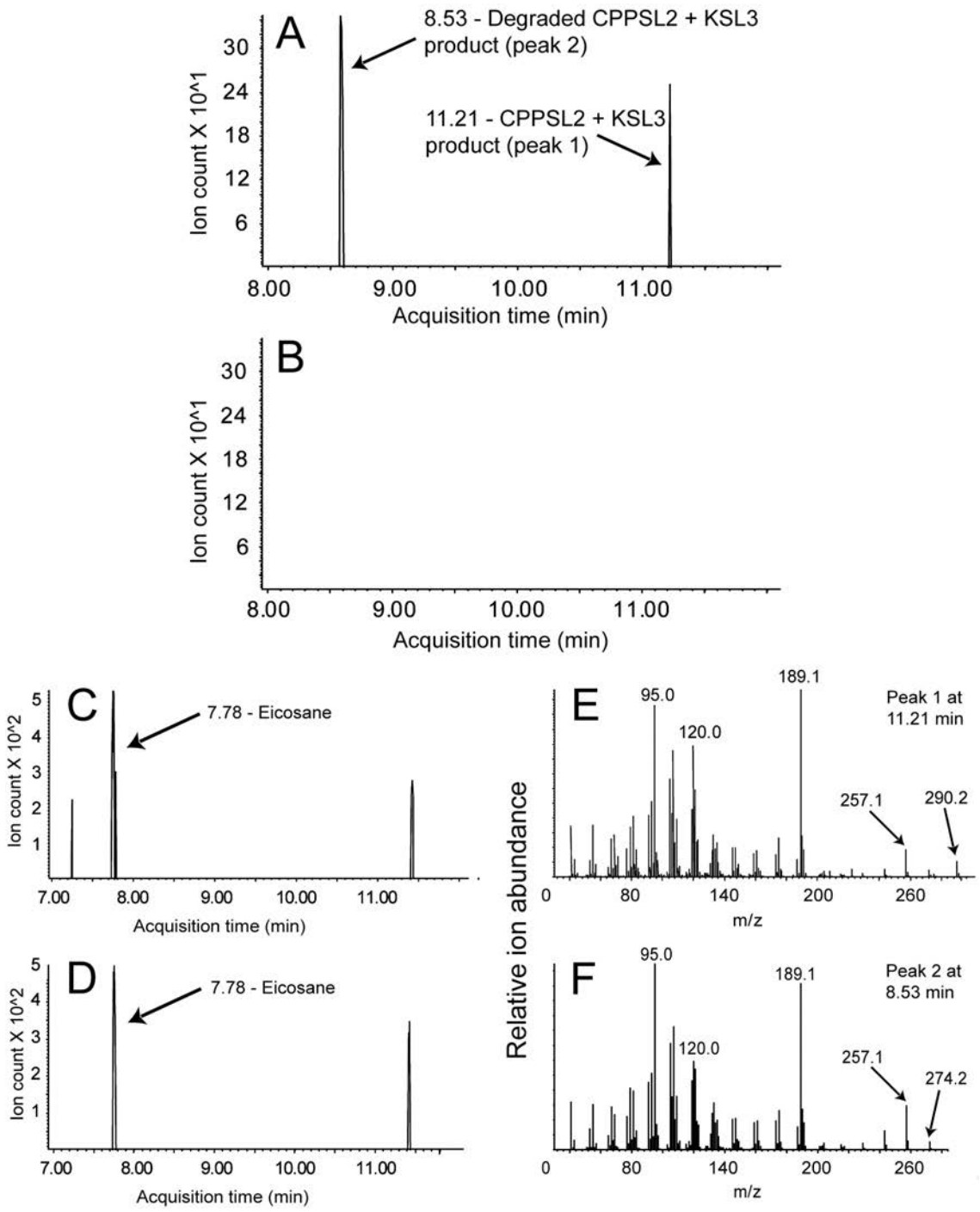


Figure 30. GC-MS analysis of CPPSL2 and KSL3 co-incubation *in vitro* assays

To evaluate possible cross-reactivity of KSL2 and KSL3 with other diterpene diphosphate substrates, previously characterized type II diterpene synthases were used to prepare two diterpene diphosphate substrates. For this purpose, pseudomature *Zea mayez ent*-CPPS (AAT70084.1 [96]) and *Picea abies* (+)-CPPS (a variant of AAS47691.1 [97]) proteins were expressed in *E. coli* and affinity purified. *In vitro* assays using GGPP substrate and the purified proteins were then performed to confirm that the enzymes were active. These diterpene synthase constructs were supplied by Dr. Philipp Zerbe (Bohlmann laboratory, University of British Columbia). In both assays, expected diterpene alcohols were detected after de-phosphorylation by phosphatase, and product mass spectra were identical to published data [95] (Figure 31B and E). The structures of these *ent*-CPP and (+)-CPP products are shown in Figure 31C and F, respectively.

Figure 31. GC-MS analysis of *ent*-CPPS and (+)-CPPS *in vitro* assays

Chromatograms of *ent*-CPPS (AAT70084.1 [96]) (A) and (+)-CPPS (a variant of AAS47691.1 [97]) (D) *in vitro* assays. Mass spectra of the peak at 10.90 min in the *ent*-CPPS assay chromatogram, and the peak at 13.52 min in the (+)-CPPS assay chromatogram are shown in B and E, respectively. The structures of *ent*-CPP (C) and (+)-CPP (F) are also shown. 10 μM 1-eicosane (Sigma) was added to each assay as an internal control. 100 μg of each purified enzyme, 25 μM GGPP and a 1 hr incubation at 30°C was used to perform each *in vitro* assay. Assays were de-phosphorylated at 37°C for 12 hr, and semi-purified by pentane extract. Analysis of the extractions was performed on an Agilent DB-5ms column using a 15.5 min oven program.

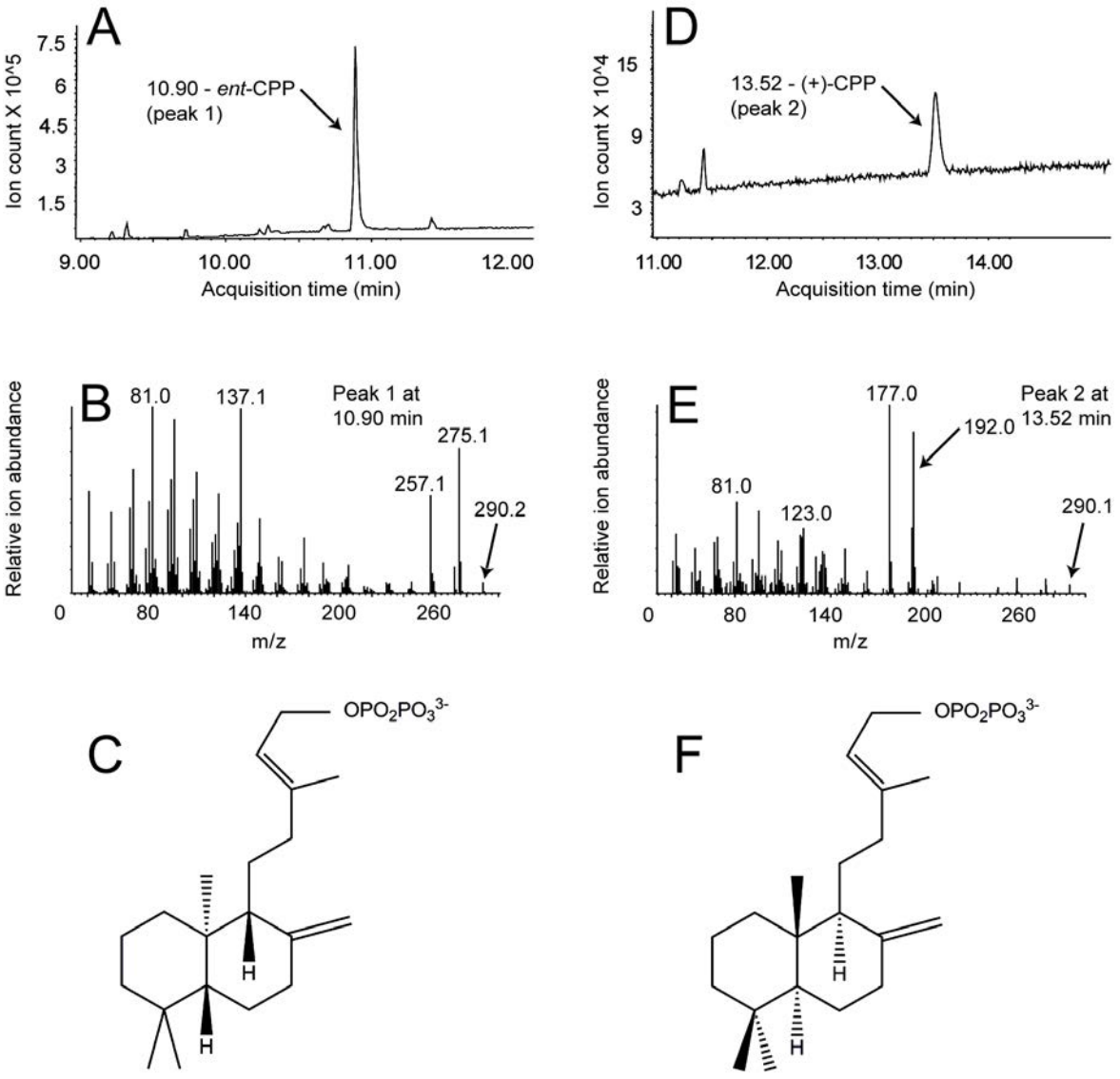


Figure 31. GC-MS analysis of *ent*-CPPS and (+)-CPPS *in vitro* assays

Purified pseudomature *ent*-CPPS and (+)-CPPS enzymes were incubated with recombinant KSL2 and KSL3 in co-incubation *in vitro* assays, and metabolite profiles were analyzed by GC-MS. The assay results showed that KSL2 was not able to use *ent*-CPP or (+)-CPP as a substrate, and KSL3 also could not render

ent-CPP into an enzymatic product. Interestingly, co-incubation of KSL3 and (+)-CPPS resulted in the synthesis of small amounts of two novel compounds, which eluted very closely but displayed different mass fragmentation patterns (Figure 32E and F). The mass fragmentation patterns of these compounds were found to be substantially different from the de-phosphorylated substrate, (+)-CPP (Figure 33C versus Figure 33D and E). A possible interpretation of this result is that KSL3 catalyzes the skeletal rearrangement of (+)-CPP by de-phosphorylation initiated carbocation reaction. In contrast, the mass fragmentation patterns of the de-phosphorylated CPPSL2 product and the KSL3 enzymatic conversion of the CPPSL2 product were identical (Figure 33A and B), indicating that KSL3 likely does not catalyze the structural rearrangement of the CPPSL2 product. The abundance of these KSL3 enzymatic products was however small, and further structural study could not be achieved..

Figure 32. GC-MS analysis of (+)-CPPS and KSL3 co-incubation *in vitro* assays

Chromatograms of native (+)-CPPS and KSL3 *in vitro* assay (A and C) and boiled KSL3 control (B and D). A and B chromatograms were extracted at a *m/z* of 275 to show the KSL3 enzymatic products, and C and D chromatograms were extracted at a *m/z* of 280 (parental ion of 1-eicosane) to show the 10 μM 1-eicosane (Sigma) internal control. Mass spectra of the peaks at 8.83 min and 8.77 min in the native (+)-CPPS and KSL3 *in vitro* assay chromatogram are shown in E and F, respectively. 100 μg (+)-CPPS and KSL3 enzyme, 25 μM GGPP and a 1 hr incubation at 30°C was used to perform each *in vitro* assay. Assays were semi-purified by pentane extract and analyzed on an Agilent DB-5ms column using a 15.5 min oven program.

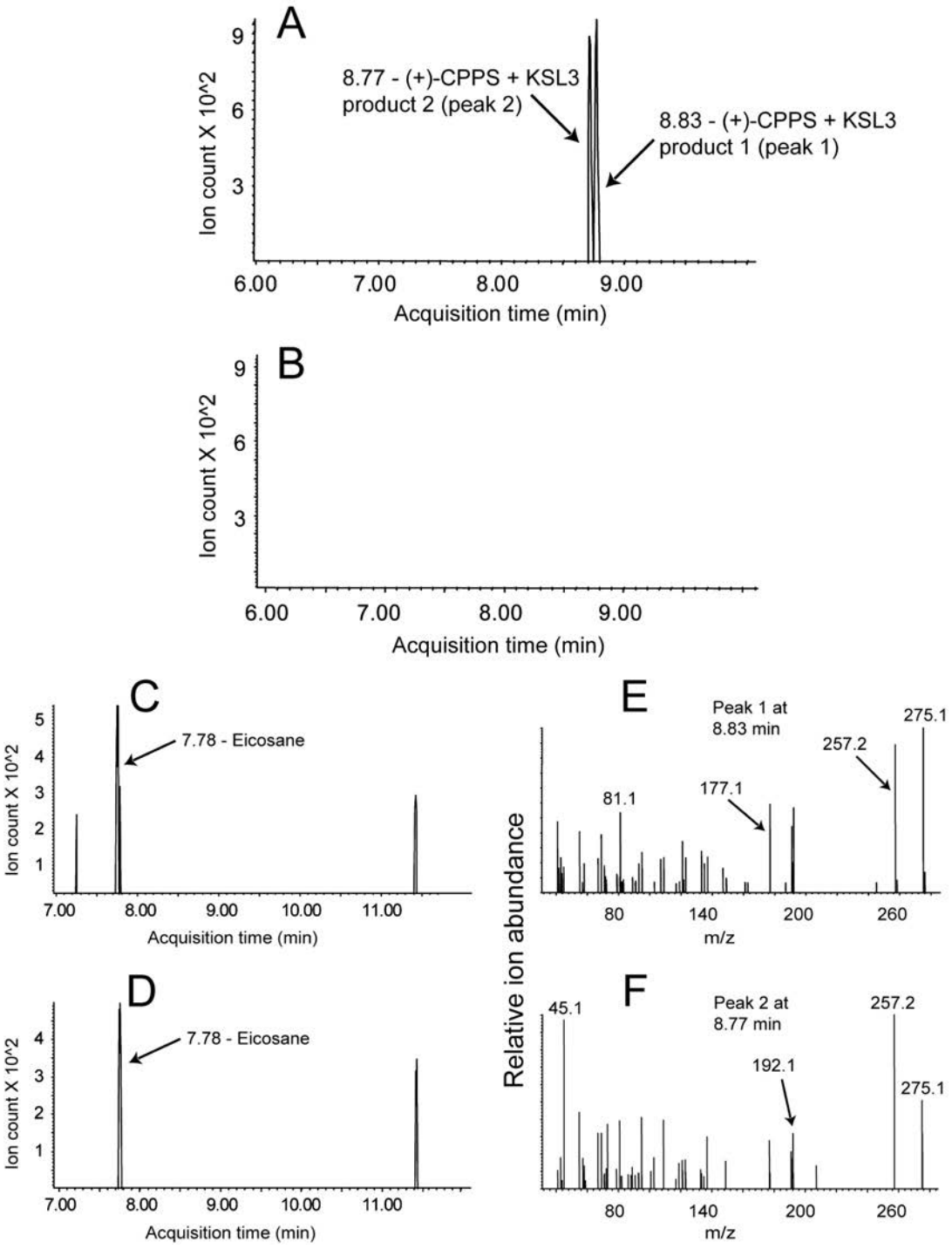
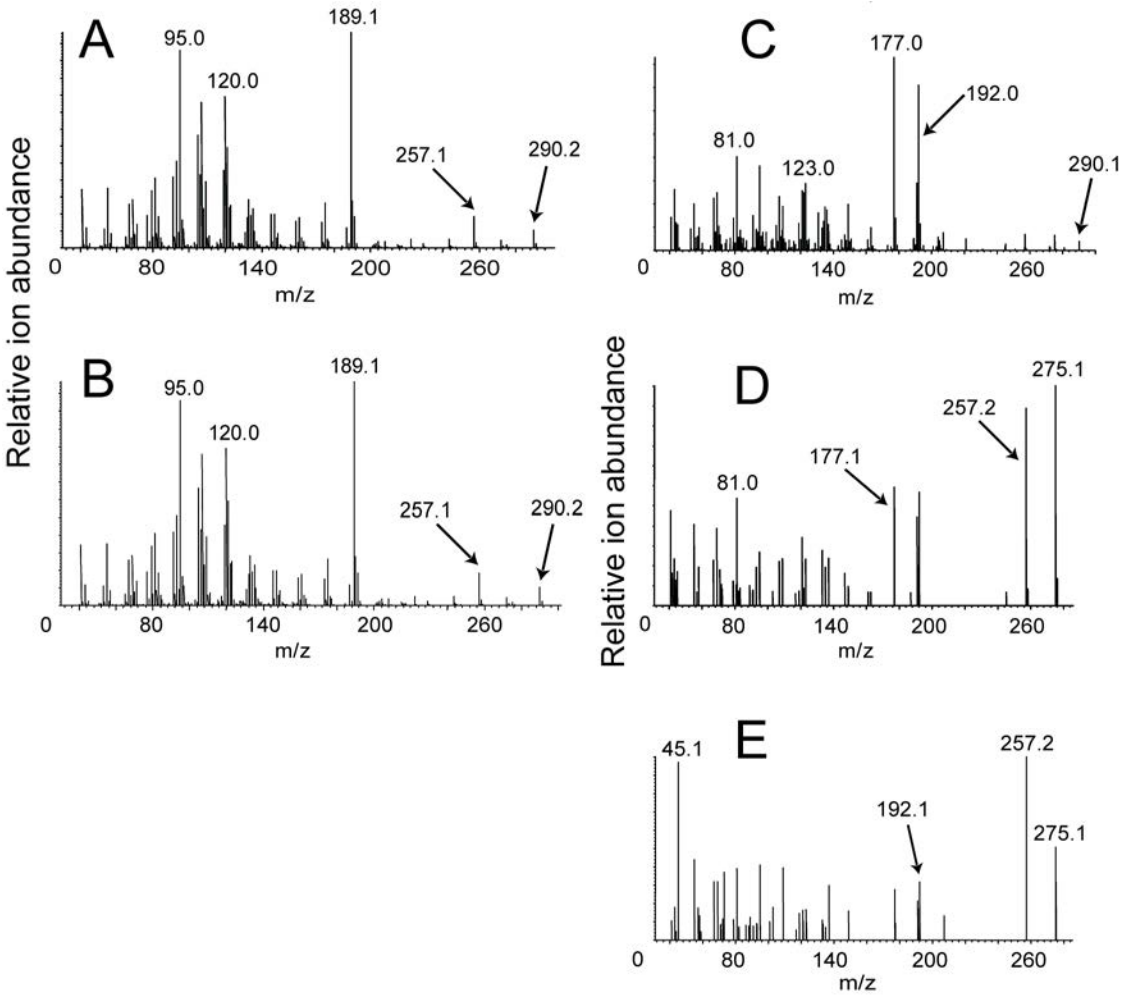


Figure 32. GC-MS analysis of (+)-CPPS and KSL3 co-incubation *in vitro* assays

Figure 33. GC-MS mass spectra of KSL3 enzymatic substrates and products

Mass spectra of the CPPSL2 enzymatic product (A), KSL3 enzymatic conversion of the CPPSL2 product (B), (+)-CPP (C), and KSL3 enzymatic conversions of (+)-CPP (D and E). Analysis was performed on an Agilent DB-5ms column using a 15.5 min oven program.



3.10 NMR analysis of CPPSL2 product

One key question is if the structure of the CPPSL2 product fits to the salvinorin A skeleton. In order to elucidate the structure of CPPSL2 product, a

large scale *in vitro* enzyme assay was performed using 6 mg of recombinant CPPSL2 enzyme and 7 mg GGPP substrate. From this reaction, approximately 500 µg of the CPPSL2 product was synthesized and purified through a silica column. The purified compound was then subjected to NMR analysis. The purified diterpenoid appeared to be >95% pure by GC-MS assessment, but the NMR data showed the presence of a contaminant which interfered with precise interpretation.

Based on the GC-MS data and hypothetical salvinorin A biosynthetic pathway, it was conjectured that kolavenol is the dephosphorylated form of the CPPSL2 product (Figure 27). ^{13}C-NMR chemical shifts for kolavenol are available in published works [93, 94], and thus the published kolevanol ^{13}C-NMR chemical shifts were compared with the ^{13}C-NMR signals from the CPPSL2 product. A total of thirty four ^{13}C NMR signals were detected from the purified CPPSL2 product. Among them, twenty ^{13}C-NMR signals closely matched the kolavenol chemical shifts from the literature (Table 12). There are however subtle variations in ^{13}C-NMR signals. These signal differences are usually <1 part per million (ppm), and are likely due to the different types of NMR solvents used ($CDCl_3$ for references versus C_6D_6 for the sample). Different solvents can cause small changes in the chemical shift of a metabolite. In contrast, one characteristic ^{13}C signal for kolavelool, 111.8 ppm [98], could not be detected from the sample. Overall, this result suggested that the CPPSL2 product is likely to be kolavenol, although additional work is required for unambiguous structural determination.

Table 12. ^{13}C NMR chemical shifts (ppm) from CPPSL2 product and two kolavenol references

Carbon number	^{13}C δ reference[1] ($CDCl_3$ solvent)	^{13}C δ reference[2] ($CDCl_3$ solvent)	^{13}C δ sample[3] (C_6D_6 solvent)
1	18.4	18.3	19.0
2	27.0	26.9	27.6
3	120.5	120.4	121.3
4	144.6	144.5	144.5
5	38.3	38.2	38.8
6	36.9	36.9	37.5
7	27.6	27.6	28.2
8	36.4	36.3	36.9
9	38.7	38.6	39.2
10	46.5	46.5	47.1
11	36.6	36.8	37.5
12	33.0	32.9	33.5
13	140.9	140.9	142.4
14	123.1	122.9	124.9
15	59.5	59.5	59.8
16	16.5	16.5	16.8

Carbon number	^{13}C δ reference[1] ($CDCl_3$ solvent)	^{13}C δ reference[2] ($CDCl_3$ solvent)	^{13}C δ sample[3] (C_6D_6 solvent)
17	16.1	16.0	16.5
18	18.5	18.4	19.0
19	20.0	20.0	20.5
20	18.1	18.0	18.7

[1]Monti H. *et al.* (1999) [93]; [2]Hubert T. D. *et al.* (1985) [94]; [3]~500 μg of diterpene diphosphate was enzymatically synthesized by recombinant pseudomature CPPSL2, dephosphorylated, and extracted by pentane for ^{13}C-NMR analysis. $CDCl_3$ was used as NMR solvent in the two references, whereas C_6D_6 was used for CPPSL2 product.

3.11 Yeast *in vivo* expression of candidate diterpene synthases

The CPPSL2 product generated by large scale *in vitro* assay was unfortunately rapidly degraded in solution, and less than 50 μg remained after several weeks of storage. It was determined that additional amounts of the compound would have to be produced, as 50 μg was insufficient for further NMR analysis. GGPP substrate is costly, and a yeast *de novo* system of GGPP synthesis was therefore developed. Yeast natively produces only small amounts of GGPP, and metabolic flux engineering was used to increase yeast GGPP production. This metabolic engineering strategy involved a “GGPP helper” plasmid which encodes two isoprenoid pathway enzymes. These are *S. cerevisiae* GGPP

synthase (*BTS1*, U31632.1) and a truncated HMG-CoA reductase. HMG-CoA reductase catalyzes a rate-limiting step in the MVA pathway, and hence the overexpression of this gene increases production of IPP and DMAPP [85].

In order to evaluate the efficacy of the helper plasmid, a plasmid encoding levopimaradiene synthase (*LVPS*) (supplied by Dr. Kristala Prather, Massachusetts Institute of Technology [99]) was co-transformed with the helper plasmid, and synthesis of levopimaradiene (LVP) by the transgenic yeast was measured by GC-MS analysis. LVPS is a bi-functional (catalyzing both type II and type I reactions) diterpene synthase that converts GGPP directly into LVP. No production of LVP was observed in yeast cultures expressing the pseudomature *LVPS* without the helper plasmid, but a significant LVP peak was observed when the helper plasmid was co-expressed with *LVPS* (Figure 34C and D). GGOH was also present in each culture extraction, as endogenous yeast phosphatases de-phosphorylate a certain percentage of the GGPP the yeast produces. This amount of GGOH was significantly increased in the helper plasmid co-expressing culture (Figure 34A). These results strongly suggested that expression of the “GGPP helper” plasmid in yeast significantly increased GGPP production.

Figure 34. GC-MS analysis of *LVPS* and "GGPP helper" plasmid *in vivo* yeast expression assays

Chromatograms of *LVPS* and "GGPP helper" plasmid co-expression *in vivo* assays (A and C) and *LVPS* only *in vivo* assays (B and D). C and D chromatograms were extracted at a *m/z* of 272 to show the LVPS enzymatic product. Mass spectra of the peak at 17.37 min in the *LVPS* and helper plasmid co-expression *in vivo* assay chromatogram is shown in E. 1.8% galactose was used to induce expression of *LVPS* and the helper plasmid in *S. cerevisiae* cultures. Cultures were induced for 2 days at 30°C, and semi-purified by hexane extract. Extractions were analyzed on an Agilent DB-5ms column using a 27 min oven program.

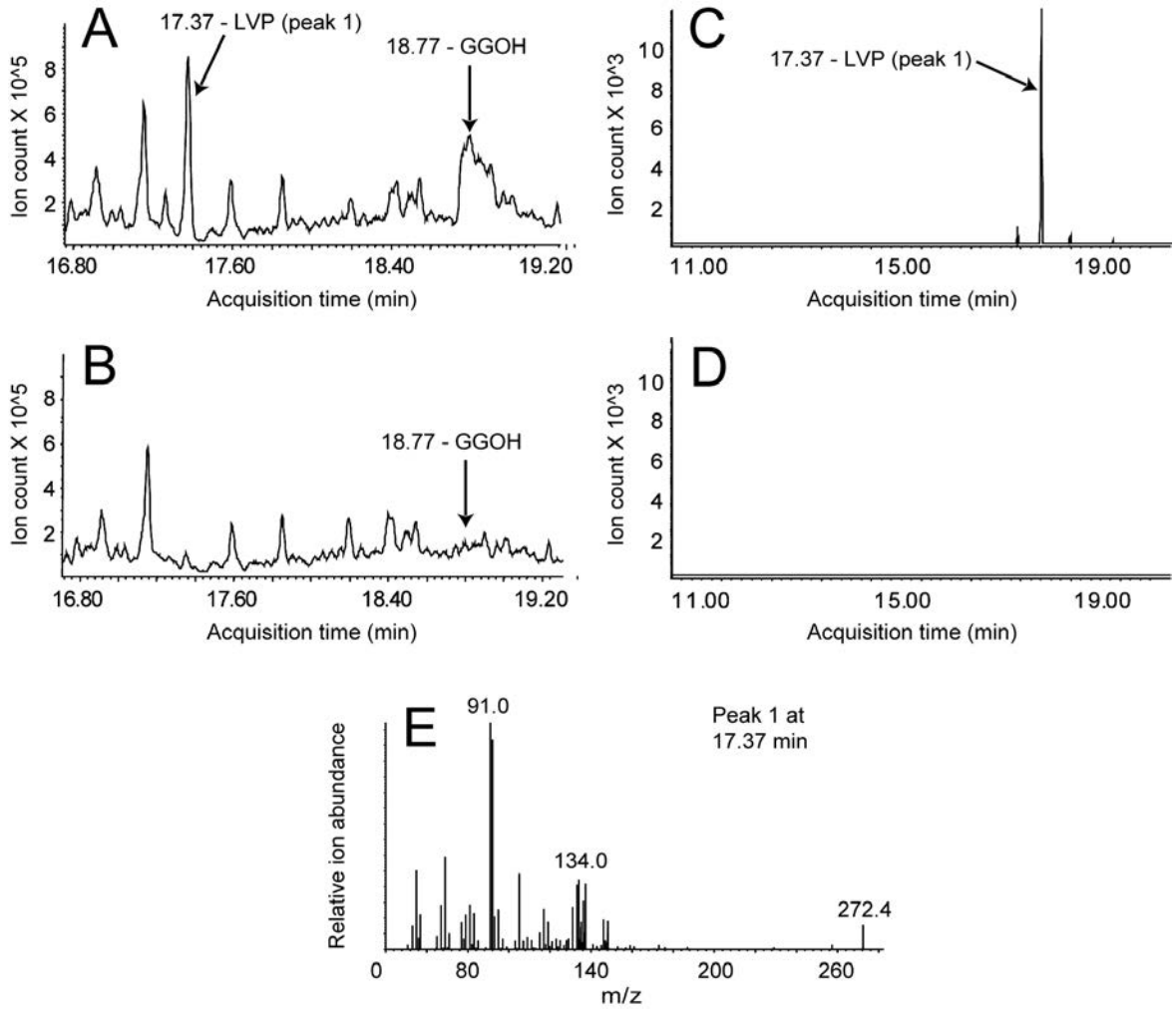


Figure 34. GC-MS analysis of *LVPS* and "GGPP helper" plasmid *in vivo* yeast expression assays

After confirming the usefulness of the helper plasmid in over-producing GGPP in yeast, three constructs were generated to co-express *CPPSL2* with *KSL1, KSL2,* or *KSL3* (Figure 35). FLAG and cMYC epitope tags were added to the C-terminus of CPPSL2 and KSL1/2/3, respectively, after removal of their putative transit peptides (FC predictions in Figure 20). Unfortunately, the initial

GC-MS analysis showed no change in metabolite profile in any of the *in vivo* assays compared to an empty vector control. Accordingly, expression of the genes was evaluated by anti-FLAG and anti-cMyc antibodies. In the immonoblot analysis using anti-cMYC antibody, recombinant KSL1, KSL3, and BTS1 enzymes were clearly detected (Figure 35B). However, no CPPSL2 protein was detected in the immunoblot using anti-FLAG antibody (Figure 35A), and no KSL2 enzyme was detected in the anti-cMYC immunoblot.

Figure 35. Assay of *CPPSL2, KSL1/2/3,* and *BTS1 in vivo* yeast expression by anti-c-MYC and anti-FLAG immunoblot

Upper section = Partial map of pESC-Leu vectors encoding pseudomature (FC prediction in Figure 20) *CPPSL2* and *KSL1/2/3*, and the "GGPP helper" plasmid encoding the *S. cerevisiae* GGPP synthase *BTS1* (U31632.1) and a truncated HMG-CoA reductase. P GAL1 = Galactose-inducible promoter 1, P GAL10 = Galactose-inducible promoter 10.

Lower section = Anti-FLAG epitope tag immunoblots (A), and anti-cMYC epitope tag immunoblots (B) for *CPPSL2* and *KSL1/2/3* co-expression *in vivo* assays. *1 = Protein sample known to contain a C-terminus FLAG epitope tag. 1.8% galactose was used to induce expression of *CPPSL2, KSL1/2/3,* and the "GGPP helper" plasmid in *S. cerevisiae* cultures. Cultures were induced for 2 days at 30°C, and protein was extracted by bead beater abrasion. Epitope tags were bound by mouse monoclonal anti-FLAG or anti-cMYC primary antibodies (Sigma), and primary antibodes were bound by sheep anti-mouse secondary antibodies (linked to horseradish peroxidase) (Sigma). The ECL plus acridan kit (GE Healthcare) was used to visualize bound secondary antibodies.

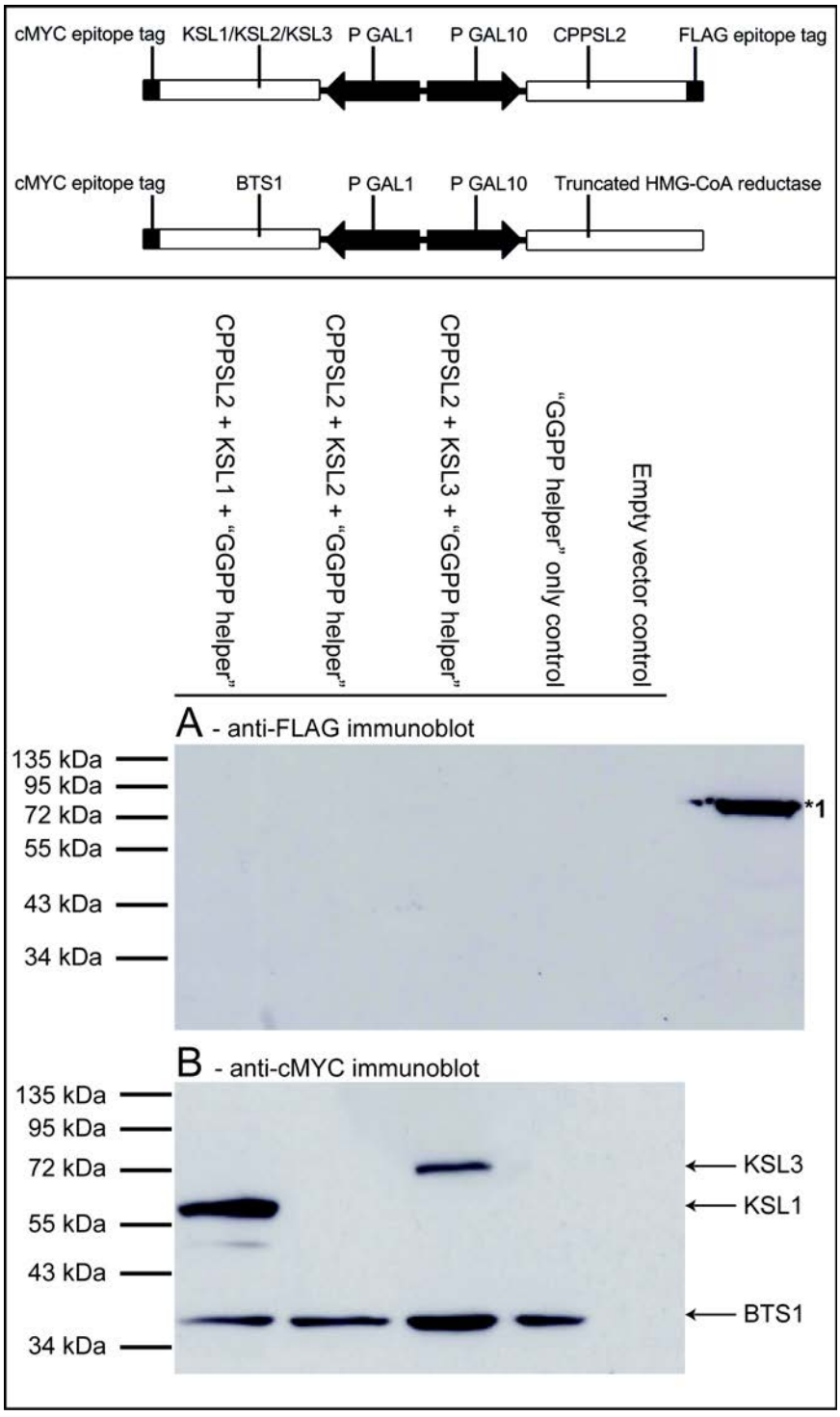


Figure 35. Assay of *CPPSL2, KSL1/2/3,* and *BTS1 in vivo* yeast expression by anti-cMYC and anti-FLAG immunoblot

In summary, a yeast *in vivo* expression system capable of producing an increased level of GGPP was developed in an effort to further characterize *S. divinorum* candidate diterpene synthases, and to economically produce novel diterpenoids for NMR analysis. As judged from the immunoblot analysis, *CPPSL2* was not successfully expressed in this yeast system, and no new diterpenoids were detected in *in vivo* assays as a result. Expression optimization of *CPPSL2* is required to utilize the yeast system.

CHAPTER FOUR: DISCUSSION

4.1 Genomics of non-model medicinal plants

The diverse range of plant secondary metabolites that are found in nature represent an excellent resource for discovery of novel drugs and commercial compounds. The elucidation of genes required for biosynthesis of these secondary metabolites is an important area of research, as it facilitates both the characterization of enzymatic pathways, and the heterologous large scale production of useful metabolites. Plants that produce secondary metabolites of interest are however commonly non-model organisms that lack significant sequencing data. Gene discovery in non-model plants is traditionally difficult, as sequence data for these organisms was virtually non-existent prior to the advent of NGS technologies such as 454 and Illumina sequencing. Conventional Sanger sequencing involves subcloning of DNA strands into vectors and amplification in microbial hosts, and large scale Sanger sequencing efforts are labour and cost intensive as a result. Nucleotide sequences of previously characterized enzymes can be used to design degenerate PCR primers for isolation of novel genes in the same family, but this approach is also limited in its scope. Novel genes for secondary metabolism can sometimes have low sequence homology to characterized enzymes, as was observed in the *S. divinorum KSL3* diterpene synthase. NGS technologies allow the qualitative and quantitative analysis of entire transcriptomes at a reasonable cost, and therefore bypass these gene

elucidation obstacles.

NGS and Sanger sequencing instruments use detection of the fluorescence or luminescence created by nucleotide incorporation to determine the base composition of a DNA strand. NGS techniques however have capabilities that are far beyond that of Sanger sequencing, as they involve the sequencing of several hundred thousand templates in parallel. 454 pyrosequencing, for example, can be used to sequence 20 million base pairs in only 4.5 hr, at a fraction of the cost of Sanger sequencing. NGS thus enables sequence-based biological projects that would not be possible through use of Sanger sequencing alone. The 454 pyrosequencing of *S. divinorum* cDNA preparations was used in this project to generate significant coverage of the secondary metabolism transcriptome, and to facilitate the identification of three novel genes that are believed to act in the salvinorin A secondary metabolite pathway. More than 35 million base pairs of *S. divinorum* sequencing data was economically generated through NGS technology, creating a sequence resource that was used to effectively drive the study of the salvinorin A biosynthetic pathway.

Normalization of cDNA populations is an additional tool that can be used to support identification of genes for secondary metabolism. Qualitative and quantitative analysis of the *S. divinorum* cDNA libraries revealed that the normalization procedure significantly reduced repeated sequencings of transcripts that are highly abundant in the plant. This resulted in an increase in the

transciptome coverage of the normalized cDNA library, as evidenced by the more than 20% of normalized library unigenes (versus 7.1% in the standard library) that lack NCBI BLAST hits and therefore possibly represent novel genes. The normalization procedure also significantly increased the relative abundance of scarce *S. divinorum* transcripts, such as those involved in secondary metabolism. Representation of four of the five *S. divinorum* candidate diterpene synthases was increased in the normalized library, with only *KSL1* having a higher number of unigene members in the standard library. Coverage of isoprenoid biosynthetic pathways in the normalized and standard cDNA libraries was complementary in some ways, indicating that the creation of both a standard and normalized cDNA library can be a worthwhile pursuit. NGS and the techniques that support it like cDNA normalization are improving to the point where any organism can become a "model" organism. These methods will therefore open the door to the discovery and characterization of vast numbers of new plant secondary metabolism genes. The *S. divinorum* sequencing resource that was generated by use of NGS and cDNA normalization will additionally enable future studies into metabolic and developmental processes in the plant.

4.2 CPPSL2 catalyzes the synthesis of terpentedienyl diphosphate

Various analytical techniques, including GC-MS, LC-MS, and NMR, were used in an attempt to elucidate the structure of the de-phosphorylated form of the diterpenoid synthesized by CPPSL2. However, due to low abundance and contamination of the sample, unambiguous structural elucidation of the CPPSL2

product could not be achieved. Nevertheless, multiple lines of direct and indirect evidence suggest that the de-phosphorylated compound is highly likely to be kolavenol, and hence the direct enzymatic product from CPPSL2 is TPP (a neoclerodane diterpene) (Figure 27). First, the electron ionization-mass spectrum (EI-MS) of the de-phosphorylated compound showed a >95% match to that from kolavelool in the mass spectrometry library [95] (Figure 25). Kolavenol mass spectral data is not deposited in the database, and hence the compound could not be directly compared to the metabolite. However, we can conclude from the EI-MS data that the carbon skeleton of the CPPSL2 product must be very closely related to the kolavelool backbone. Kolavenol and kolavelool contain different tail structures from C13 to C16 (i.e., double-bond and oxygen-linking region-specificities), but their core skeletons are identical (Figure 36). Therefore, the EI-MS of kolevanol is expected to be almost identical to kolavelool. Second, the location of salvinorin A's furan ring precludes kolavelool from being the precursor of the metabolite. The oxygen atom in kolavelool is attached to C13, whereas the oxygen in salvinorin A's furan ring is attached to C15 and C16 (Figure 36). Kolavenol is a better fit for the salvinorin A biosynthetic pathway, as the oxygen atom in the compound is attached to C15. In principle, the furan ring can be formed by the oxidation and cyclization reactions shown in Figure 37 [100]. In this reaction scheme, C15 and C16 are further oxidized, possibly by cytochrome P450s, yielding a reactive ketone and an alcohol moiety. Subsequently, a 5-ring hemiacetal structure is formed, followed by double-bond migrations and dehydration. The final furan ring structure will be identical to the one found in

salvinorin A. Third, ^{13}C-NMR analysis identified all twenty δ ^{13}C-signals which matched well with the published data, although fourteen additional δ ^{13}C-signals were detected and ^{1}H-signals could not be clearly resolved due to unknown contaminants. Unfortunately, the reference NMR spectra were measured in $CDCl_3$ solvent while that of the CPPSL2 product was measured in C_6D_6 solvent, causing minor variations (usually <1 ppm) in the ^{13}C chemical shifts.

Further experiments are necessary to draw a conclusion regarding the chemical identity of the CPPSL2 product. Possible experimental approaches are to re-purify the compound from the *in vitro* enzyme assays, to improve and develop an *in vivo* yeast system for *de novo* synthesis of the product, or to acquire a standard from a kolavenol accumulating plant, such as *Aristolochis galeata* [101]. The data presented here however strongly suggests that the type I diterpene synthase reaction proposed in Figure 9 is correct, namely that TPPS renders GGPP into TPP in the first committed step of the salvinorin A biosynthetic pathway.

Figure 36. Oxygen geometry of kolavenol and kolavelool

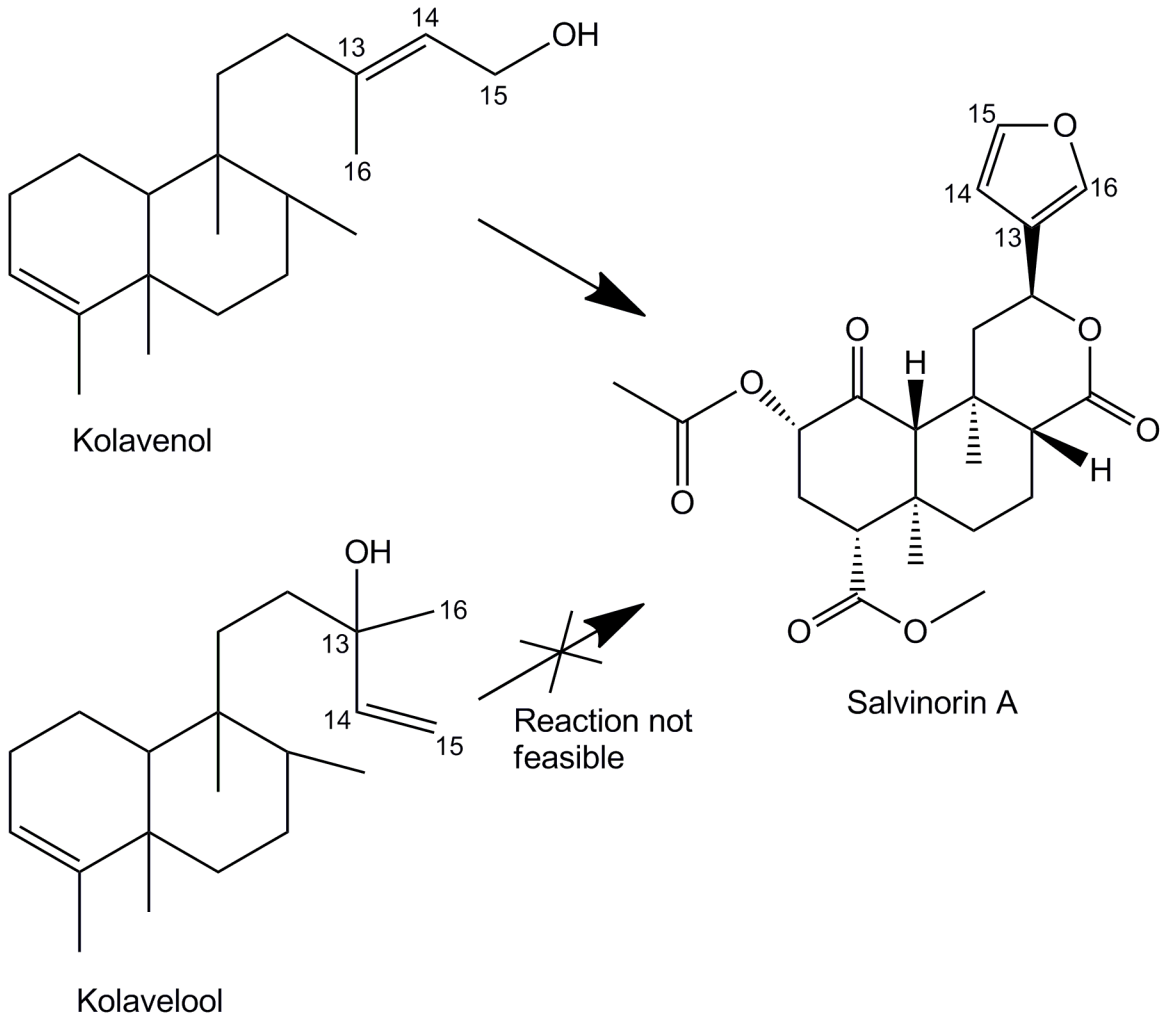
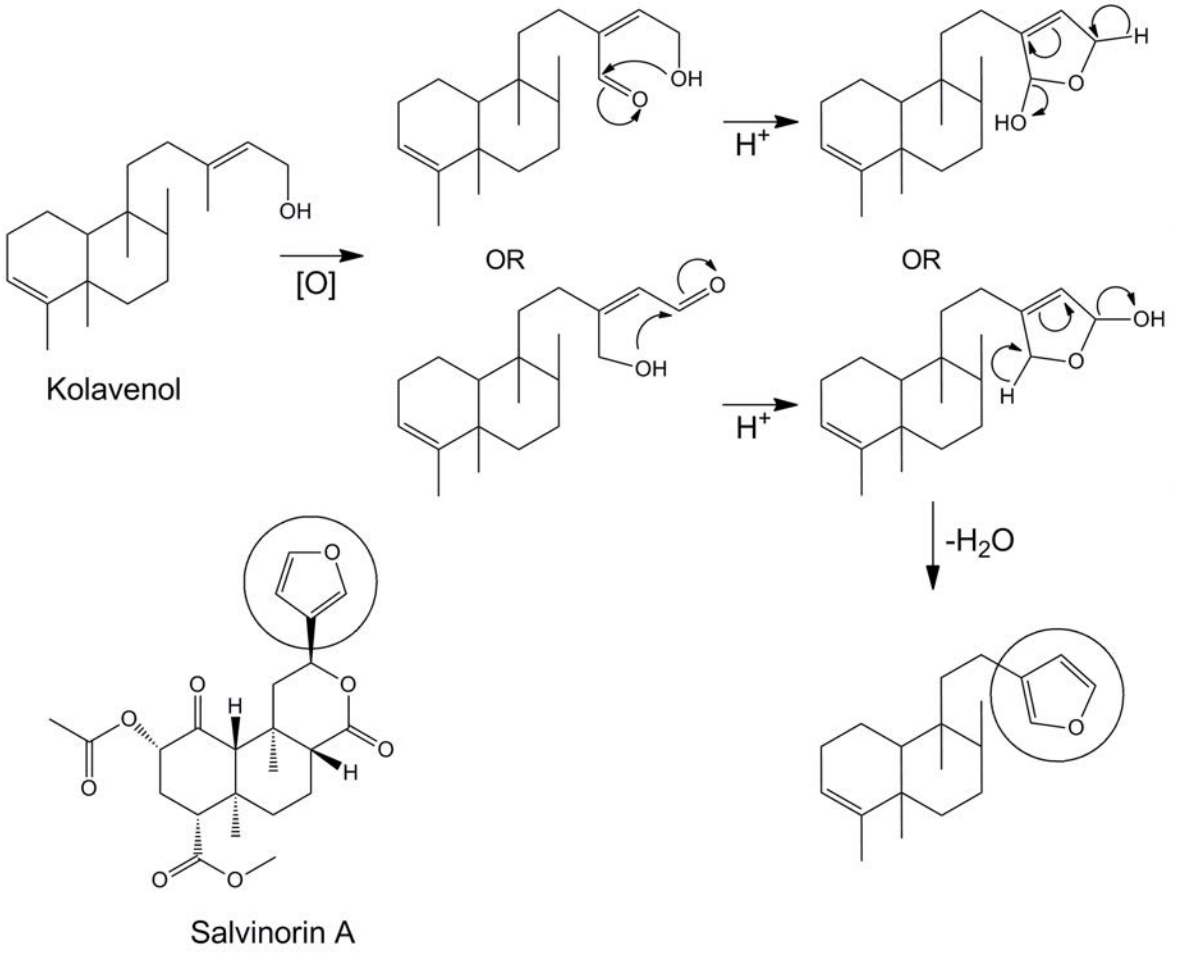


Figure 37. Potential mechanisms of salvinorin A furan ring formation



4.3 Biochemical functions of KSL2 and KSL3

Although CPPSL2 most likely acts to render GGPP into TPP, it is not clear if KSL2 or KSL3 is the enzyme which subsequently utilizes the TPP to produce an intermediate in the salvinorin A biosynthetic pathway. None of the seven N-terminus deleted recombinant KSL2 enzymes produced in *E. coli* displayed the ability to use TPP as a substrate. It is difficult to conclude whether this lack of activity was a result of artifacts caused by N-terminus deletion or the unsuitability of

TPP as a substrate for the enzyme. On the other hand, one of the seven pseudomature KSL3's (+8 chloroplast transit peptide prediction) was successfully expressed and purified, and was putatively found to be able to render TPP into kolavenol. Additionally, (+)-CPP was rearranged into two new diterpenes by the recombinant KSL3. However, conversion of TPP into kolavenol and (+)-CPP into the two novel diterpenes by KSL3 was very slow, as only small amounts of the diterpenes were synthesized in *in vitro* assays. Weak and broad substrate utilization is a typical feature of promiscuous enzyme activity, and it is possible that the KSL3 activities observed here are non-specific catalytic activities. Some type I diterpene synthases however also demonstrate promiscuity in ligand binding. Taxadiene synthase, for example, binds GGPP and 13-aza-13,14-dihydrocopalyl diphosphate [106]. As a result, it is difficult to conclude KSL3's *bona fide* biochemical role in *S. divinorum.*

An incorrect chloroplast transit peptide prediction could be the cause of the KSL2's lack of activities and the weak KSL3 activities. Future efforts to characterize the enzymes should therefore focus on *in planta* expression. For example, *KSL2* and *KSL3* could be expressed in model plants such as *A. thaliana* and *Nicotiana tabacum*. The advantage of a model plant enzyme production system is that no N-terminus deletions are necessary, as chloroplast transit peptides will be cleaved when the enzymes enter the organelle. The altering of enzyme function by incorrect removal of the chloroplast transit peptide is therefore not a concern in a these expression systems. RNAi-mediated silencing of *KSL2* or

KSL3 in *S. divinorum* could also be used to assay the function of the enzymes. A significant decrease in the salvinorin A content of *S. divinorum* peltate glandular trichomes would be observed subsequent to RNAi-silencing if KSL2 or KSL3 are involved in the salvinorin A biosynthetic pathway. Further characterization of KSL2 and KSL3 will shed light on their role in *S. divinorum*, and add to our understanding of the rules and trends that govern the diterpene synthase enzyme family.

4.4 Development of a yeast *in vivo* expression system

Expression of candidate diterpene synthases *in vivo* in *S. cerevisiae* was also attempted, as this expression system has several advantages. Aggregation of protein into inclusion bodies does not occur in yeast, as the organism is a eukaryote and thus able to perform quality control of unfolded proteins. *In vivo* expression can also be used to economically generate large amounts of an isoprenoid of interest, whereas metabolite production by *in vitro* assay requires the purchase of costly substrates [85]. More than 500 µg of a compound is commonly required for NMR analysis, and structure elucidation is therefore significantly facilitated by the development of an *in vivo* production system. Yeast IPP and DMAPP biosynthetic pathways are tightly regulated, and minimal amounts of these isoprenoid pathway precursors accumulate in native yeast cells. For this reason, a method of increasing GGPP production in yeast by “GGPP helper” plasmid was developed. Overexpression of a truncated HMG-CoA reductase and GGPP synthase was demonstrated to significantly increase biosynthesis of diterpenoid pathway precursors, and therefore enhance flux into heterologous diterpenoid

pathways. There is a large body of work that focuses on increasing the *in vivo* production of sesquiterpenes and tetraterpenes, and it is probable that these methods can be used to further increase GGPP production in this yeast *in vivo* system. Notable techniques include the expression of a mutant transcription factor [85] and the engineering of global transcription machinery [102].

Initial efforts to produce salvinorin A pathway intermediates in "GGPP helper" yeast were unfortunately not successful. Protein immunoblots revealed that pseudomature recombinant KSL1 and KSL3 protein was produced by the *in vivo* system, but the activity of these enzymes could not be assayed as no CPPSL2 protein was synthesized by the transgenic yeast. Expression of only the *KSL*s will not generate novel metabolites, as these enzymes use the products of type II diterpene synthases like the CPPSLs as substrates. Analysis of *CPPSL2*'s amino acid composition showed that the enzyme contains a large number of codons that are not well represented in *S. cerevisiae's* genomic tRNA pool. Codon optimization of *CPPSL2* for expression in *S. cerevisiae* is therefore likely necessary for protein production. The partial reconstitution of the salvinorin A biosynthetic pathway in GGPP overproduction yeast will be possible once *CPPSL2* expression is optimized, and this advancement will significantly reduce the complexity of salvinorin A chemical synthesis and the cost of structure elucidation for novel metabolites in the pathway. The "GGPP helper" plasmid that was characterized here can possibly also be used to increase production of other medically and commercially important diterpenoids.

4.5 Significance and future directions

Metabolites that act at all three major opioid receptors (KOR, DOR and MOR) can be generated through modification of a salvinorin A scaffold [60]. These novel opioid receptor ligands can potentially be used to treat a wide range of human disorders, including fever, diarrhea, addiction, depression, and schizophrenia [30]. Production of salvinorin A pathway intermediates by partial reconstitution of the salvinorin A pathway in a microbial system will facilitate research into these novel drugs. Pathway intermediates generated by *in vivo* expression of *CPPSL2, KSL2,* and *KSL3* could potentially be used in salvinorin A semi-synthesis efforts. Functional group decoration could also be applied to the diterpene skeletons these enzymes produce, in an effort to create novel opioid receptor agonists and antagonists.

Directed mutation of *CPPSL2, KSL2,* and *KSL3* could likely also be used to generate novel KOR-specific ligands that are not present in nature. Aspects of salvinorin A's ring structure are integral to opioid receptor affinity [60], and the novel diterpene skeletons these mutated enzymes would produce would potentially have a wide range of activity at the KOR, DOR and MOR. Terpene synthase mutations can result in the generation of entirely new metabolic products, sometimes as the result of a single residue switch [103, 104]. The ability of KSL3 to use (+)-CPP as a substrate suggests that the enzyme is catalytically promiscuous, and combinatorial expression of *KSL3* and type II diterpene synthases could possibly also be used to produce novel diterpenoids. Directed mutation has additionally been shown to

result in increased specificity and activity levels in terpene synthases [105]. Application of these mutation techniques to KSL2 and KSL3 could potentially be used to increase their activity, enabling the characterization of the enzymes. An increase in CPPSL2 enzyme activity would also facilitate use of the enzyme in industrial efforts to synthesize salvinorin A. The identification and characterization of these novel *S. divinorum* diterpene synthases has opened up many opportunities for the development of novel drugs, and for the study of the chemical and biological processes that shape this class of enzyme.

Bibliography

1. Bentley R. *Secondary metabolite biosynthesis: the first century.* Crit Rev Biotech, 1999. **19**: p. 1-40.
2. Paech K. *Biochemie und Physiologie der Sekundären*, in *Pflanzenstoffe* 1950. Springer-Verlag: Berlin.
3. Folkers K, Shunk CH, Linn BO, Robinson FM, Wittreich PE, Huff JW, Gilfillan JL, and Skeggs HR. *Discovery and elucidation of mevalonic acid*, in *Biosynthesis of Terpenes and Sterols* 1959. Wolstenholme GE and O'Connor M, Editors. Little, Brown and Company: Boston, MA. p. 20-45.
4. Tamura G. *Hiochic acid, a new growth factor for Lactobacillus homohiochi and Lactobacillus heterohiochi.* J Gen Appld Microbiol (Tokyo), 1956. **2**: p. 431-434.
5. Staunton J, and Weissman KJ. *Polyketide biosynthesis: a millennium review.* Nat Prod Rep, 2001. **18**: p. 380-416.
6. Vogt T. *Phenylpropanoid Biosynthesis.* Mol Plant, 2009. **3**: p. 2-20.
7. Winterstein E, and Trier G. *Die Alkaloide,* in Eine Monographie der natürlichen Basen, 1910. Gebrüder Borntraeger: Berlin.
8. Robinson R. *A theory of the mechanism of the phytochemical synthesis of certain alkaloids.* J Chem Soc, 1917. **111**: p. 876–899.
9. Leete E. *Alkaloid biogenesis,* in Biogenesis of Natural Compounds, 1963. Bernfield P, Editor. Pergamon Press: New York, NY. p. 739–796.

10. Novak BH, Hudlicky T, Mulzer J, Trauner D, and Reed JW. *Morphine Synthesis and Biosynthesis - An Update.* Curr Org Chem, 2000. **4**: p. 343-362.

11. Leete E. *Recent developments in the biosynthesis of the tropane alkaloids.* Planta Med, 1990. **56**: p. 339-352.

12. Uefuji H, Ogita S, Yamaguchi Y, Koizumi N, and Sano H. *Molecular cloning and functional characterization of three distinct N-methyltransferases involved in the caffeine biosynthetic pathway in coffee plants.* Plant Physiol, 2003. **132**: p. 372-380.

13. Brückmann M, Trigo JR, Foglio MA, and Hartmann T. *Storage and metabolism of radioactively labeled pyrrolizidine alkaloids by butterflies and larvae of Mechanitis polymnia (Lepidoptera: Nymphalidae, Ithomiinae)* Chemoecology, 2000. **10**: p. 25-32.

14. Connolly JD. *Dictionary of terpenoids*. Vol. 1, 1991. Chapman & Hall: New York, NY.

15. De Pasquale R, Circosta C, Occhiuto F, De Stefano S, and De Rosa S. *Central nervous system activity of terpenoids from marine sponge.* Pharmacol Res Comm, 1988. **20**: p. 23-26.

16. Kirby J, and Keasling JD. *Biosynthesis of plant isoprenoids: Perspectives for microbial engineering.* Annu Rev Plant Biol, 2009. **60**: p. 335-355.

17. Meckes M, Calzada F, Tortoriello J, Gonzalez JL and Martinez M. *Terpenoids isolated from Psidium guajava hexane extract with depressant activity on central nervous system.* Phytother Res, 1996. **10**: p. 600-603.

18. Dubey VS, Bhalla R, and Luthra R. *An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants.* J Biosci, 2003. **28**: p. 637-646.

19. Kuzuyama T. *Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units.* Biosci Biotechnol Biochem, 2002. **66**: p. 1619-1627.

20. Rodwell VW, Beach MJ, Bischoff M, Bochar DA, Darnay BG, Friesen JA, Gill JF, Hedl M, Jordan-Starck T, Kennelly PJ, Kim DY and Wang Y. *3-Hydroxy-3-methylglutaryl-CoA reductase.* Meth Enzymol, 2000. **324**: p. 259-280.

21. Rohmer M, Knani M, Simonin P, Sutter B, and Sahm H. *Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate.* Biochem J, 1993. **295**: p. 517-524.

22. Lichtenthaler HK, Schwender J, Disch A, and Rohmer M. *Biosynthesis of isoprenoids in higher plant chloroplasts proceeds via a mevalonate-independent pathway.* FEBS Lett, 1997. **400**: p. 271-274.

23. Chappell J. *The genetics and molecular genetics of terpene and sterol origami.* Curr Opin Plant Biol, 2002. **5**: p. 151-157.

24. Bick JA, and Lange BM. *Metabolic cross talk between cytosolic and plastidial pathways of isoprenoid biosynthesis: unidirectional transport of intermediates across the chloroplast envelope membrane.* Arch Biochem Biophys, 2003. **415**: p. 146-154.

25. Laule 0, Furholz A, Chang HS, Zhu T, Wang X, Heifetz PB, Gruissem W and Lange M. *Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in Arabidopsis thaliana.* Proc Natl Acad Sci USA, 2003. **100**: p. 6866-6871.

26. Epling C, and Játiva MC. *A new Species of Salvia from Mexico.* Botanical Museum Leaflets, Harvard University, 1962. **20**: p. 75-76.

27. Valdes LJ, Diaz JL, and Paul AG. *Ethnopharmacology of Ska María Pastora.* J. Ethnopharmacol, 1983. **7**: p. 287-312.

28. Clebsch B, and Barner CD. *The New Book of Salvias.* Vol 1, 2003. Timber Press: Portland, OR.

29. Wasson RG. *Notes on the Present Status of Ololiuhqui and the Other Hallucinogens of Mexico.* Botanical Museum Leaflets, Harvard University, 1963. **20**: p. 161-193.

30. Vortherms TA, and Roth BL. *Salvinorin A: From natural product to human therapeutics.* Mol Interv, 2006. **6**: p. 259-267.

31. Valdes LJ. *The Pharmacognosy of Salvia divinorum (Epling and Játiva-M): An Investigation of Ska María Pastora (Mexico).* J Ethnopharm, 1983. **7**: *p. 287-312.*

32. Valdes LJ, Hatfield GM, Koreeda M and Paul AG. *Studies of Salvia divinorum (Lamiaceae), an Hallucinogenic Mint from the Sierra Mazateca in Oaxaca, Central Mexico.* Eco Bot, 1987. **41**: p. 283-291.

33. DEA Office of Diversion Control. *S. divinorum legal status*, in *Drugs and Chemicals of Concern* 2007, U.S. Drug Enforcement Administration: www.justice.gov/dea/

34. Ortega A, Blount JF, and Manchand PS. *Salvinorin, a new transneoclerodane diterpene from Salvia divinorurn (Labiatae).* J Chem Soc, Perkin Transactions, 1982. **1**: p. 2505-2508.

35. Siebert DJ. *Salvia divinorum and salvinorin A: New pharmacologic findings.* J Ethnopharmacol, 1994. **43**: p. 53-56.

36. Scheerer JR, Lawrence JF, Wang GC, and Evans DA. *Asymmetric synthesis of salvinorin A, a potent K opioid receptor agonist.* J Am Chem Soc, 2007. **129**: p. 8968-8969.

37. Hooker JM, Xu Y, Schiffer W, Shea C, Carter P, and Fowler JS. *Pharmacokinetics of the potent hallucinogen, salvinorin A in primates parallels the rapid onset and short duration of effects in humans.* NeuroImage, 2008. **41**: p. 1044-1050.

38. Gonzalez D, Riba J, Bouso, JC, Gomez-Jarabo G, and Barbanoj, MJ. *Pattern of use and subjective effects of Salvia divinorum among recreational users.* Drug Alcohol Depend 2006. **85**: p. 157-162.

39. Fantegrossi WE, Kugle KM, Valdes LJ 3rd, Koreeda M, and Woods JH *Kappa-opioid receptor-mediated effects of the plant-derived hallucinogen, salvinorin A, on inverted screen performance in themouse.* Behav Pharmacol, 2005. **16**: p. 627-633.

40. Carlezon WA, Beguin C, DiNieri JA, Baumann PL, Richards MR, Todtenkopf MS, Rothman RB, Ma Z, Lee DY, and Cohen BM. *Depressive-like effects of the K-opioid receptor agonist salvinorin A on behavior and neurochemistry in rats.* J Pharmacol Exp Therapeut, 2006. **316**: p. 440-447.

41. Braida D, Limonta V, Capurro V, Fadda P, Rubino T, Mascia P, Zani A, Gori E, Fratta W, Parolaro D, and Sala M. *Involvement of K-opioid and endocannabinoid system on salvinorin A-induced reward.* Biol Psychiatr, 2008. **63**: p. 286-292.

42. Zhang Y, Butelman ER, Schlussman SD, Ho A, and Kreek MJ. *Effects of the plant-derived hallucinogen salvinorin A on basal dopamine levels in the caudate putamen and in a conditioned place aversion assay in mice: agonist actions at kappa opioid receptors.* Psychopharmacology, 2005. **179**: p. 551-558.

43. Braida D, Limonta V, Pegorini S, Zani A, Guerini-Rocco C, Gori E, and Sala M. *Hallucinatory and rewarding effect of salvinorin A in zebrafish: K-opioid and CB1 -cannabinoid receptor involvement.* Psychopharmacology, 2007. **190**: p. 441-448.

44. Roth BL, Baner K, Westkaemper R, Siebert D, Rice KC, Steinberg S, Ernsberger P, and Rothman RB. *Salvinorin A: A potent naturally occurring nonnitrogenous κ opioid selective agonist.* PNAS, 2002. **99**: p. 11934-11939.

45. Glennon RA, Titeler M, and McKenney JD. *Evidence for 5-HT2 involvement in the mechanism of action of hallucinogenic agents.* Life Sci, 1984. **35**: p. 2505-2511.

46. Kutrzeba LM, Karamyan VT, Speth RC, Williamson JS, and Zjawiony JK. *In vitro studies on metabolism of salvinorin A.* Pharmaceut Biol, 2009. **47**: p. 1078-1084.

47. Schmidt MS, Prisinzano TE, Tidgewell K, Harding W, Butelman ER, Kreek MJ, and Murry DJ. *Determination of salvinorin A in body fluids by high performance liquid chromatography-atmospheric pressure chemical ionization.* J Chrom B, 2005. **818**: p. 221-225.

48. Siebert DJ. *Localization of salvinorin A and related compounds in glandular trichomes of the psychoactive sage, Salvia divinorum.* Ann Bot, 2004. **93**: p. 763-771.

49. Aldrich JV, and Vigil-Cruz SC. *Narcotic analgesics,* in Burger's Medicinal Chemistry and Drug Discovery, 2003. John Wiley & Sons: New York. p. 329-481.

50. Brownstein MJ. *A brief history of opiates, opioid peptides, and opioid receptors.* Proc Natl Acad Sci, 1993. **90**: p. 5391-5393.

51. Pfeiffer A, Brantl V, Herz A, and Emrich HM. *Psychotomimesis mediated by kappa opiate receptors.* Sci, 1986. **15**: p. 774-776.

52. Walsh SL, Strain EC, Abreu ME, and Bigelow GE. *Enadoline, a selective kappa opioid agonist: comparison with butorphanol and hydromorphone in humans.* Psychopharmacology, 2001. **157**: p. 151-162.

53. Fahn A. *Structure and function of secretory cells.* Adv Bot Res, 2000. **31**: p. 37-75.

54. Levin D. *The role of trichomes in plant defense* Q Rev Biol, 1973. **48**: p. 3-15.

55. Hallahan DL. *Monoterpenoid biosynthesis in the glandular trichomes of labiate plants.* Adv Bot Res, 2000. **31**: p. 77-120.

56. Croteau R, and Johnson MA. *Biosynthesis of terpenoids in glandular trichomes.* Biology and chemistry of plant trichomes. New York: Plenum Press, 1984: p. 133-185.

57. Gruber JW, Siebert DJ, Der Marderosian AH, and Hock RS. *High performance liquid chromatographic quantification of salvinorin a from tissues of Salvia divinorum epling and játiva-m.* Phytochem Anal, 1999. **10**: p. 22-25.

58. Harding WW, Tidgewell K, Schmidt M, Shah K, Dersch CM, Snyder J, Parrish D, Deschamps JR, Rothman RB, and Prisinzano TE. *Salvinicins A and B, New Neoclerodane Diterpenes from Salvia divinorum.* Org Lett, 2005. **7**: p. 3017-3020.

59. Munro TA, Rizzacasa MA, Roth BL, Toth BA, and Yan F. *Studies toward the pharmacophore of salvinorin A, a potent kappa opioid receptor agonist.* J Med Chem, 2005. **48**: p. 345-348.

60. Prisinzano TE, and Rothman RB. *Salvinorin A analogs as probes in opioid pharmacology.* Chem Rev, 2008. **108**: p. 1732-1743.

61. Beguin C, Richards MR, Li JG, Wang Y, Xu W, Liu-Chen LY, and Carlezon WA Jr. *Cohen BM Synthesis and in vitro evaluation of salvinorin A analogues: Effect of configuration at C2 and substitution at C18.* Bioorg. Med Chem Lett, 2006. **16**: p. 4679-4685.

62. Chavkin C, Sud S, Jin W, Stewert J, Zjawiony JK, Siebert DJ, Toth BA, Hufeisen SJ, and Roth BL. *Salvinorin A, an active component of the hallucinogenic sage Salvia divinorum is a highly efficacious kappa-opioid receptor agonist: structural and functional considerations.* J Pharmacol Exp, 2004. **308**: p. 1197-1203.

63. John TF, French LG, and Erlichman JS. *The antinociceptive effect of Salvinorin A in mice.* Eur J Pharmacol, 2006. **545**: p. 129-133.

64. Ansonoff MA, Zhang J, and Czyzyk T. *Antinociceptive and hypothermic effects of salvinorin A are abolished in a novel strain of {kappa}-opioid receptor-1 knockout mice.* J Pharmacol Exp Ther, 2006. **318**: p. 641-648.

65. Capasso R, Borrelli F, Capasso F, Siebert DJ, Stewart DJ, Zjawiony JK, and Izzo AA. *The hallucinogenic herb Salvia divinorum and its active ingredient salvinorin A inhibit enteric cholinergic transmission in the guinea-pig ileum. .* Neurogastroenterol Motil, 2006. **18**: p. 69-75.

66. Capasso R, Borrelli F, Zjawiony J, Kutrzeba L, Aviello G, Sarnelli G, Capasso F, and Izzo AA. *The hallucinogenic herb Salvia divinorum and its active ingredient salvinorin a reduce inflammation-induced hypermotility in mice.* Neurogastroenterol Motil, 2008. **20**: p. 142-148.

67. Mello NK, and Negus SS *Interactions between kappa opioid agonists and cocaine.* Ann New York Acad Sci, 2000. **909**: p. 104-132.

68. Chartoff EH, Potter D, Damez-Werno D, Cohen BM and Carlezon WA J. *Exposure to the selective kappa-opioid receptor agonist salvinorin A modulates the behavioral and molecular effects of cocaine in rats.* Neuropsychopharmacology, 2008. **33**: p. 2550-2562.

69. Mague SD, Pliakas AM, Todtenkopf MS, Tomasiewicz HC, Zhang Y, Stevens WC Jr, Jones RM, Portoghese PS, and Carlezon WA Jr. *Antidepressant-like effects of kappa-opioid receptor antagonists in the forced swim test in rats.* J Pharmacol Exp Ther, 2003. **305**: p. 323-30.

70. Todtenkopf MS, Marcus JF, Portoghese PS, and Carlezon WA Jr. *Effects of kappa-opioid receptor ligands on intracranial self-stimulation in rats.* Psychopharmacology, 2004. **41**: p. 463-470.

71. Hanes K. *Salvia divinorum: Clinical and Research Potential.* Maps, 2003. **13**.

72. Marchesi GF, Santone G, Cotani P, Giordano A, and Chelli F. *The therapeutic role of naltrexone in negative symptom schizophrenia.* Prog Neuropsychopharmacol Biol Psychiatry, 1995. **19**: p. 1239-1249.

73. Barg J, Belcheva M, Rowinski J, Ho A, Burke WJ, Chung HD, Schmidt CA, and Coscia CJ. *Opioid receptor density changes in Alzheimer amygdala and putamen.* Brain Res, 1993. **632**: p. 209-215.

74. Nichols DE. *Hallucinogens.* Pharmacol Therapeut, 2004. **101**: p. 131-181.

75. Mowry M, Mosher M, and Briner W. *Acute physiologic and chronic histologic changes in rats and mice exposed to the unique hallucinogen salvinorin A.* J Psychoactive Drugs, 2003. **35**: p. 379-382.

76. Kutrzeba L, Dayan FE,Howell JL, Feng J, Giner JL, and Zjawionyad JK. *Biosynthesis of salvinorin A proceeds via the deoxyxylulose phosphate pathway.* Phytochemistry, 2007. **68**: p. 1872-1881.

77. Yamaguchi S. *Gibberellin metabolism and its regulation.* Plant Biol, 2008. **59**: p. 225-251.

78. Tholl D. *Terpene synthases and the regulation, diversity and biological roles of terpene metabolism.* Cur Opin Plant Biol, 2006. **9**: p. 297-304.

79. Schmiderer C, Grassi P, Novak J, Weber M, and Franz C. *Diversity of essential oil glands of clary sage.* Plant Biol, 2008. **10**: p. 433-440.

80. Schalk Michel (Collonges-Sous-Saleve, FR), *Method for producing sclareol*, 2011, FIRMENICH SA (CH): United States. http://www.freepatentsonline.com/y2011/0041218.html

81. Hamano Y, Kuzuyama T, Itoh N, Furihata K, Seto H, and Dairi T. *Functional analysis of eubacterial diterpene cyclases responsible for biosynthesis of a diterpene antibiotic, terpentecin.* J Biol Chem, 2002. **277**: p. 37098-104.

82. Lu T, Vargas D, Franzblau SG, and Fischer NH. *Diterpenes from Solidago rugosa.* Phytochemistry, 1995. **38**: p. 451-456.

83. Kawakami S, Harinantenaina L, Matsunami K, Otsuka H, Shinzato T, and Takeda Y. *Macaflavanones A-G, prenylated flavanones from the leaves of Macaranga tanarius.* J Nat Prod, 2008. **71**: p. 1872-1876.

84. Hayashi Y, Toyomasu T, Hirose Y, Onodera Y, Mitsuhashi W, Yamane H, Sassa T, and Dairi T. *Comparison of the Enzymatic Properties of ent-Copalyl Diphosphate Synthases in the Biosynthesis of Phytoalexins and Gibberellins in Rice.* Biosci Biotechnol Biochem, 2008. **72**: p. 523-530.

85. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, Ndungu JM, Ho KA, Eachus RA, Ham TS, Kirby J, Chang MC, Withers ST, Shiba Y, Sarpong R, and Keasling JD. *Production of the antimalarial drug precursor artemisinic acid in engineered yeast.* Nature, 2006. **440**: p. 940-943.

86. Hillwig ML, Xu M, Toyomasu T, Tiernan MS, Wei G, Cui G, Huang L, and Peters RJ *Domain loss has independently occurred multiple times in plant terpene synthase evolution.* Plant J, 2011. **68**: p. 1051-1060.

87. Morgenstern B. *DIALIGN: Multiple DNA and Protein Sequence Alignment at BiBiServ.* Nucleic Acids Res, 2004. **32**: p. W33-W36.

88. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, and Gascuel O. *New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0.* Syst Biol, 2010. **59**: p. 307-321.

89. Cheong JJ, and Choi YD. *Methyl jasmonate as a vital substance in plants.* Trends Genet, 2003. **19**: p. 409-413.

90. Boughton AJ, Hoover K, and Felton GW. *Methyl jasmonate application induces increased densities of glandular trichomes on tomato, Lycopersicon esculentum.* J Chem Ecol, 2005. **31**: p. 2211-2216.

91. Keegstra K, Olsen LJ, and Theg SM. *Chloroplastic Precursors and their Transport Across the Envelope Membranes.* Annu Rev Plant Physiol Plant Mol Biol, 1989. **40**: p. 471-501.

92. Emanuelsson O, Nielsen H, and von Heijne G. *ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites.* Protein Sci, 1999. **8**: p. 978-984.

93. Monti H, Tiliacos N, and Faure R. *Copaïba oil: isolation and characterization of a new diterpenoid with the dinorlabdane skeleton.* Phytochemistry, 1999. **51**: p. 1013-1015.

94. Hubert TD, and Wiemer DF. *Ant-repellent terpenoids from Melampodium divaricatum.* Phytochemistry, 1985. **24**: p. 1197-1198.

95. McLafferty FW. NIST2/Wiley 5 Mass Spectra Library, ed. McLafferty FW2009, Software: Wiley-Interscience.

96. Harris LJ, Saparno A, Johnston A, Prisic S, Xu M, Allard S, Kathiresan A, Ouellet T, and Peters RJ. *The maize An2 gene is induced by Fusarium attack and encodes an ent-copalyl diphosphate synthase.* Plant Mol Biol, 2005. **59**: p. 881-894.

97. Zerbe P, Chiang A, Yuen M, Hamberger Bjorn, Hamberger Britta, Draper JA, Britton R, and Bohlmann J. *Bifunctional cis-Abienol Synthase from Abies balsamea Discovered by Transcriptome Sequencing and Its Implications for Diterpenoid Fragrance Production.* J Biol Chem **287**: p. 12121-12131.

98. Bomm MD, Zukerman-Schpector J, and Lopes LMX. *Rearranged (4→2)-abeo-clerodane and clerodane diterpenes from Aristolochia chamissonis.* Phytochemistry, 1999. **50**: p. 455-461.

99. Leonarda E, Ajikumara PK, Thayerb K, Xiaoa WH, Moa JD, Tidorb B, Stephanopoulosa G, and Prathera KL. *Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control.* PNAS, 2010. **107**: p. 13654-13659.

100. Gaikwad NW, and Madyastha KM. *Biosynthesis of b-Substituted Furan Skeleton in the Lower Furanoterpenoids: A Model Study.* Biochem Biophys Res Comm, 2002. **290**: p. 589-594.

101. Pacheco AP, Oliveira PM, Pilo-Veloso D, and Alcantara AF. *13C-NMR Data of Diterpenes Isolated from Aristolochia Species.* Molecules, 2009. **14**: p. 1245-1262.

102. Alper H, and Stephanopoulos G. *Global transcription machinery engineering: a new approach for improving cellular phenotype.* Metab Eng, 2007. **9**: p. 258-267.

103. Segura MJ, Jackson BE, and Matsuda SP. *Mutagenesis approaches to deduce structure-function relationships in terpene synthases.* Nat Prod Rep, 2003. **20**: p. 314-317.

104. Xu M, Wilderman PR, and Peters RJ. *Following evolution's lead to a single residue switch for diterpene synthase product outcome.* PNAS, 2007. **104**: p. 7397-7401.

105. Yoshikuni Y, Martin VJ, Ferrin TE, and Keasling JD. *Engineering cotton (+)-δ-cadinene synthase to an altered function: germacrene D-4-ol synthase.* Chem Biol, 2006. **13**: p. 91-98.

106. Köksal M, Jin Y, Coates RM, Croteau R, and Christianson DW. *Taxadiene Synthase Structure and Evolution of Modular Architecture in Terpene Biosynthesis.* Nature, 2011. **6**: p. 116-120.

APPENDIX A: NUCLEOTIDE AND PROTEIN SEQUENCES OF *S. divinorum* CANDIDATE DITERPENE SYNTHASES

A.1 *CPPSL1* nucleotide sequence

ATGGCCTCTCTTTCCTCTACAATCCTCAGCCGCCCTCCGGCGACCGGATGTAG
AATTCCGTCGTCCTCGGCTAAGCTTCCCTTGCCGGAATGTTTCGCCACCAATG
CCTGGATGAGCAGCACTAAAAACCTCTCTCTCACTTACCAACTTAATCACAAGA
AAAATTCAAAAGTAGTTGCAGTTGCGACAGTAGATGCTCCACAGGTGCATGAT
CACGACGATTCCACTGTTCTTCAAGGCCATGATGCGGTTAACAATATCGATGAT
CCGATTGAGTACATCAGAACGTTGCTGAAGACGACGGGGGACGGGCGGATAA
GCGTGTCGCCCTACGACACTGCATGGGTGGCGATGATAAAGGACGTGGGGG
GGCGCGACGCCCCAGAGTTTCCCTCCAGCCTGGAGTGGATCGTGCAGAACC
AGCTCCAGGACGGATCTTGGGGCGATCAGAAGCTTTTCTGCGTCTATGATCGC
CTCGTTAACACCATTGCGTGTGTCGTGGCCTTGAGATCGTGGAACGTTCACGA
CGATAAGGTCAAAAGAGGAGTGGCGTACATCAACGAAAACGTCGATAAGCTCA
CGGAGGGAAACGAGGAGCACATGACGTGTGGTTTCGAGGTCGTGTTTCCGG
CCCTTCTTCAAAACGCCAAGAGCATGGGCATCGAAGATCTTCCCTACGATGCG
CCCGCGGTGCACGAGGTTTACCACACCAGAGAGCAAAAGTTGAAAAGGATTC
CACTGGAGATTATGCACAAAATGCCGACATCATTATTATTCAGTTTGGAAGGGTT
AGAAAATTTGGATTGGGACAAGCTTTTGAAACGGCAATCAGCAGATGGTTCCT
TCCTCACTTCTCCCTCCTCCACCGCCTTCCCATTCATGCAAACCAGAGATAAAA
AATGCTACCAATTCATCAAGAACACTGTAGACACCTTTAATGGGGGAGCGCCA
CACACTTATCCTGTCGACGTGTTCGGGAGACTCTGGGCGATCGACCGGCTGC

AGCGCCTCGGGATTTCTCGCTTTTTCGAGACGGAGATTGCTGAATGCTTGAGC

CACATCCACAAGTTTTGGACAGATAAGGGGGTTTTCAGTGGGAGAGATTCTGA

GTTTTGTGATATTGATGATACATCCATGGGAGTGAGGCTTCTCAGAATGCATGG

ATATGATGTTGATCCAAATGTTCTGAGGAATTTCAAGCAGAAAGATGGTAAATTC

TCTTGCTACGGCGGGCAGATGATCGAGTCGCCTTCTCCGATATACAATCTCTAC

AGAGCTTCTCAACTCCGGTTTCCCGGAGAGGAAATCTTGGAAGATGCAGACAA

ATTTGCCTACGATTTCTTGAAAGAAAAATTAGCTAATAATCAGATTCTGGATAAGT

GGGTTATATCTAAGCATTTGCCTGATGAGATAAAGTTGGGGATAGAGATGCCAT

GGCTCGCCACCCTTCCCCGCGTGGAGGCCAAGTACTACATCCAGTACTACGC

CGGCTCCGGCGATGCATGGATCGGAAAGACACTATACAGAATGCCGGAGATCA

GCAACGACACGTATCACGACCTAGCCAAGACGGATTTCAAGAGATGCCAAGC

GCAACATCAGTTTGAATGGATCTACATGCAAGAGTGGTACGAGAGCTGCGGCA

TCGAAGAATTCGGGATAAGCAGAAAGGATCTTCTCCTGTATTACTTCATGGCAG

CTGCAAGCATCTTCGAGCTCGAGAGGACCAACGAGAGACTGGCGTGGGCCA

AATCCCAAGTCATTTCCAAGATGATCACTTGTTTCTTCAACAGCGAGACTAGAT

CGTTCCAGGAGAAACAAGACCTTTTGAATGAGCTCAAAAATATCAATGGCCTC

GACAGCAGAAACAGTGGCGAGAGAGAAGATGGGGCCGCAAGCATGGTTGTA

GCAACCCTCACTCAATTCCTCGGGGGATTCAACAGATACACCAGACACCAGTT

AAAAAACACCTGGAGTGTGTGGCTGACGAAGTTGCAAGATGGCAAAGTAGAC

GACGCAGAGCTCCTAACCAACACGTTGACCATCTGCGCCGGCCACACAAAAG

AGACACTGTCGCACAGCGAATACAAGACTCTCTCCAACCTAACCACCAAAATTT

GTCAGCAGCTTTCTCTGATTCAAACCCAGAAGGAGATGGGAGTAGATGGGCAT

ATCGCAGCGATAAAAAATAAGGAACTCGAACAAGACATGCAGATGCTGGTGAA

GTTGGTGCTCGAGAAATCTGGGGGCATCAACAGAAATATCAAGAAAACGTTTT
TGGCGGTTGCAAAGACTTATTATTACAGAGCATATCATGCTGCCAACACCATTG
ATACCCACATGTTCAAAGTGCTTTTTGAACCCGTTTGA

A.2 *CPPSL1* amino acid sequence

MASLSSTILSRPPATGCRIPSSSAKLPLPECFATNAWMSSTKNLSLTYQLNHKKNS
KVVAVATVDAPQVHDHDDSTVLQGHDAVNNIDDPIEYIRTLLKTTGDGRISVSPYDT
AWVAMIKDVGGRDAPEFPSSLEWIVQNQLQDGSWGDQKLFCVYDRLVNTIACVV
ALRSWNVHDDKVKRGVAYINENVDKLTEGNEEHMTCGFEVVFPALLQNAKSMGIE
DLPYDAPAVHEVYHTREQKLKRIPLEIMHKMPTSLLFSLEGLENLDWDKLLKRQSA
DGSFLTSPSSTAFPFMQTRDKKCYQFIKNTVDTFNGGAPHTYPVDVFGRLWAIDR
LQRLGISRFFETEIAECLSHIHKFWTDKGVFSGRDSEFCDIDDTSMGVRLLRMHG
YDVDPNVLRNFKQKDGKFSCYGGQMIESPSPIYNLYRASQLRFPGEEILEDADKF
AYDFLKEKLANNQILDKWVISKHLPDEIKLGIEMPWLATLPRVEAKYYIQYYAGSGD
AWIGKTLYRMPEISNDTYHDLAKTDFKRCQAQHQFEWIYMQEWYESCGIEEFGIS
RKDLLLYYFMAAASIFELERTNERLAWAKSQVISKMITCFFNSETRSFQEKQDLLN
ELKNINGLDSRNSGEREDGAASMVVATLTQFLGGFNRYTRHQLKNTWSVWLTKL
QDGKVDDAELLTNTLTICAGHTKETLSHSEYKTLSNLTTKICQQLSLIQTQKEMGV
DGHIAAIKNKELEQDMQMLVKLVLEKSGGINRNIKKTFLAVAKTYYYRAYHAANTID
THMFKVLFEPV*

A.3 *CPPSL2* nucleotide sequence

ATGTCATTCGCCACCTCCCTCCCCCGACCAACCACCACCGGCGCCGCCGGAT

TTGGCCTTCCGCTAGCAACATGTATCTCTCTCTCTGTCTCTCATTCCTTTTCCC

CAAAATTTGGCATTTGTAACAACACAAGTTTGAGACTCAAATCAAAGGCTGGGA

GCGGATGTTATGAGGGGATTCACAGAAGTCAATTAGCAGCATCAACAATTTTGG

AGGGTCACACTCCGATTAATCCGGAGGTCGAATCGGAGAAGATACGGCTGATT

GAAAGGATTCGTTTGATGTTCCGAAGCATGGACGATGGCGAGATAAGTGTGTC

ACCATACGACACAGCATGGGTGGCGCTCGTGGAAGATATTGGTGGCAGCGGG

GGGCCACAGTTTCCGACGAGCCTAGAGTGGATTTCTAACAACCAGTTGGATGA

TGGATCGTGGGGGGATCGCAAATTTGTTCTCTACGACCGGATACTCAATACGTT

AGCATGTGTTGTCGCACTCACGACTTGGAAAATGCATCCTAACAAATGCGAAA

AAGGGTTGAGGTTTATAAGTGATAATATTGAGAAACTCGCGGATGAAGACGAG

GAGCTCATGCCCGTAGGATTCGAAATCGCTCTGCCATCACTCATTGATTTAGCT

AAAAGATTGTGTATAGAAATCCCAGATAATTCTGCAAGCATAAAAAATATTTATGC

AAAGAGAGATTCAAAACTCAAAAGAATACCAATGGATTTAATGCACAAAAAGCC

AACATCACTACTCTTCAGCTTGGAAGGCATGGAAGGCCTTAACTGGGACAAAC

TATTGGATTTTCAATCCGAGGGTTCGTTTCTTTCATCGCCGTCGTCCACTGCCT

ACGCTCTCCACCACACCAAGGATGAGCTATGCCTAGAGTATCTGCTCAAGGCT

GTCAAGAAATTCCACGGTGGAGTTCCAAATGCATACCCTGTCGACATGTTTGA

GCATCTGTGGTCCGTGGATCGCTTGCGGAGATTAGGAATTTCTCGGTATTTTCA

AGTTGAAATCGATGAATGCCTCGACTATGTTTACAGATACTGGACAAATAAAGG

AATTTGCTGGGCACGAAATATGTGTGTCCAAGATAGTGACGACTCCTCAATGG

GATTCAGGCTATTAAGGTTGTACGGTTACGATGTTTCTATAGATGTTTTCAAACA

ATTCGAGGAAGGTGGACAATTCTGCAGCATACCAGGACAGATGACTCACGCTA

TTACAGGAATGTACAACTTGTATAGAGCTTCTCAACTTATGTTCCCTCAAGAACA

CATACTTGCCGATGCCAGAAACTTCACAGCTAACCTCTTGCATCAAAAAAGAGT

TACTAACTCAATAGTAGACAAGTGGATCATTACCAAAGACCTTCCAGGCGAGGT

GGCATATGCATTGGATGTGCCATTCTACGCCAGTCTGCCACGACTGGAAGCAC

GATTCTTCTTAGAACAATATGGGGGTGATGATGATGTTTGGATCGGGAAAACTT

TGTACAGGATGCTATATGTAAACTGCAACACATACCTTGAGCTGGCAAAATTAGA

TTACAAGCATTGCCAGACTGTGCATCAGTTAGAGTGGAATAGCATGCAAACATG

GTATAGAGAATGCAATCTAGGAGAGTTTGGGTTGAGCGAAAGAAGCCTTCTCC

TAGCTTACTATATAGCAGCCTCAACTGCATTTGAGCCGGAAAAATCAAGTGAGC

GGCTGGCTTGGGCTATAACAACAATTTTAGTCGAAACAATCATGTCCCAAGAAC

TCTCCGATGAACAAAAGAGAGAGTTTGTTGATGAATTTGTAAACATAAGCATCAT

CAATAACCAAAATGGAGGAAGATATAAACCAGGTAACAGATTGGTTGAGGTTTT

GATCAACACTGTAACACTGATGGCAGAAGGCAGAGGCACAGATCAGCTGTTGT

CTAATGCATGGAAAAATTGGCTAAAGACATGGGAAGAAGGAGGTGACCTGGG

GGAAGCAGAAGCACGGCTTCTCCTGCACACGATACATTTGAGCTCCGGATTG

GATGAATCATCATTTTCCCATCCAAAATATCAGCAGCTCTTGGAGGCGACCAGC

AAAGTCTGCCACCAACTTCGCCTATTCCAGAATCTAAAGCAGGCGAATGATGC

CCAAGGGTCTACAAGCCGTTTGGTAACTGTGACAACTTTCCAAATAGAAGCAG

GCATGCAAGAACTAGTGAAATTAATCTTCACAAAAACCTTGGAAGATTTGACTT

CTGCTACCAAACAAAGCTTTTTTAATATTGCTAGAAGCTTCTATTACACTGCCTAT

TGTCCTGCAGACACTATAGACTCTCACATAAACAAAGTATTGTTTGAAAAAATTG

TTTAG

A.4 *CPPSL2* amino acid sequence

MSFATSLPRPTTTGAAGFGLPLATCISLSVSHSFSPKFGICNNTSLRLKSKAGSGC
YEGIHRSQLAASTILEGHTPINPEVESEKIRLIERIRLMFRSMDDGEISVSPYDTAWV
ALVEDIGGSGGPQFPTSLEWISNNQLDDGSWGDRKFVLYDRILNTLACVVALTTW
KMHPNKCEKGLRFISDNIEKLADEDEELMPVGFEIALPSLIDLAKRLCIEIPDNSASI
KNIYAKRDSKLKRIPMDLMHKKPTSLLFSLEGMEGLNWDKLLDFQSEGSFLSSPS
STAYALHHTKDELCLEYLLKAVKKFHGGVPNAYPVDMFEHLWSVDRLRRLGISRYF
QVEIDECLDYVYRYWTNKGICWARNMCVQDSDDSSMGFRLLRLYGYDVSIDVFK
QFEEGGQFCSIPGQMTHAITGMYNLYRASQLMFPQEHILADARNFTANLLHQKRV
TNSIVDKWIITKDLPGEVAYALDVPFYASLPRLEARFFLEQYGGDDDVWIGKTLYR
MLYVNCNTYLELAKLDYKHCQTVHQLEWNSMQTWYRECNLGEFGLSERSLLLAY
YIAASTAFEPEKSSERLAWAITTILVETIMSQELSDEQKREFVDEFVNISIINNQNGG
RYKPGNRLVEVLINTVTLMAEGRGTDQLLSNAWKNWLKTWEEGGDLGEAEARLL
LHTIHLSSGLDESSFSHPKYQQLLEATSKVCHQLRLFQNLKQANDAQGSTSRLVT
VTTFQIEAGMQELVKLIFTKTLEDLTSATKQSFFNIARSFYYTAYCPADTIDSHINKVL
FEKIV*

A.5 *KSL1* nucleotide sequence

ATGTCGCTCGCCTTCAACGCCGAAGTCATGGCTTTCTCCGTTAACAGAGTGAG
AGAGGTTTTTCCGGTGAGGCATGTTACAGTCCAAGGATTTCCGGTGATCACCA
ATAAGTCATCTTTCGCCGTAAAATGCAATCTTACTACGACAGATTTGATGGGGAA
GATAGCAGAGAAGTTCAAGGGAGAAGACTGTCATTTTCCGCCAGCTGCCGCTT
TTGAACAAGCTGCGGATATACCCTCTTATCTGTGTATAATCGACACCCTCCAGA

GGTTGGGAGTCGACCGATACTTCCAGTCTGAAATCGACACCATTCTAGAGGAC

ACATACAGGTTATGGCAACAGAAAGAGAGAGATATATTTTCGGATATTACTATTC

ATGCAATGGCATTTAGACTCTTGCGAGTCAAAGGATACCAAGTTTCATCAGAGG

AACTGGCTCCGTATGCTGACCAAGAGCGCATTGACCTGCAAACGATTGATGTG

GCTACCGTTATTGAGCTTTACAGAGCAGCAGAGGAGAGAATAGATGAAGAAGA

GAGCAGTCTTAAGAAACTACATGCTTGGACCACCACTTTTCTCAAGCAGCAGT

TGCTCACTAATTCCATTCCTGACAAGAAGTCGCGGAAACTGGTAAATTACTACT

TGAAGAACTATCACGGGATATTAGATAGAATGGGAGTTAGACAAAACCTCGACC

TATTTGACATAACTCATTATCAAACACTAAAAGGTGCAGATAGAAACATGTTTTAT

CTTTTGTTTTAGATTCTCTAATCTACGCAATGAAGATTTTCTAGCATTTGCGAGG

AAAGATTTTAACATTTGCCAAGCCCAACACCAAAAAGAACTTCAGCAATTACAA

AGGTGGTATGCAGATTGTAGGTTGGACACCTTGAAGTATGGAAGAGATGTAGT

GCGTGTTTCTAATTTTCTAACTTCTGCAATTATCGGCGATCCTGAATTGTCTGAT

GTTCGTTTAGTCTTCGCCAAACATATTGTGCTTGTGACACGTATTGATGATTTTT

TCGATCATGCGGGGTCTAGGGAAGACTCCTACAAGATCCTTGAATTAGTAAAAG

AATGGAAAGAGAAGCCAGCTGCAGAATATGGCTCCGAGGAAGTTTAAATTCTT

TTTACAGCAGTATACAATACAGTTAACGAGTTGGCAGAGATGGCTCATGTCACA

CAAGAACGTAAAGAATTTCTAATAGAGCTGTGGGTTCAAATACTATCATGTTTCA

AGAAAGAATTAGATACGTGGAGCGATGACACTGCACTAACCTTGGATAGTTACT

TGTCTTCGTCATGGGTGTCCATTGGTTGCAGAATCTGCATTCTCATGTCAATGC

ACTTCATCGGAATGAAACTATCCGATGAAATGCTTCTGAGTGAAGAGTGCGTTG

ATCTGTGTAGGCTTGTCTCCATGGTTGACCGGCTGCTCAACGATGTGCAGACG

TTTGAGAAGGAACGCAAGGAAAATACGGGAAACAGTGTGAGCCTTCTGCTAG

CAGCTAATAAAGATGACAGTGCATTCACTGAAGAGGAAGCTAATACAAAGATAA
AAGAAATGGCTGAATGTAACAGAAGGAAATTGATGCAGATTGTCTATAAAACAG
GAACCCTCTTCCCGAGAAAATGCAAAGATATGTTTCTGAAGGTATGCAGGATTG
GGTGTTATTTGTATGCAAGCGGCGATGAGTTTACATCTCCACAACAGATGATGG
AAGACGTGAAATCCCTGGTTTACGAACCCTTAACAATACATCCTCTTGAAGCTA
ACAATGTGAGTGGCAAATGAAATGA

A.6 *KSL1* amino acid sequence

MSLAFNAEVMAFSVNRVREVFPVRHVTVQGFPVITNKSSFAVKCNLTTTDLMGKI
AEKFKGEDCHFPPAAAFEQAADIPSYLCIIDTLQRLGVDRYFQSEIDTILEDTYRLW
QQKERDIFSDITIHAMAFRLLRVKGYQVSSEELAPYADQERIDLQTIDVATVIELYRA
AEERIDEEESSLKKLHAWTTTFLKQQLLTNSIPDKKLRKLVNYYLKNYHGILDRMG
VRQNLDLFDITHYQTLKGADRFSNLRNEDFLAFARKDFNICQAQHQKELQQLQR
WYADCRLDTLKYGRDVVRVSNFLTSAIIGDPELSDVRLVFAKHIVLVTRIDDFFDHA
GSREDSYKILELVKEWKEKPAAEYGSEEVEILFTAVYNTVNELAEMAHVTQERKEF
LIELWVQILSCFKKELDTWSDDTALTLDSYLSSSWVSIGCRICILMSMHFIGMKLSD
EMLLSEECVDLCRLVSMVDRLLNDVQTFEKERKENTGSSVSLLLAANKDDSAFTE
EEANTKIKEMAECNRRKLMQIVYKTGTLFPRKCKDMFLKVCRIGCYLYASGDEFTS
PQQMMEDVKSLVYEPLTIHPLEANNVSGK*

A.7 *KSL2* nucleotide sequence

ATGTCGCTCGCCTTCACCATCGGAGTCACCCCTTTCTCCGGCCACCGAGTTG
GGTGTTTGAGCCAAACTTTTCCAGTTGTCAAAGCCATCGCCAAAAAGTCATATG

CCATCGGTAAATCCAGCCTTACTACAACAGATTTGTTGGGGGAATTAAAGGAGA

AGTTGAAGGGGGAAGAAGACGATGATTCTCGAGCAGCTGCAACTGTTCCACC

CTCGGATATACCCTCGAGTCTGTGTATAATCGACACCCTCGAAAGATTGGGAGT

CGACCGACACTTCCAATCCGAAATCGATACCATTCTTGAAGACACATATAGGTT

GTGGCAAAAGAAACACAAAGTTATATATTCAAGTGTCACTACTCATGCCATGGC

GTTTAGACTTCTGCGAGTGAAAGGATATGAAGTTTCATCAGAGGAACTGGCTC

CATACGCGGACCAAGAGCGCGTTAGCCGGCAAACAACTGATGTGGCGATGGT

TATCGAGCTTTACCGAGCAGCGTACGAGAGAATGTACGAAGACGAGAGCGGT

CTTGAGAAAATACTTGCTTGGACAACCACTTTTCTCAAGCACCAGTTGCAGAGT

AACTCCATTCCTGATAAGAAATTGCACCAACTGGTGGAATTTTATTTGAAGAACT

ATAATGGCATAACCAAGAGAATAGCAGCTCGACGAAACCTTGACCTATATGACA

TGAGCTATTATCAAGCACTAAAACCTACAGATAGGTTCTCTAATCTGTGTAATGA

AGATTTTCTAGTATTTGCAAGGCATGATTTTAATATCTGCCAAGCCCAGAACCAG

AAAGATCTTCAGCAACTACAGAGATGGTATGCAGATTGTAGGTTGGGCACATTG

AAATATGGAAGAGAAGCAATACTCATGTGTTATTTTTTGGCTTCATTAGCAATGG

AAGATCCAGAATTATCTTATGTTCGTGTAGCCTCTGCCAGCAATATGGTGCTAGT

AACATGTTTCGATGATTTTTTCGACGTTGGTGGCTCTAGACAAGAGTCATACAA

GATCATAGAACTAGTGAAAGAATGGAAAGAGAAGCCAGTTGCAGAGTATGGAT

CCAAGGAAGCTGAAATCCTTTTTCAAGCTTTGTACAATACAGTGAATGAGGTTG

CTGAGATGGCTGGTGTCGAACAAGGACGAGGTGTGAAAGGATTTCTTGTTGA

ACTGTGGATTGAAATGCTTTTAGCTATGAAGATAGAGCTGGATACATGGACTGA

CGGCGTGCAACTAAGCTTGGATGAGTACATGTCTAAATCGTGGATGTCGACCG

GTTGCAGAATATCACTTCTCATGTCGTCAATGTTCCTCGGTGTTAAATTATCTGA

AGAAATGCTTAGAAGTGAAGAGTGTAGCGATCTATGTAGGCATGTTTCGATAGT
TCGTCGTCTGCTCAACGATGTGACGACTTTCGAGAGGGAACGCAAGGAAAATA
AGGGAAGCAGTGTGAGCATTCTCATTGCAGAGGGAAGAAGTGCCACAGAGGA
GGAAGCTATTGCAGAGATAAAAGAAGTAGTTGACTATTACACAAGAAAACAGAA
GCAGATTGTCTACAAACAAGCCACCATTTTCCCAAGAAAATTCAAAGACATATTT
TTGGAGGGATGCAAGATTATTTCTCATATGTATTCGAGCAGCGACGAATTCACT
TCTCCTCAGCAACTGATGGCGGATGTGAAATCCTTGATATACGAGCCTCTTACA
CTTTCGAGTTAA

A.8 *KSL2* amino acid sequence

MSLAFTIGVTPFSGHRVGCLSQTFPVVKAIAKKSYAIGKSSLTTTDLLGELKEKLKG
EEDDDSRAAATVPPSDIPSSLCIIDTLERLGVDRHFQSEIDTILEDTYRLWQKKHKVI
YSSVTTHAMAFRLLRVKGYEVSSEELAPYADQERVSRQTTDVAMVIELYRAAYER
MYEDESGLEKILAWTTTFLKHQLQSNSIPDKKLHQLVEFYLKNYNGITKRIAARRNL
DLYDMSYYQALKPTDRFSNLCNEDFLVFARHDFNICQAQNQKDLQQLQRWYADC
RLGTLKYGREAILMCYFLASLAMEDPELSYVRVASASNMVLVTCFDDFFDVGGSR
QESYKIIELVKEWKEKPVAEYGSKEAEILFQALYNTVNEVAEMAGVEQGRGVKGFL
VELWIEMLLAMKIELDTWTDGVQLSLDEYMSKSWMSTGCRISLLMSSMFLGVKLS
EEMLRSEECSDLCRHVSIVRRLLNDVTTFERERKENKGSSVSILIAEGRSATEEEAI
AEIKEVVDYYTRKQKQIVYKQATIFPRKFKDIFLEGCKIISHMYSSSDEFTSPQQLM
ADVKSLIYEPLTLSS*

A.9 *KSL3* nucleotide sequence

ATGTCGCTTCCTCTCTCCACTTGCTTGTTACTTTCCCCCAGTGAATCACATTTT

CGGAGATTCCGTTTCTCTACTGCTTCAGCTTCTTTGAAAGCTGGGCTTCAAACT

ACTTCACCAAAAATCACCTCAATGCCAGCGTGCTTCGAGGAGACGAGAGGAA

GGATAGCAAAGCTGTTTGATAACGATGAACTTTCTGTTTCGACGTATGATACAG

CATGGGTTGCTATGGTCCCTTCCCCGACTTCGTTGGAGGAGCCTTGCTTCCC

GGACAGTCTAAACTGGTTGCTGGAGAACCAGTGCCCTGACGGTTCGTGGGCT

CGTCCCCACCACCACCCTTTGCTAAAGAAAGATGTCCTTTCTTCTACCCTGGC

CTGCATTCTTGCCCTTAAAAAATGGGGAGTTGGTGAAGAACAAATCATCAGGG

GTTTGCATTTTATGGAGTTGAATTTTGCTTCAGCTACCGATGAGTGTCAGATTAC

TCCTGTGGGATTCGATATTGTGTTTCCAGCCATGCTTGATTATGCCAGAGGTTC

CTCTTTAAACCTGCGGCTAGAGCCAACTACATTTAATGATTTGATGCATAAGAGA

GAATTGGAGCTTATAAGGGGCAACCAAAATCACACATCGGAGAGGGAAGCATA

CTGGGCATACATAGCTGAGGGAATGGGAGAGTTACAGAACTGGGAATCAATTA

TGAAATATCAAAGAAAAAATGGATCTTTTTTCAACTGTCCTTCTACGACAGCAG

CTGCTTTTATTGCCCTGCGTAATTCTGACTGCCTCAACTACTTACATTTAGCCCT

AAAGAAGTTTGGAAACGCAGTTCCTGCAGTTTATCCTCTAGATAAATACTCTCA

GTTGTGCATAGTTGACAATCTTGAAAGGTTGGGGATCAGACAGTATTTTTTGAC

GGAGATTCAAAATGTGTTGGATGAAACACACAGATGTTGGATGCAGGGCAATG

AAGAGATATTCATGGACGCCTCAACTTGCGCCTTAGCATTCCGGACGTTGCGA

TGGAATGGATATGATGTGACTTCAGATCCAGTGACGAAAATCCTACAACAATGC

TTTCCGAGTTCCATTATGACAGACATTAAAACAACTCTTGAATTATATAGGGCTT

CTGAACTCACACTTTATCCTGATGAGAGAGATCTTGATGAACAAAATTTAAGGC

TTAAACTTTTACTTGAGCAAGAATTAGAGACTGGTTTGAGTCAATCATGTCAACT

TGGAAGAGGAATTAATGCTGAGGTGAACCAGGCTATCGAGTATCCATTTTATGC

AATCATGGACAGGGTGGCAAAGCAGAAAAGTATAGAGCACTACAACCTCGATA

ATACAAGAATTCTGAAAACTTCGTTTTGTTTGCCTAATTTTAGCAACAAGGATTT

CCTTTTTCTGTCCATAGAAGACTTCAATCGATGTCAAGCTGCACATCGTGAAGA

ACTCAGAGAACTTGAAAGATGGGTTGTGGAGAATAGATTGGACGAGCTAAAGT

TTGCAAGGAGTAAGTCTGCCTACTGTTATTTTTCTGCAGCAGCAACCTTTTTTG

CTCCAGAACTATCAGATGCACGCATGTCGTGGGCCAAAAATGGTGTTCTAACG

ACTGTGGTAGATGACTTTTTTGATCTTGGAGGTTCAATGGAGGAATTGAAGAAC

TTAATCAAGTTGGTTGAATTATGGGACGTGGATGTTAGCACTGAATGCTCTTCC

AACAATGTCCAGATAATATTTTCAGCACTTAAGCACACAATCTGTGAGATTGGA

GACAAAGCATTCAAGCTACAAGGGCGTTGTGTTACCAACCATATAATAGGCATT

TGGCTAGATTTGCTAAATTCTATGATGAGAGAAACTGAATGGGCCAGAGACAAC

TTTGTCCCAACTATTGATGAATATATGAGCAACGCACATGTATCATTTTCTCTGG

GGCCAACTGTCCTACCAGCTCTTTATCTTGTGGGGCCGAAGCTCTCGGAAGA

GATGGTTAACCAACCTGAGTACCATAACCTATTCAAATTGATGAGCACGTGCGG

ACGCCTTCTGAATGACGTCCGTGGCTATGAGAGGGAGCTCAAAGATGGTAAAC

TAAACGCATTATCTCTGTACATTAATAATCATGGCGGTGAAGTAAGTAAAGAAGC

TGCCATTTCTGTGATTGAAAGCTGGATTGAGACACAAAGGAGAGAATTACTGA

GAATGGTTTTGGAGGGGAAGAAGAGTGTTCTACCAAAGCCGTGCAAGGCATT

GTTTTGGCATATGTGCTCTGTGGTGCACCAGTTCTACAGTAAAGACGACGGAT

TCACCAAGCAAGATTTGCTTGAGGTTGTAAATGCAATTATTCATCAACCTATTCT

TCTCAAGGAGCAAAAGTTGTGA

A.10 *KSL3* amino acid sequence

MSLPLSTCLLLSPSESHFRRFRFSTASASLKAGLQTTSPKITSMPACFEETRGRIA
KLFDNDELSVSTYDTAWVAMVPSPTSLEEPCFPDSLNWLLENQCPDGSWARPHH
HPLLKKDVLSSTLACILALKKWGVGEEQIIRGLHFMELNFASATDECQITPVGFDIV
FPAMLDYARGSSLNLRLEPTTFNDLMHKRELELIRGNQNHTSEREAYWAYIAEGM
GELQNWESIMKYQRKNGSFFNCPSTTAAAFIALRNSDCLNYLHLALKKFGNAVPA
VYPLDKYSQLCIVDNLERLGIRQYFLTEIQNVLDETHRCWMQGNEEIFMDASTCAL
AFRTLRWNGYDVTSDPVTKILQQCFPSSIMTDIKTTLELYRASELTLYPDERDLDE
QNLRLKLLLEQELETGLSQSCQLGRGINAEVNQAIEYPFYAIMDRVAKQKSIEHYN
LDNTRILKTSFCLPNFSNKDFLFLSIEDFNRCQAAHREELRELERWVVENRLDELK
FARSKSAYCYFSAAATFFAPELSDARMSWAKNGVLTTVVDDFFDLGGSMEELKNL
IKLVELWDVDVSTECSSNNVQIIFSALKHTICEIGDKAFKLQGRCVTNHIIGIWLDLL
NSMMRETEWARDNFVPTIDEYMSNAHVSFSLGPTVLPALYLVGPKLSEEMVNQP
EYHNLFKLMSTCGRLLNDVRGYERELKDGKLNALSLYINNHGGEVSKEAAISVIES
WIETQRRELLRMVLEGKKSVLPKPCKALFWHMCSVVHQFYSKDDGFTKQDLLEV
VNAIIHQPILLKEQKL*