Investigating the evolution of cytochromes P450 involved in GA biosynthesis

by

Sibongile Mafu

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Biochemistry

Program of Study Committee:

Reuben J. Peters Major Professor
David Hannapel
Mark Hargrove
Alan M. Myers
Basil Nikolau

Iowa State University

Ames, Iowa

2013

TABLE OF CONTENTS

List of Abbreviations

CPP	copalyl diphosphate
CPS	copalyl diphosphate synthases
CPR	cytochrome P450 reductase
CYP	cytochromes P450
DMAPP	dimethylallyl pyrophosphate
GA	gibberellic acid
GC	gas chromatography
GC-FID	gas chromatography – flame ionization detector
GC-MS	gas chromatography – mass spectrometry
GGPP	geranyl geranyl diphosphate
IPP	isopentenyl diphosphate
KO	kaurene oxidase
KAO	kaurenoic acid oxidase
KS	kaurene synthase
KSL	kaurene synthase -like
LRD	labdane related diterpenoids
MEP	methyl erythritol pathway
MVA	mevalonic acid
MW	molecular weight
NAD(P)H	nicotinamide adenine dinucleotide (phosphate) reduced
ORF	open reading frame
TPS	terpene synthase

Abstract

Vascular plants invariably contain a class II diterpene cyclase (EC 5.5.1.x), as an *ent*-copalyl diphosphate synthase is required for gibberellin phytohormone biosynthesis. This has provided the basis for evolution of a functionally diverse enzymatic family. A bifunctional diterpene synthase was characterized from the lycophyte *Selaginella moellendorffii*. The structure of its product, labda-7,13*E*-dien-15-ol, demonstrates that this enzyme catalyzes a novel class II diterpene cyclization reaction, and clarifies the biosynthetic origins of the family of derived natural products.

All higher plants contain kaurene oxidases (KO), which are multifunctional cytochromes P450 that catalyze oxidation at the C4α methyl, converting *ent*-kaurene to *ent*-kaurenoic acid, an early step in gibberellin phytohormone biosynthesis. Arabidopsis produces no labdane-related diterpenoids other than gibberellins, whereas rice produces a wide range of such natural products as antibiotic phytoalexins or allelochemicals. While rice contains several kaurene oxidase homologs, only OsKO2 (CYP701A6) is required for gibberellin biosynthesis. Here, we demonstrate that the KO from *Arabidopsis thaliana* (CYP701A3 or AtKO) exhibits significantly greater promiscuity than does OsKO2. To further characterize this plasticity of AtKO, we determined the structure of the resulting products, whereas OsKO2 only hydroxylates its substrates on C19, AtKO reacts with labdane-related diterpenes of varied stereochemistry, which further leads to altered hydroxylation regiochemistry. Thus, our data are consistent with the hypothesis that enzymatic plasticity is shaped, at least in part, by evolutionary context such as that noted here.

CHAPTER I: INTRODUCTION

Overview and Significance

Giberellins (GAs) are complex diterpene phytohormones produced by plants, fungi and bacteria utilized for a variety of natural developmental and regulatory purposes in plant development. GA has played an important role in agriculture leading to semi-dwarf varieties of rice and wheat that were integral part of the Green Revolution and it continues to be used in agricultural settings today [1]. Evolutionary, duplication and alternative metabolism of enzymes from the GA pathway has led to the development of biosynthetic pathways of labdane related compounds which have pharmaceutical and industrial application. Because of the absolute requirement for GA in plants and its role in production of natural products, it is important to continually advance our understanding of its evolution, biosynthesis and metabolism.

Terpenoids

Terpenoids are a structurally diverse group of natural compounds involved in various plant functions. The primary terpenoids include photosynthetic pigments, plant sterols, electron carriers and phytohormones such as GA and abscicic acid [2]. Secondary terpenoids -monoterpenes, sesquiterpenes, diterepenes and triterpenes- have vital physiological roles in plant communication and defense and applied uses in pharmaceutical and industrial setting. Terpenoids all originate from condensation of isopentenyl diphosphate (IPP) and dimethyallyl diphosphate DMAPP which are synthesized through two different pathways the mevalonate pathway (MVA) in the cytosol and the methylerythritol pathway (MEP) in the plastids and also found in prokaryotes. In brief, the MVA pathway initiates with successive

condensation of three molecules of acetyl-CoA to form HMG-CoA which is reduced to mevalonate by HMG-CoA reductase. Mevalonate undergoes double phosphorylation and decarboxylation resulting in IPP which can be converted DMAPP by IPP isomerase. The MEP pathway initiates with condensation of pyruvate and glyceraldehyde-3-phosphate to form DXP (1-deoxy-D-xylulose-5-phosphate), an intermediate for IPP and DMAPP formation [2, 3]. Prenyltransferase catalyzes head to tail condensation to produce precursors of terpenoids metabolism (C10, C15 and C20) (Figure 1). Triterpenes and tetraterpenes are produced by head to head condensation of two C15 or C20 molecules respectively**.**

GA biosynthesis

GA is important in regulation of growth and development of plants ranging from seed germination, stem elongation flower initiation and development, fruit and seed development [4, 5]. Its complex biosynthesis consists of provision of upstream precursors of DMAPP and IPP which undergo head to tail condensation to form (*E*,*E*,*E*) geranyl geranyl diphosphate (GGPP) in proplastids. Through ^{13}C labeling studies it was shown that GA is synthesized predominantly through the MEP pathway and MVA pathway plays a major role in the production of sterols but can still contribute to GA pathway under limiting conditions[3]. Thereafter, it is divided into three stages comprising of terpene synthases (TPS) which catalyze the reaction of GGPP to *ent*-kaurene, two functionally independent P45O groups which convert ent-kaurene to GA_{12} and the formation of bioactive GAs in plants is catalyzed by 2-oxoglutarate dependent dioxygenases (2ODDs) [5].

The initial step in GA biosynthesis is the proton-initiated cyclization of GGPP to *ent*-labdadienyl/copalyl diphosphate (*ent*-CPP) mediated by a stereospecific CPP synthase (CPS).

CPS are class II enzymes that produce the initial carbocation using a conserved DXDD motif. Other stereochemistry, *syn* and normal, are found within plants and are intermediates for secondary metabolites [6]. CPS is demonstrated to be a rate limiting step in GA biosynthesis through substrate inhibition exerted by GGPP and magnesium ion enzymatic co factor [7, 8]. Kaurene synthase is a class I enzyme that catalyzes ionization- initiated cyclization of the CPS products mediated by divalent magnesium ions bound by conserved DDXXD and NDX_2TX_3E motifs [9]. TPS reactions are localized in the plastids [10] and loss of function of CPS and KS genes results in dwarfism[4].

Further elaboration of *ent*-kaurene to GA_{12} is carried out by downstream NADPH-dependent membrane bound cytochromes P450 monooxygenase enzymes. Conversion of *ent*-kaurene to *ent*-kaurenoic acid is catalyzed by kaurene oxidase (KO) in the CYP701 family in a multi-step sequential reaction [11, 12] . The catalytic action of the kaurene oxidase occurs through consecutive hydroxylation reactions through a gem-diol intermediate as demonstrated utilizing isotopic labeling studies [12]. Functional KO's have been characterized in angiosperms notably *Arabidopsis*, pea, cucumber, rice and lower non vascular land plant *Physcomitrella patens* belonging to the CYP701A and CYP701B subfamilies respectively [11, 13, 14]. Mutations in AtKO and OsKO2 result in severe dwarfism [15, 16]. In rice OsKO2 is found in a gene cluster in tandem with four other KOL of which CYP701A8/OsKOL4 is involved in specialized metabolism where it hydroxylates kaurene, *ent*-cassadiene and *ent*-sandaracopimaradiene at the C3α position (Figure 3) [17]. This is another example of evolutionary duplication of primary metabolism genes for use in secondary metabolism.

A second cytochrome P450, kaurenoic acid oxidase (CYP88 family) converts *ent*-kaurenoic acid to GA_{12}. Biosynthesis of gibberellins in higher plants has been obtained from the studies of cell free systems where *ent*-7α-hydroxykaurenoic acid , GA_{12} aldehyde and GA_{12} were identified as conversion products from mevalonic acid [18]. Later, two genes for kaurenoic acid oxidase were cloned from *Arabidopsis* and demonstrated to carry out the conversion of *ent*-kaurenoic acid to GA_{12} [19]. The gene was confirmed to carry out the three step sequential reaction by feeding intermediates, *ent*-7α-hydroxykaurenoic acid and GA_{12} aldehyde, which still produced GA_{12}. KAO is localized on the cytosolic face of the endoplasmic reticulum [10].

GA_{12} is converted to GA_4 through oxidations on C-20 by GA20 oxidase and on C-3 by GA3 oxidase GA_{12} is also a substrate for GA13-oxidase which then undergoes C-20 and C-3 oxidation resulting in the production of 13-hydroxylated bioactive GAs (Figure 2).

There are 136 identified GA compounds (http://www.planthormones.info/gibberellin_nomenclature.htm) of which there are four bioactive GAs (GA_1, GA_3 GA_4 and GA_7) used in regulation of growth and development. These four have a 6-5-6-5 tetracyclic ring, , 3β-hydroxyl, carboxyl on C6 and a lactone ring between C4and C10 [4]. GA_1 and GA_3 are in addition hydroxylated on C13 (Figure 2). Many non-bioactive GAs exist in plants as precursors for bioactive forms or deactivated metabolites. Most common deactivation mechanism is 2β-hydroxylation catalyzed by GA2 oxidases though epoxidation and methylation also occur [20-22].

Terpene biosynthesis in early land plants

*S. moellendorffii (*model system for lycophytes) has a functional GA signaling system whereas that of the moss, *P. patens* (model system for bryophytes), did not exhibit functional

conservation [23]. *P. patens* contains a bifunctional terpene synthase PpCPSKS and KO enzymes to produce *ent*- kaurenoic acid which was demonstrated to have a physiological role in *P. patens* as protonematal defects of bifunctional *cps/ks* mutants could be rescued by application of kaurene or kaurenoic acid [13, 24]. Whether it's actually *ent*-kaurenoic acid or downstream metabolites which play a role in development is yet to be clarified. *P. paten* does not, however, contain the downstream enzymes involved in GA biosynthesis (Hirano 2007). *S. moellendorffii* contains a range of terpene synthase genes of which a bifunctional SmCPSKS and a monofunctional SmCPS and SmKS were reported to produce *ent*-kaurene (Ueno *et al*. 2010). The downstream enzymes for GA biosynthesis GA20 and GA3 oxidases have been functionally characterized and are functionally similar to those in angiosperms which have been identified and functionally characterized [23].

The *Selaginella* genome has been reported to encode 18 TPS genes which are homologous to plant diterpene synthases. Of these, 4 have been characterized SmCPSKSL1 which produces labdadienol, SmMDS which produces miltradiene synthase and SmTPS9 and SmTPS10 *ent*-copalyl diphosphate producers [25-27]. Interestingly, 48 TPS genes have also been identified in *Selaginella*, 6 of which have been characterized as monoterpene and sesquiterpene synthases [27]. It appears SmMTPS genes have been attained from microbes through horizontal gene transfer and evolved to have new functions. This results in a new class of TPS in land plants which is evolutionarily distinct from that of higher plants.

Cytochromes P450

Cytochromes P450 are heme-thiolate monooxygenase existing in mammalian, plants, bacteria, and yeast, that exhibit characteristic reduced CO binding absorption band maximum near 450nm [28]. They participate in the production of primary and secondary

metabolites and detoxification of xenobiotics. P450s are often associated with hydroxylation reactions mediated by insertion of oxygen into a carbon-hydrogen bond but they can catalyze diverse reactions such as desaturation, dehydration, epoxidation, dealkylation, coupling reactions, decarboxylation, ring contraction, ester cleavage and methyl transfer [29]. In catalyzing their mono-oxygenation reactions, the P450 requires two one-electron transfer steps using NADH or NADPH as electron donor. The general reaction can be expressed as:

$$RH + O_2 + NAD(P)H+H^+ \rightarrow ROH + H_2O+NAD(P)^+$$

To acquire electrons from NAD(P)H, most P450s need a separate electron transport partner. Generally, bacteria and mitochondrial P450s utilize a two-component shuttle system consisting of an iron-sulfur protein and ferrodoxin reductase, whereas microsomal P450s employ NADPH-cytochrome P450 reductase (CPR), a single flavoprotein containing both FAD and FMN domains which shuttle electrons to the heme-iron center of the P450 [30].

Metabolic engineering of terpenoids

To aid in the functional characterization of enzymes in the biosynthetic pathways of model systems, the Peters' group developed a versatile modular metabolic engineering system. The system enables co-transformation of GGPPs, CPP synthase, KSL to produce a wide range of olefins [31]. As the N-terminal end plastid targeting sequence interferes with correct protein folding, truncated versions of the TPS genes are utilized. The system has also been successfully optimized for the bacterial expression of CYPs and their corresponding reductase, through whole gene codon optimization and N-terminal modification [12, 17, 32-35]. The need to produce increased quantities of compound sufficient for structural identification by NMR led to co-expression including a pMBI plasmid which contains genes that encode for bottom half of the MVA pathway from *S. cerevisiae* or pIRS which encodes

for the MEP pathway both of which increase production of IPP and DMAPP improving flux [36, 37].

Metabolites are extracted by hexanes and analyzed by GC-MS, (diterpenes have a molecular weight of 272 Da, alcohols 288 Da, aldehydes 286 Da, and 316 Da MW corresponds to acid products derivatized with diazomethane). The modular metabolic engineering system enables functional characterization, determination of stereochemistry of TPS genes, production of sufficient quantities for structural determination by NMR as well as comprehensive substrate specificity screening.

Organization of the dissertation and author contribution

The dissertation describes the characterization of a diterpene synthase cloned from *S. moellendorffii*. I carried out cloning and functional characterization and Dr. Matthew Hillwig carried out NMR analysis. The work was published in ChemBioChem 2011. Chapter III describes biochemical evolution of KO genes in higher plants. The work was initiated by Dr. Dana Morrone and carried on by Dr. Yisheng Wu. I performed the comparative study and isolated and purified 6 of the 7 resultant products described. Dr. Jiachen Zi and Dr. Matthew Hillwig carried out the NMR analysis. The project is near completion with the exception of relative kinetic constants. Chapter IV describes incomplete work, which continues to explore evolution of KO genes in land plants. The final chapter discusses the research, its implications and future directions.

References

1. Hedden P: **The genes of the Green Revolution**. *Trends Genet* 2003, **19**(1):5-9.
2. McGarvey DJ, Croteau R: **Terpenoid metabolism**. *Plant Cell* 1995, **7**(7):1015-1026.

3. Kasahara H, Hanada A, Kuzuyama T, Takagi M, Kamiya Y, Yamaguchi S: **Contribution of the mevalonate and methylerythritol phosphate pathways to the biosynthesis of gibberellins in Arabidopsis**. *J Biol Chem* 2002, **277**(47):45188-45194.
4. Yamaguchi S: **Gibberellin metabolism and its regulation**. *Annu Rev Plant Biol* 2008, **59**:225-251.
5. Hedden P, Thomas SG: **Gibberellin biosynthesis and its regulation**. *Biochem J* 2012, **444**(1):11-25.
6. Peters RJ: **Two rings in them all: the labdane-related diterpenoids**. *Nat Prod Rep* 2010, **27**(11):1521-1530.
7. Fleet CM, Yamaguchi S, Hanada A, Kawaide H, David CJ, Kamiya Y, Sun TP: **Overexpression of AtCPS and AtKS in Arabidopsis confers increased ent-kaurene production but no increase in bioactive gibberellins**. *Plant Physiol* 2003, **132**(2):830-839.
8. Prisic S, Peters RJ: **Synergistic substrate inhibition of ent-copalyl diphosphate synthase: a potential feed-forward inhibition mechanism limiting gibberellin metabolism**. *Plant Physiol* 2007, **144**(1):445-454.
9. Zhou K, Peters RJ: **Investigating the conservation pattern of a putative second terpene synthase divalent metal binding motif in plants**. *Phytochemistry* 2009, **70**(3):366-369.
10. Helliwell CA, Sullivan JA, Mould RM, Gray JC, Peacock WJ, Dennis ES: **A plastid envelope location of Arabidopsis ent-kaurene oxidase links the plastid and endoplasmic reticulum steps of the gibberellin biosynthesis pathway**. *Plant J* 2001, **28**(2):201-208.
11. Helliwell CA, Poole A, Peacock WJ, Dennis ES: **Arabidopsis ent-kaurene oxidase catalyzes three steps of gibberellin biosynthesis**. *Plant Physiol* 1999, **119**(2):507-510.
12. Morrone D, Chen X, Coates RM, Peters RJ: **Characterization of the kaurene oxidase CYP701A3, a multifunctional cytochrome P450 from gibberellin biosynthesis**. *Biochem J* 2010, **431**(3):337-344.
13. Miyazaki S, Katsumata T, Natsume M, Kawaide H: **The CYP701B1 of Physcomitrella patens is an ent-kaurene oxidase that resists inhibition by uniconazole-P**. *FEBS Lett*, **585**(12):1879-1883.
14. Ko KW, Lin F, Katsumata T, Sugai Y, Miyazaki S, Kawaide H, Okada K, Nojiri H, Yamane H: **Functional identification of a rice ent-kaurene oxidase, OsKO2, using the Pichia pastoris expression system**. *Biosci Biotechnol Biochem* 2008, **72**(12):3285-3288.
15. Helliwell CA, Sheldon CC, Olive MR, Walker AR, Zeevaart JA, Peacock WJ, Dennis ES: **Cloning of the Arabidopsis ent-kaurene oxidase gene GA3**. *Proc Natl Acad Sci U S A* 1998, **95**(15):9019-9024.
16. Itoh H, Tatsumi T, Sakamoto T, Otomo K, Toyomasu T, Kitano H, Ashikari M, Ichihara S, Matsuoka M: **A rice semi-dwarf gene, Tan-Ginbozu (D35), encodes the gibberellin biosynthesis enzyme, ent-kaurene oxidase**. *Plant Mol Biol* 2004, **54**(4):533-547.

17. Wang Q, Hillwig ML, Wu Y, Peters RJ: **CYP701A8: a rice ent-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism**. *Plant Physiol* 2012, **158**(3):1418-1425.
18. GRAEBE JE, DENNIS DT, UPPER CD, WEST CA: **BIOSYNTHESIS OF GIBBERELLINS. I. THE BIOSYNTHESIS OF (-)-KAUREN-19-OL, AND TRANS-GERANYLGERANIOL IN ENDOSPERM NUCELLUS OF ECHINOCYSTIS MACROCARPA GREENE**. *J Biol Chem* 1965, **240**:1847-1854.
19. Helliwell CA, Chandler PM, Poole A, Dennis ES, Peacock WJ: **The CYP88A cytochrome P450, ent-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway**. *Proc Natl Acad Sci U S A* 2001, **98**(4):2065-2070.
20. Rieu I, Eriksson S, Powers SJ, Gong F, Griffiths J, Woolley L, Benlloch R, Nilsson O, Thomas SG, Hedden P *et al*: **Genetic analysis reveals that C19-GA 2-oxidation is a major gibberellin inactivation pathway in Arabidopsis**. *Plant Cell* 2008, **20**(9):2420-2436.
21. Zhu Y, Nomura T, Xu Y, Zhang Y, Peng Y, Mao B, Hanada A, Zhou H, Wang R, Li P *et al*: **ELONGATED UPPERMOST INTERNODE encodes a cytochrome P450 monooxygenase that epoxidizes gibberellins in a novel deactivation reaction in rice**. *Plant Cell* 2006, **18**(2):442-456.
22. Varbanova M, Yamaguchi S, Yang Y, McKelvey K, Hanada A, Borochov R, Yu F, Jikumaru Y, Ross J, Cortes D *et al*: **Methylation of gibberellins by Arabidopsis GAMT1 and GAMT2**. *Plant Cell* 2007, **19**(1):32-45.
23. Hirano K, Nakajima M, Asano K, Nishiyama T, Sakakibara H, Kojima M, Katoh E, Xiang H, Tanahashi T, Hasebe M *et al*: **The GID1-mediated gibberellin perception mechanism is conserved in the Lycophyte Selaginella moellendorffii but not in the Bryophyte Physcomitrella patens**. *Plant Cell* 2007, **19**(10):3058-3079.
24. Hayashi K, Horie K, Hiwatashi Y, Kawaide H, Yamaguchi S, Hanada A, Nakashima T, Nakajima M, Mander LN, Yamane H *et al*: **Endogenous diterpenes derived from ent-kaurene, a common gibberellin precursor, regulate protonema differentiation of the moss Physcomitrella patens**. *Plant Physiol*, **153**(3):1085-1097.
25. Mafu S, Hillwig ML, Peters RJ: **A novel labda-7,13e-dien-15-ol-producing bifunctional diterpene synthase from Selaginella moellendorffii**. *Chembiochem* 2011, **12**(13):1984-1987.
26. Sugai Y, Ueno Y, Hayashi K, Oogami S, Toyomasu T, Matsumoto S, Natsume M, Nozaki H, Kawaide H: **Enzymatic (13)C labeling and multidimensional NMR analysis of miltiradiene synthesized by bifunctional diterpene cyclase in Selaginella moellendorffii**. *J Biol Chem* 2011, **286**(50):42840-42847.
27. Li G, Köllner TG, Yin Y, Jiang Y, Chen H, Xu Y, Gershenzon J, Pichersky E, Chen F: **Nonseed plant Selaginella moellendorffi [corrected] has both seed plant and microbial types of terpene synthases**. *Proc Natl Acad Sci U S A* 2012, **109**(36):14711-14715.
28. Omura T, Sato R, Cooper DY, Rosenthal O, Estabrook RW: **Function of cytochrome P-450 of microsomes**. *Fed Proc* 1965, **24**(5):1181-1189.
29. Isin EM, Guengerich FP: **Complex reactions catalyzed by cytochrome P450 enzymes**. *Biochim Biophys Acta* 2007, **1770**(3):314-329.

30. Sono M, Roach MP, Coulter ED, Dawson JH: **Heme-Containing Oxygenases**. *Chem Rev* 1996, **96**(7):2841-2888.
31. Cyr A, Wilderman PR, Determan M, Peters RJ: **A modular approach for facile biosynthesis of labdane-related diterpenes**. *J Am Chem Soc* 2007, **129**(21):6684-6685.
32. Swaminathan S, Morrone D, Wang Q, Fulton DB, Peters RJ: **CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice**. *Plant Cell* 2009, **21**(10):3315-3325.
33. Wang Q, Hillwig ML, Peters RJ: **CYP99A3: functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice**. *Plant J* 2011, **65**(1):87-95.
34. Wang Q, Hillwig ML, Okada K, Yamazaki K, Wu Y, Swaminathan S, Yamane H, Peters RJ: **Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster**. *J Biol Chem* 2012, **287**(9):6159-6168.
35. Wu Y, Hillwig ML, Wang Q, Peters RJ: **Parsing a multifunctional biosynthetic gene cluster from rice: Biochemical characterization of CYP71Z6 & 7**. *FEBS Lett* 2011, **585**(21):3446-3451.
36. Morrone D, Lowry L, Determan MK, Hershey DM, Xu M, Peters RJ: **Increasing diterpene yield with a modular metabolic engineering system in E. coli: comparison of MEV and MEP isoprenoid precursor pathway engineering**. *Appl Microbiol Biotechnol* 2010, **85**(6):1893-1906.
37. Martin DN, Proebsting WM, Parks TD, Dougherty WG, Lange T, Lewis MJ, Gaskin P, Hedden P: **Feed-back regulation of gibberellin biosynthesis and gene expression in Pisum sativum L**. *Planta* 1996, **200**(2):159-166.

Figures

Figure 1: Methyl erythritol pathway and mevalonate pathway for production of IPP and DMAPP. Schematic illustration for backbone synthesis of terpenoids

Figure 2: GA Biosynthetic Pathway: Bioactive GAs are highlighted in rectangles. GA_7 is biosynthesized from GA_9 in a similar manner as GA3 from GA20 but is not shown. The properties of bioactive GAs are highlighted in red. CPS, *ent*-copalol diphosphate synthase; KS, *ent*-kaurene synthase; KO, *ent*-kaurene oxidase; KAO, *ent*-kaurenoic acid oxidase

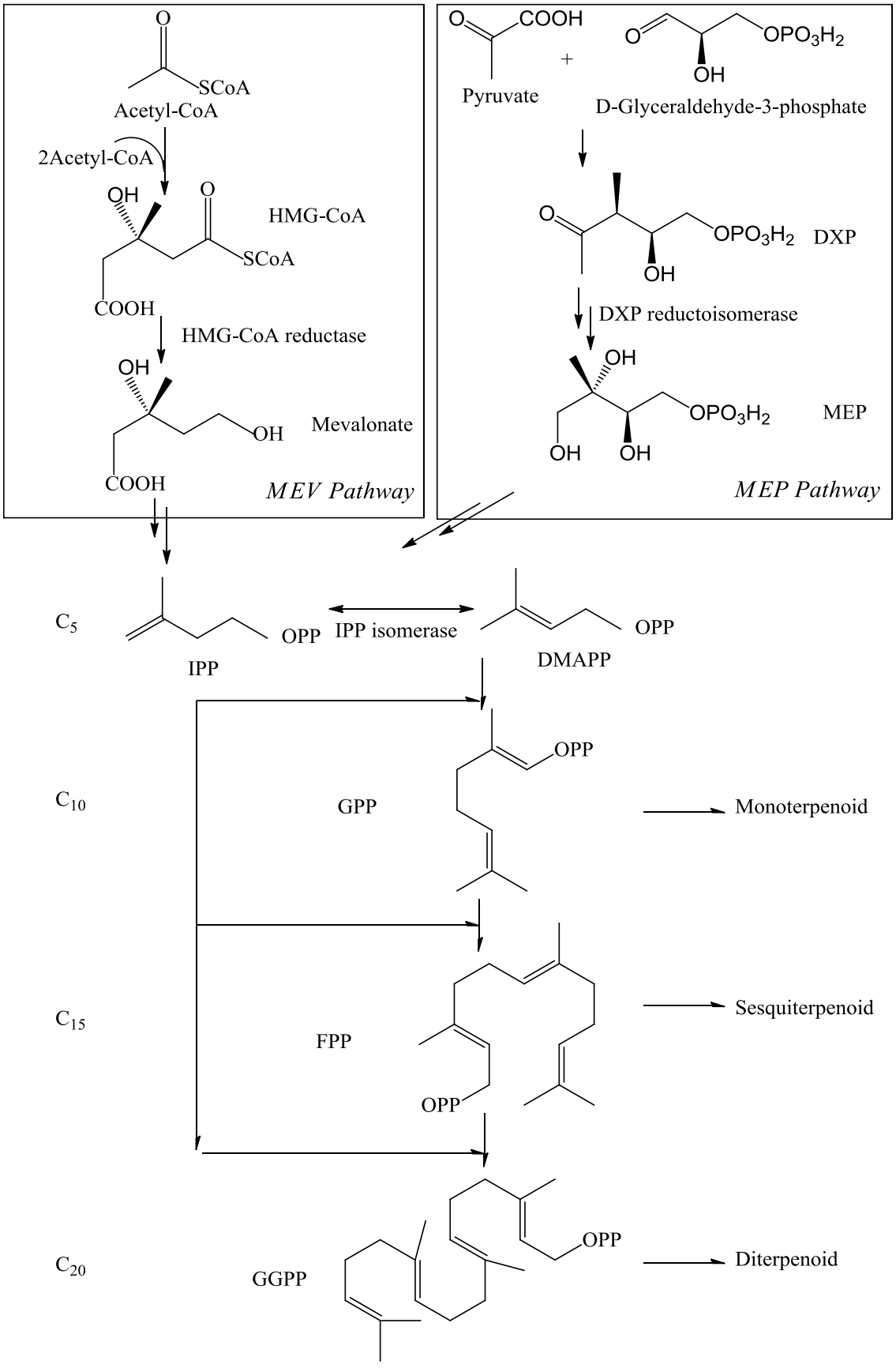

Acetyl-CoA
2Acetyl-CoA
HMG-CoA
HMG-CoA reductase
Mevalonate
MEV Pathway
Pyruvate
D-Glyceraldehyde-3-phosphate
DXP
DXP reductoisomerase
MEP
MEP Pathway
C5
IPP
IPP isomerase
DMAPP
C10
GPP
Monoterpenoid
C15
FPP
Sesquiterpenoid
C20
GGPP
Diterpenoid

Figure 2

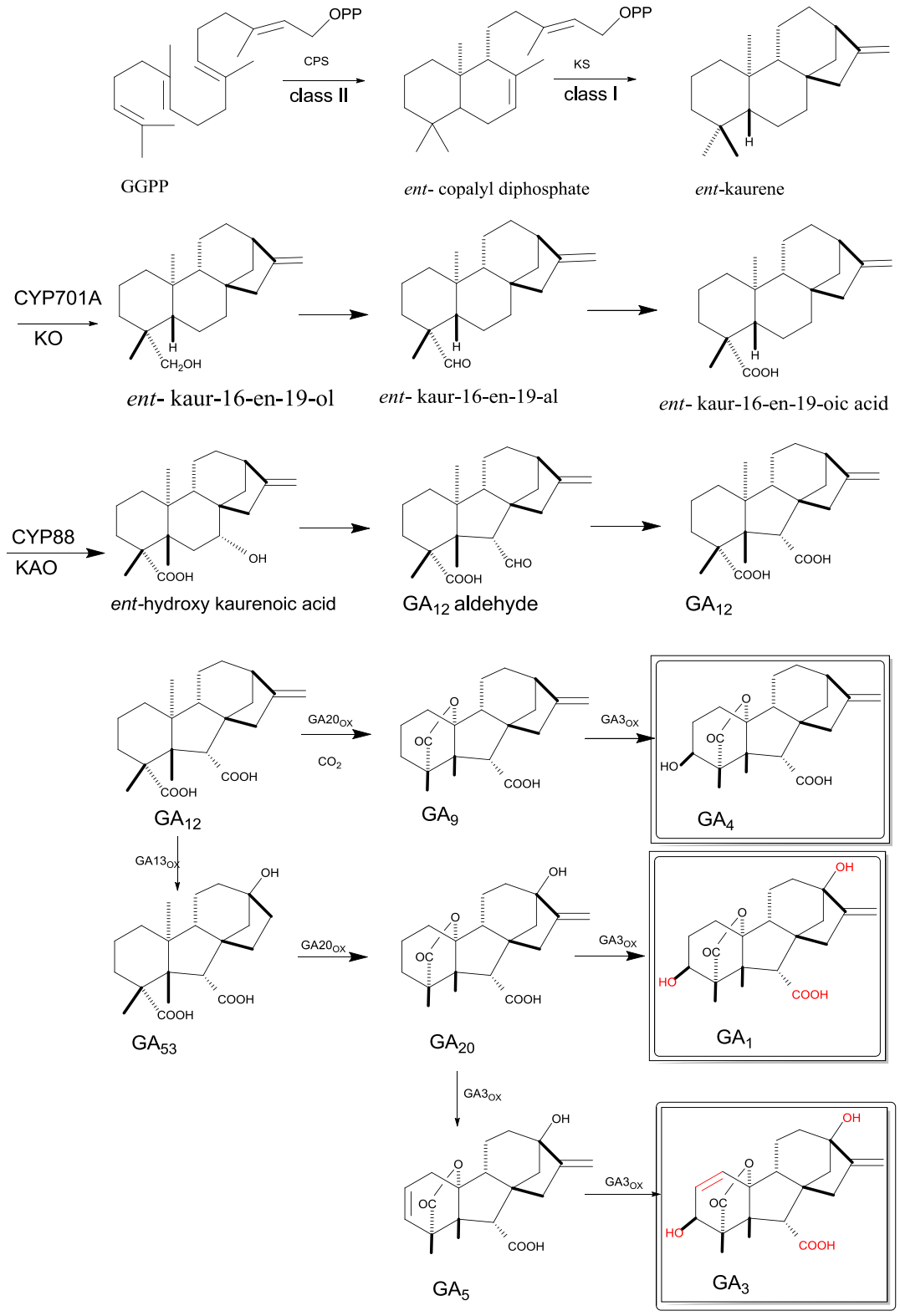
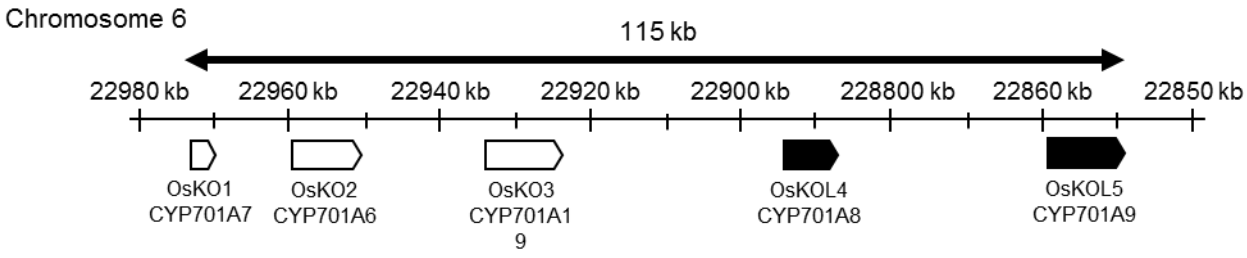

OPP
CPS
class II
GGPP
OPP
KS
class I
ent- copalyl diphosphate
ent-kaurene
CYP701A
KO
CH2OH
ent- kaur-16-en-19-ol
CHO
ent- kaur-16-en-19-al
COOH
ent- kaur-16-en-19-oic acid
CYP88
KAO
OH
ent-hydroxy kaurenoic acid
GA12 aldehyde
GA12
GA20OX
CO2
GA9
GA3OX
GA4
GA13OX
GA53
GA20
GA1
GA5
GA3

Figure 3: Biosynthetic gene cluster of CYP701A subfamily in rice (non-inducible KO in white and inducible KOL in black) from Wang et al. (2010) Plant Phys.



CHAPTER II: A NOVEL LABDA-7,13*E*-DIEN-15-OL PRODUCING BIFUNCTIONAL DITERPENE SYNTHASE FROM *SELAGINELLA MOELLENDORFFII*

A paper published in ChemBioChem (2011) 12(13): 1984-1987

Sibongile Mafu, Matthew L. Hillwig, and Reuben J. Peters*[1]

Abstract

A bifunctional diterpene synthase from the lycophyte *Selaginella moellendorffii* has been characterized. The structure of its product, labda-7,13*E*-dien-15-ol, demonstrates that this enzyme catalyzes a novel class II diterpene cyclization reaction, and clarifies the biosynthetic origins of the family of derived natural products.

Introduction

Vascular plants invariably contain a class II diterpene cyclase (EC 5.5.1.x), as an *ent*-copalyl diphosphate synthase is required for gibberellin phytohormone biosynthesis. This has provided the basis for evolution of a functionally diverse enzymatic family[1]. These biocatalysts fold their substrate, the general diterpenoid precursor (*E*,*E*,*E*)-geranylgeranyl diphosphate (GGPP), to bring the terminal three carbon-carbon double bonds into proximity with each other, and then carry out bicyclization via a protonation-initiated carbocation cascade reaction. The resulting labda-15-en-8-yl^{+} diphosphate intermediate is most

[1] Department of Biochemistry, Biophysics and Molecular Biology, Iowa State University, Ames, IA, 50011, USA

*Corresponding author: 4216 Mol. Biol. Bldg., Ames, IA 50011,

FAX: 515-294-8580, E-mail: rjpeters@iastate.edu

commonly quenched by deprotonation at an exocyclic methyl, as in the production of labdadienyl/copalyl diphosphate (Scheme 1). Alternatively, the bicyclized labda-15-en-8-yl^{+} diphosphate intermediate can be captured by water prior to deprotonation, to form hydroxylated compounds such as labda-15-en-8-ol diphosphate[2]. In addition, this intermediate can undergo subsequent rearrangement via 1,2-hydride and/or methyl shifts, starting with the hydrogen substituent on the neighboring endocyclic methine (C9)[3, 4]. However, terminating deprotonation at the neighboring endocyclic methylene (C7) has not previously been observed. Here we report that the lycophyte *Selaginella moellendorffii* contains a bifunctional diterpene synthase, SmCPSKSL1, which catalyzes just such a class II cyclization reaction. In particular, SmCPSKSL1 produces an endocyclic double bond isomer of copalyl diphosphate (CPP), as well as carries out subsequent replacement of the diphosphate by a hydroxyl group to form labda-7,13*E*-dien-15-ol. Although this is a known plant metabolite, [5, 6] and a small family of bioactive derived natural products is known from a phylogenetically diverse group of plants, [5-11] its biosynthesis has not been previously investigated. Our results demonstrate that this diterpenoid can be generated by a single bifunctional diterpene synthase that directly generates the endocyclic double bond, as well as hydroxyl group.

S. moellendorffii is a basal vascular land plant whose genome has been sequenced, and which contains a number of putative diterpene synthases.[12] At least one of these appears to be a bifunctional diterpene synthase (corresponding to the Joint Genome Institute gene accession Selmol 112927), as it contains both the DXDD motif associated with class II activity and DDXXD motif associated with allylic diphosphate ionization initiated class I activity. In an attempt to determine the biochemical function of the encoded enzyme, a

corresponding synthetic gene was obtained (Genscript), with codon-optimization for expression in *Escherichia coli*. Unfortunately, no diterpene synthase activity was observed with the resulting recombinant protein. The corresponding gene was then cloned from the lycophyte itself, revealing a few key differences from the predicted gene sequence, most notably a nine-nucleotide insertion. Expression of this native gene (SmCPSKSL1) in *E. coli* provided diterpene synthase activity. Specifically, upon co-expression with a GGPP synthase, using a previously described metabolic engineering system,[13] this bifunctional diterpene synthase predominantly led to production of an unknown diterpenoid by the recombinant bacteria, albeit with a few other minor products also observed (Figure 1).

The structure of this predominant unknown diterpenoids, which represented ~80% of the total product output, was determined by NMR spectroscopy. Sufficient amounts of material for this purpose was produced by increasing metabolic flux into isoprenoid/terpenoid metabolism in the metabolically engineered bacteria, as previously described,[14] followed by extraction from 2 L cultures of the resulting recombinant bacteria. Following purification by silica gel and reverse-phase high-performance liquid chromatography, 18 mg of compound were obtained, thus enabling straightforward characterization. In particular, the structure was determined using HMBC, HSQC, and COSY data, to be labda-7,13*E*-dien-15-ol, which was confirmed by comparison to data in the literature.[5, 6]

The presence of the primary C15 hydroxyl group essentially represents hydrolysis of the allylic diphosphate ester bond. While similar replacement of the diphosphate has been observed with other class I terpene synthases, [15] here it was possible that this was a result of the endogenous phosphatases from the *E. coli* based metabolic engineering system. To

verify that this was catalyzed by SmCPSKSL1, the recombinant enzyme was expressed with a His_6 tag, purified, and shown to catalyze direct production of labda-7,13*E*-dien-15-ol *in vitro*, even in the presence of phosphatase inhibitor. Thus, SmCPSKSL1 is bifunctional, catalyzing both class II bicyclization of GGPP and class I ionization of the resulting intermediate (Scheme 2). Notably, SmCPSKSL1 provides the first example of a class I diterpene synthase that catalyzes such direct replacement of the diphosphate with a hydroxyl, although others that incorporate water following cyclization are known.[16-19] In addition, the mass spectra of the minor products suggest that these (e.g., the apparent molecular ion peaks at *m/z* 272 or 290, respectively) might correspond to labdatriene(s) and labda-7,14-dien-13-ol(s) – albeit of unknown configuration. While these will be more precisely characterized in future studies, the production of such diterepenes is also consistent with class 1-mediated ionization of the postulated labda-7,13*E*-dienyl diphosphate intermediate produced by the class ll active site, as analogous compounds are produced by other class l diterpene cyclases[3, 4, 20, 21].

Labdanes can be produced as several different stereoisomers, which presumably result from folding of GGPP into alternative prochiral conformations prior to class II diterpene cyclase-catalyzed bicyclization[1]. While we cannot assign the absolute stereochemistry of the SmCPSKSL1 labda-7,13*E*-dien-15-ol product, the observed NOE between the hydrogen substituents on C20 and C11 demonstrates that these C10 and C9 substituents exhibit a *cis* configuration. This then implies a pro-chair-chair conformation of GGPP in the class II active site. Based on analogous chemical shifts relative to those observed with other labdane related diterpenes, as well as that of the previously characterized

natural product, we tentatively assign the SmCPSKLS1 product to the normal absolute configuration.

Regardless of absolute configuration, the production of an endocyclic double bond bicycle strongly indicates that SmCPSKSL1 catalyzes a unique class II reaction. Specifically, that the terminal labda-15-en-8-yl^{+} diphosphate intermediate is quenched by deprotonation at the neighboring C7 methylene. It is relatively easy to envision a concerted bicyclization reaction in the commonly catalyzed class II diterpene cyclase production of CPP, due to the evident ability to align the departing proton from the rotatable methyl group with the π orbitals of the relevant carbon-carbon double bonds (Scheme 1). Given the production of a C7 endocyclic double bond isomer observed here, the relative positioning of the C7 hydrogen substituents was examined. Notably, molecular modeling demonstrates that one of these hydrogens also can be closely aligned with the neighboring π orbital (e.g. C7α in the assumed "normal" configuration; Figure 2), thereby enabling a similarly concerted bicyclization reaction, and providing insight into the conformation of the GGPP substrate in this class II active site.

In conclusion, characterization of a bifunctional diterpene synthase from *S. moellendorffii* has revealed novel enzymatic activity. Specifically, this SmCPSKSL1 defines a new class II reaction, which further enabled some insight into GGPP substrate conformation, as well as providing the first example of a class I diterpene synthase carrying out direct hydrolysis of the allylic diphosphate ester bond. In addition, given the uncertain origins of the endocyclic double bond, along with that of the hydroxyl group, the biosynthesis of labda-7,13*E*-dien-15-ol was previously unclear, and potentially required several enzymatic steps. The results reported here demonstrate that labda-7,13*E*-dien-15-ol

can instead be produced by a single (albeit bifunctional) enzyme. This highlights such simple biosynthesis, only requiring changes in a single enzymatic gene, as a potential mechanism enabling the repeated convergent evolution that presumably then underlies the observed phylogenetically scattered production of the derived family of natural products.

Experimental Section

SmCPSKSL1 was cloned from *S. moellendorffii* maintained in a terrarium under a 14/10 hr light/dark cycle at 22 °C. Detached plant material was induced by using methyl jasmonate (5 mM) for 36 hours, RNA extracted by using a Plant RNA Purification kit (Invitrogen, Carlsbad CA), and cDNA produced with the Superscript system (Invitrogen). Primers were designed from putative bifunctional diterpene synthases (Selmol 120716 and 112927) found in the *S. mollendorffii* genome (genome.jgi-psf.org/Selmo1/Selmo1.home.html). SmCPSKSL1 was cloned using ATGAT AGAGG AAATG AGAAA ATTGC TTGC and TCATT CAGCT GCTTT ATACA ACACA TT as the forward and reverse primers, respectively. The PCR product was cloned into pENTR/SD/D-Topo (Invitrogen), completely sequenced, and deposited into Genbank (accession no. JN001323). SmCPSKSL1 was then transferred by directional recombination into pDEST15 and pDEST17 expression vectors.

SmCPSKSL1 activity was initially assessed by co-transforming *E. coli* strain C41 (Lucigen, Middleton, WI) with the pDEST15 construct and a previously described pGG vector.[13] This recombinant bacteria was grown in a 50 mL culture of TB media to a mid-log phase (OD_{600} ~0.6) at 37 °C, transferred to 16 °C for 1 hour prior to induction with 0.5 mM isopropylthiogalactoside (IPTG). Thereafter, they were fermented for an additional 72 h, and the culture then extracted with an equal volume of hexanes. This extract was dried under

a stream of N_2 and the residue redissolved in 200 μL of hexane for analysis by gas chromatography with mass spectrometry detection (GC-MS), using a Varian 3900 GC with Saturn 2100T ion-trap MS (Varian Inc., Palo Alto, CA), as previously described [20, 22-25].

A larger amount of the unrecognized diterpenoid was produced by additional transformation with a pIRS vector that increases the endogenous isoprenoid precursor supply.[14] The resulting bacterial was grown in 2 × 1 L cultures and extracted, as described above. The extract was dried by rotary evaporation, resuspended in 10 mL hexanes, and passed over a column of silica gel (10 mL), which was then eluted with ethyl acetate in hexnes (10 mL, 20 % v/v). This eluate was dried under a stream of N_2 and resuspended in methanol, and the diterpenoid purified via reverse-phase chromatography using a C_{18} column on an Agilent 1100 series high-performance system (Santa Clara, CA), much as previously described [20, 22-25].

Structural analysis by NMR also was carried out much as previously described [20, 22-25]. Following identification of the product as labda-7,13*E*-dien-15-ol, the chemical shift and MS data was compared to that previously reported for this compound, demonstrating good agreement. Thus, only the chemical shift data is reported here. ^{1}H(700.13 MHz) ^{1}H NMR (700.13 MHz, [D]chloroform): δ=0.919 (dt, J=3.6, 13.0 Hz, 1H; H1a), 1.818 (m, 1H; H1b), 1.414 (m, 1H; H2a), 1.510 (m, 1H; H2b), 1.132 (d, J=13.4 Hz, 1H; H3a), 1.384 (t,1H; H3b), 1.158 (m, 1H; H5), 1.840 (m, 1H; H6a), 1.945 (m, 1H;H6b), 5.369 (br s, 1H; H7), 1.591 (s, 1H; H9), 1.256 (m, 1H; H11a), 1.528 (m, 1H; H11b), 1.935 (m, 1H; H12a), 2.201 (dt, J=4.7, 12.4 Hz, 1H; H12b), 5.397 (t, J=6.8 Hz, 1H; H14), 4.134 (d, J=6.9 Hz, 2H; H15a-b), 1.669 (s, 3H; H16), 1.674 (s, 3H; H17), 0.830 (s,3H; H18 or H19), 0.852 (s, 3H; H18 or H19), 0.731 ppm (s, 3H;H20); 13C NMR (174.05 MHz, [D]chloroform): d=39.35 (C1),

19.00 (C2), 42.51 (C3), 33.17 (C4), 50.35 (C5), 24.02 (C6), 122.53 (C7), 135.46 (C8), 54.66 (C9), 37.00 (C10), 25.79 (C11), 42.28 (C12), 140.61(C13), 123.55 (C14), 59.63 (C15), 16.59 (C16), 22.40 (C17), 33.36 (C18or C19), 22.05 (C18 or C19), 13.75 ppm (C20). The cis configuration of the C9–10 substituents was assigned on the basis of the observed NOE correlations in a NOESY spectrum. NOE correlations for H11a and H11b (and their relative intensities observed), are H1b (s) and H20 (s). The ssigned E configuration at C13 was based on a strong NOE observed between H15 and H16.

In vitro assays with SmCPSKSL1 were carried out using the pDEST17 construct in C41 *E. coli*. The recombinant bacteria were grown in NZY liquid media to a mid-log phase (OD_{600}=0.6) at 37 °C, transferred to 16 °C for 1 hr prior to induction with IPTG to a final concentration of 0.5 mM, and fermented overnight (14–16 hrs). Cells were harvested by centrifugation and suspended in lysis buffer (10 mM Tris–Cl, pH 6.8, 10% glycerol, 1 mM dithiothreitol) for sonication. The resulting lysates were clarified via centrifugation (15 min, 15,000*g*). The recombinant protein was then purified over Ni-NTA Superflow resin (Novagen) following the manufacturers protocols. The purified enzyme was characterized in assay buffer (50 mM HEPES, pH 7.8, 100 mM KCl, 5 mM $MgCl_2$ and 10% glycerol) with GGPP (10 μM) as substrate, the purified enzyme (50 μL), and in some cases PhosStop phosphatase inhibitor (100 mM, Roche Diagnostics) in 1 mL assays. These were incubated 10 min at 30 °C, and the reaction stopped by addition of N-ethylmaleimide (110 μL, 0.2 M) for 5 min at room temperature, followed by the addition of dithiothreitol (10 μL). The production of labda-7,13*E*-dien-15-ol was verified by extraction with an equal volume of hexanes, which was dried under a stream of N_2, resuspended in hexanes (50 μL), and then analysis by GC-MS.

Acknowledgements

We thank Prof. Robert M. Coates for his assistance with the preparation of figure 2. The work was supported by grants from the NIH (GMO76324) and NSF (MCB0919735) to R.J.P.

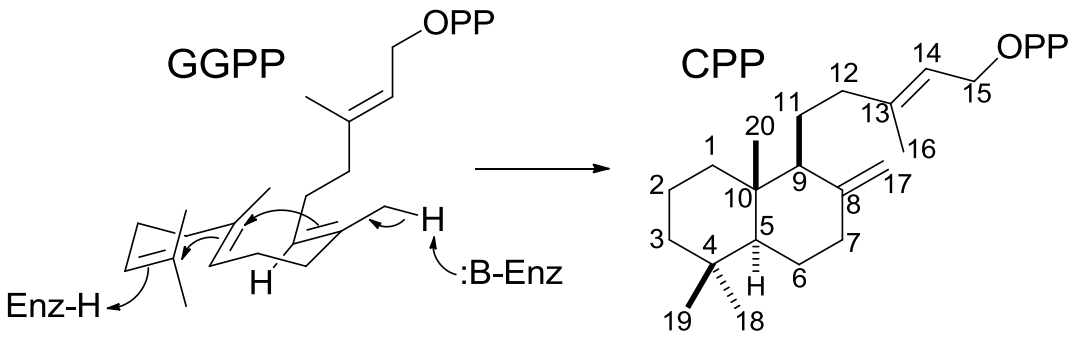
References

1. Peters RJ: **Two rings in them all: The labdane-related diterpenoids**. *Nat Prod Rep* 2010, **27**(11):1521-1530.
2. Falara V, Pichersky E, Kanellis AK: **A copal-8-ol diphosphate synthase from the angiosperm Cistus creticus subsp. creticus is a putative key enzyme for the formation of pharmacologically active, oxygen-containing labdane-type diterpenes**. *Plant Physiol* 2010, **154**(1):301-310.
3. Hamano Y, Kuzuyama Y, Itoh N, Furihata K, Seto H, Dairi T: **Functional analysis of eubacterial diterpene cyclases responsible for biosynthesis of a diterpene antibiotic, terpentecin**. *J Biol Chem* 2002, **277**(40):37098-37104.
4. Nakano C, Okamura T, Sato T, Dairi T, Hoshino T: **Mycobacterium tuberculosis H37Rv3377c encodes the diterpene cyclase for producing the halimane skeleton**. *Chem Commun (Camb)* 2005, **2005**:1016-1018.
5. Suzuki H, Noma M, Kawashima N: **Two labdane diterpenoids from *Nicotiana setchellii***. *Phytochemistry* 1983, **22**:1294-1295.
6. Demetzos C, Harvala C, Philianos SM, Skaltsounis AL: **A new labdane-type diterpene and other compounds from the leaves of *Cistus incanus* ssp. *creticus***. *J Nat Prod* 1990, **53**:1365-1368.
7. Rodilla JML, de Mendonca DIM, Urones JG, Moro RF: **Hydroxylated diterpenoids from *Halimium viscosum***. *Phytochemistry* 1998, **49**:817-822.
8. Feresin GE, Tapia A, Gimenez A, Ravelo AG, Zacchino S, Sortino M, Schmeda-Hirschmann G: **Constituents of the Argentinian medicinal plant *Baccharis grisebachii* and their antimicrobial activity**. *J Ethnopharmacol* 2003, **89**(1):73-80.
9. Aoki W, Ohtsuki T, Sadhu SK, Sato M, Koyano T, Preeprame S, Kowithayakorn T, Ishibashi M: **First isolation of three diterpenes as naturally-occurring compounds from *Sindora siamensis***. *J Nat Med* 2007, **61**:77-79.
10. Bevan CWL, Ekong DEU, Okogun JI: **West African timbers. Part XXII. The diterpenes of Oxystigma oxyphyllum Harms**. *J Chem Soc, C* 1968:1067-1070.
11. Rodriguez B: **New labdane diterpenoids from *Sideritis chamaedryfolia***. *Phytochemistry* 1978, **17**:281-286.
12. Chen F, Tholl D, Bohlmann J, Pichersky E: **The family of terpene synthases in plants: a mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom**. *Plant J* 2011, **66**(1):212-229.
13. Cyr A, Wilderman PR, Determan M, Peters RJ: **A Modular Approach for Facile Biosynthesis of Labdane-Related Diterpenes**. *J Am Chem Soc* 2007, **129**:6684-6685.

14. Morrone D, Lowry L, Determan MK, Hershey DM, Xu M, Peters RJ: **Increasing diterpene yield with a modular metabolic engineering system in E. coli: comparison of MEV and MEP isoprenoid precursor pathway engineering**. *Appl Microbiol Biotechnol* 2010, **85**:1893-1906.
15. Christianson DW: **Structural biology and chemistry of the terpenoid cyclases**. *Chem Rev* 2006, **106**:3412-3442.
16. Hayashi K, Kawaide H, Notomi M, Sakigi Y, Matsuo A, Nozaki H: **Identification and functional analysis of bifunctional ent-kaurene synthase from the moss Physcomitrella patens**. *FEBS Lett* 2006, **580**(26):6175-6181.
17. Toyomasu T, Niida R, Kenmoku H, Kanno Y, Miura S, Nakano C, Shiono Y, Mitsuhashi W, Toshima H, Oikawa H *et al*: **Identification of diterpene biosynthetic gene clusters and functional analysis of labdane-related diterpene cyclases in Phomopsis amygdali**. *Biosci Biotechnol Biochem* 2008, **72**(4):1038-1047.
18. Oikawa H, Toyomasu T, Toshima H, Ohashi S, Kawaide H, Kamiya Y, Ohtsuka M, Shinoda S, Mitsuhashi W, Sassa T: **Cloning and functional expression of cDNA encoding aphidicolan-16ß-ol synthase: A key enzyme responsible for formation of an unusual diterpene skeleton in biosynthesis of aphidicolin**. *J Am Chem Soc* 2001, **123**:5154-5155.
19. Herde M, Gartner K, Kollner TG, Fode B, Boland W, Gershenzon J, Gatz C, Tholl D: **Identification and regulation of TPS04/GES, an Arabidopsis geranyllinalool synthase catalyzing the first step in the formation of the insect-induced volatile C16-homoterpene TMTT**. *Plant Cell* 2008, **20**(4):1152-1168.
20. Morrone D, Hillwig ML, Mead ME, Lowry L, Fulton DB, Peters RJ: **Evident and latent plasticity across the rice diterpene synthase family with potential implications for the evolution of diterpenoid metabolism in the cereals**. *Biochem J* 2011, **435**(3):589-595.
21. Maugel N, Mann FM, Hillwig ML, Peters RJ, Snider BB: **Synthesis of (+/-)-nosyberkol (isotuberculosinol, revised structure of edaxadiene) and (+/-)-tuberculosinol**. *Org Lett* 2010, **12**(11):2626-2629.
22. Morrone D, Xu M, Fulton DB, Determan MK, Peters RJ: **Increasing complexity of a diterpene synthase reaction with a single residue switch**. *J Am Chem Soc* 2008, **130**(16):5400-5401.
23. Swaminathan S, Morrone D, Wang Q, Fulton DB, Peters RJ: **CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice**. *Plant Cell* 2009, **21**(10):3315-3325.
24. Gao W, Hillwig ML, Huang L, Cui G, Wang X, Kong J, Yang B, Peters RJ: **A functional genomics approach to tanshinone biosynthesis provides stereochemical insights**. *Org Lett* 2009, **11**(22):5170-5173.
25. Wang Q, Hillwig ML, Peters RJ: **CYP99A3: functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice**. *Plant J* 2011, **65**(1):87-95.

Figures

Scheme 1: Cyclization of GGPP to CPP.



Scheme 2: Reactions catalyzed by SmCPSKSL to produce labda-7,13*E*-dien-15-ol from GGPP.

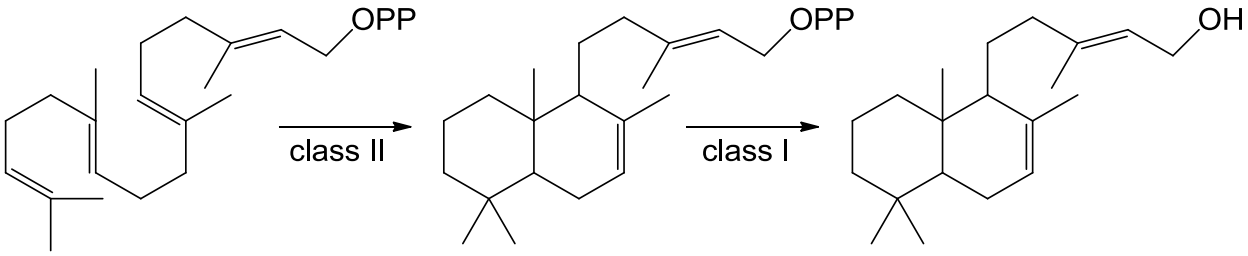


Figure 1: SmCPSKSL1 produces labda-7,13*E*-dien-15-ol (peak 1). A) GC-MS chromatogram of extract from *E. coli* engineered to co-express SmCPSKSL1 with a GGPP synthase. B) MS of peak 1 (labda-7,13*E*-dien-15-ol); the inset clearly depicts the molecular ion peak at *m/z* 290

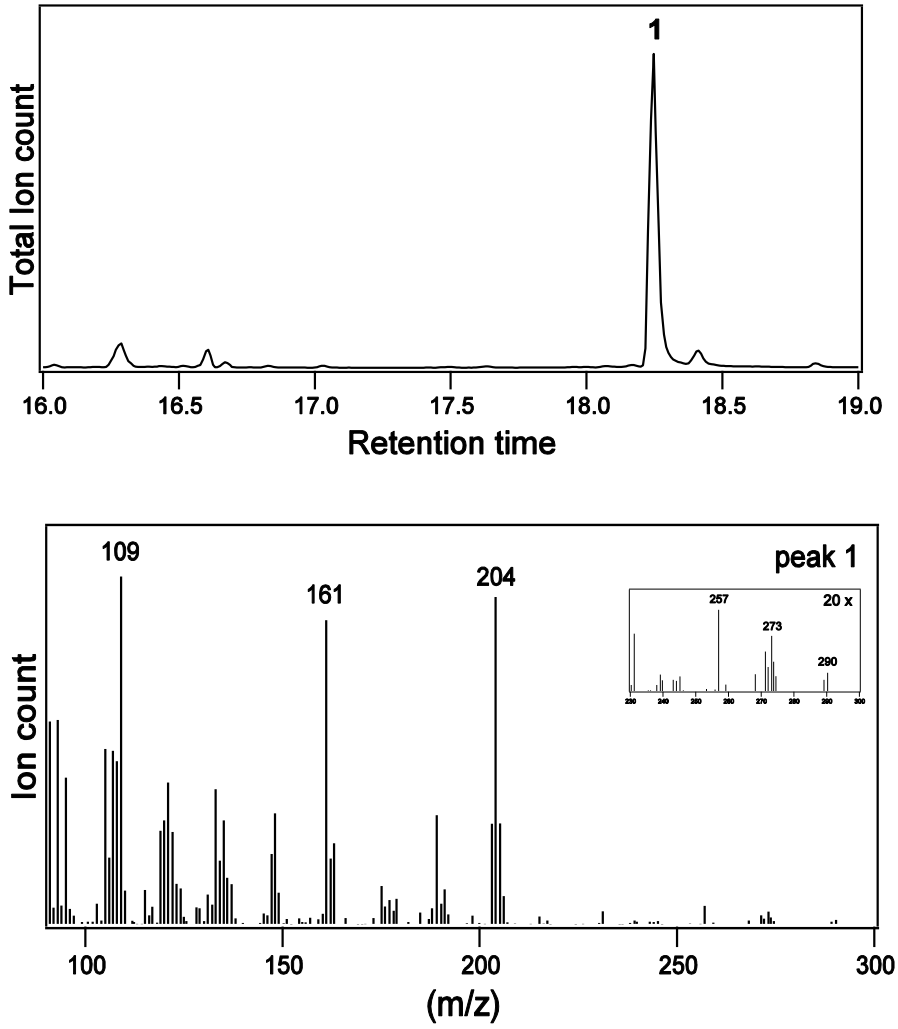
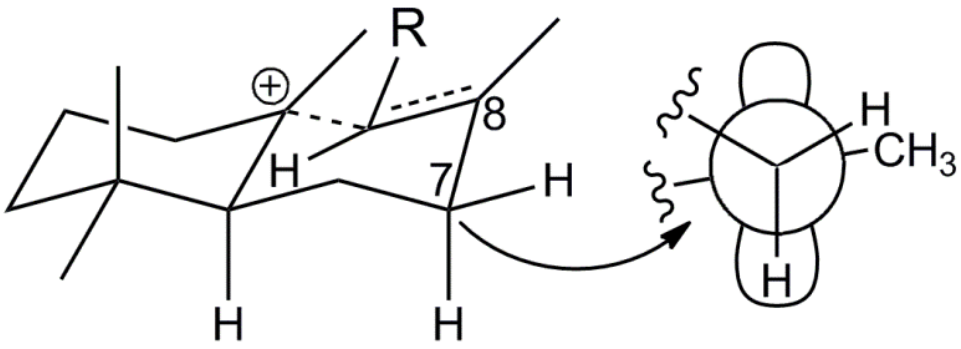


Figure 2: Relative configuration of C7 hydrogen substituents with π orbital of the sp^3 hybridized C8 in SmCPSKSL1 catalyzed class II bicyclization. A putative moncyclic intermediate (R = allylic isoprenyl diphosphate "tail") is depicted, with Newman projection along the C7-C8 bond.



CHAPTER III: EVOLUTIONARY CONTEXT DEPENDENT PLASTICITY OF *ENT*-KAURENE OXIDASES

Sibongile Mafu, Jiachen Zi, Dana Morrone, Yisheng Wu, Matthew L. Hillwig

and Reuben J. Peters

Abstract

All higher plants contain kaurene oxidases (KO), which are multifunctional cytochromes P450 that catalyze oxidation at the C4α methyl, converting *ent*-kaurene to *ent*-kaurenoic acid, an early step in gibberellin phytohormone biosynthesis. Arabidopsis produces no labdane-related diterpenoids other than gibberellins, whereas rice produces a wide range of such natural products as antibiotic phytoalexins or allelochemicals. While rice contains several kaurene oxidase homologs, only OsKO2 (CYP701A6) is required for gibberellin biosynthesis. Here, we demonstrate that the KO from *Arabidopsis thaliana* (CYP701A3 or AtKO) exhibits significantly greater promiscuity than does OsKO2. Specifically, when both were screened against a wide range of labdane-related diterpene olefins (using a modular metabolic engineering system in *E. coli*), AtKO was able to react with a much wider range of these potential substrates than was OsKO2, which seemed to be quite specific for *ent*-kaurene and closely related substances. To further characterize this plasticity of AtKO, we determined the structure of the resulting products, whereas OsKO2 only hydroxylates its substrates on C19, AtKO reacts with labdane-related diterpenes of varied stereochemistry, which further leads to altered hydroxylation regiochemistry. Our results demonstrate greater promiscuity of AtKO, which evolved in the absence of other potential labdane-related diterpene substrates, relative to OsKO2 from rice, which contains many such alternative substrates. Thus, our

data are consistent with the hypothesis that enzymatic plasticity is shaped, at least in part, by evolutionary context such as that noted here.

Introduction

Giberellins are essential for plant growth and development. The genes encoding for enzymes of the biosynthetic pathway are present in all land plants. As such all plants contain kaurene oxidases (KO), which are multifunctional cytochromes P450 that catalyze oxidation at the C4α methyl, converting *ent*-kaurene to *ent*-kaurenoic acid, an early step in GA biosynthesis with related KO genes exhibiting neofunctionalization. The biochemical selectivity of these KOs is shaped in part by its biochemical evolutionary background.

In brief, the GA biosynthetic pathway is initiated by the proton-initiated cyclization of geranyl geranyl diphosphate (*E*,*E*,*E*-GGPP) into copalyl diphosphate (CPP) through a labda-13-en-8yl^{+} intermediate whose stereochemistry is dependent on the prochiral conformation of the GGPP precursor to give normal (9S, 10S), syn (9S, 10R) or ent (9R, 10R) (Figure 1). Carbocationic cyclization is catalyzed by class I cyclase kaurene synthase (KS), resulting in kaurene. The intermediate olefin is acted upon by membrane bound cytochromes P450, KO to form *ent*-kaurenoic acid and conversion to the 6-5-6-5 complex by kaurenoic acid oxidase (KAO), the committed step in GA biosynthesis [2].

Variation in the labdane related diterpenoids arises from the evolutionary duplication and sub/neofunctonalization of TPS genes to give rise to an array of CPS and KS-like (KSL) genes to produce a range of precursors of labdane related secondary compounds (Figure 3) [2-9]. A similar theme of functional diversification originating from hormone biosynthesis has been recently demonstrated in the CYP701A subfamily [10]. KO was initially characterized in cell free extracts of *Echinocystis macrocarpa* [11, 12]. *Arabidopsis thaliana*,

contains a single kaurene oxidase homolog CYP701A3 (AtKO) [13] whereas rice (*Oryza sativa*) contains multiple kaurene oxidase paralogs of which CYP701A6/OsKO2 is involved in kaurenoic acid biosynthesis and CYP701A8/OsKOL4 has undergone neofunctionalzation with the ability to hydroxylate *ent*-kaurene, *ent*-sandaracopimaradiene and *ent*-cassadiene at the C3α position [10, 14, 15] most likely intermediates in phytoalexins biosynthesis. Rice is a great model system for LRD biosynthesis as it produces ~20 known labdane related phytoalexins [16] showing both diversification in the upstream TPS genes as well as the downstream CYPs. In contrast *Arabidopsis* produces only GA as an LRD.

KO in *E. macrocarpa* was shown to be strictly specific for kaurene, unable to even catalyze *ent*-isokaurene which differs in a single double bond [17], however, more recently AtKO was shown to be able to catalyze other labdane related compounds [18]. This raises interesting question about the relative plasticity of AtKO. Utilizing two biochemically distinct model systems, rice which produces GA and 20 other LRDs and *Arabidopsis* which produces GA as the only LRD, gives a natural comparative system to demonstrate the influence of biochemical backgrounds on KO evolutionary based selectivity. Using our modular metabolic engineering system, we demonstrate that AtKO/CYP701A3 has the ability to recognize substrates of differing stereochemistry and oxidize at multiple stereo and regiochemical positions, a function which is specialized to specific cytochromes P450 in rice. Analysis of the resultant products gives us an insight into the substrate properties required to ensure catalysis by KO. The relative plasticity of AtKO is explored further in this report to understand the substrate structure-activity relationship underlying AtKO catalysis.

Experimental Procedures

Unless otherwise noted, chemicals were purchased from Fisher Scientific and molecular biology reagents from Invitrogen. Gas chromatography (GC) was performed with a Varian 3900 GC with Saturn 2100 ion trap mass spectrometer (MS) in electron ionization (70 eV) mode. Samples (1 μL) were injected in splitless mode at 50^{o}C and, after holding for 3min at 50^{o}C, the oven temperature was raised at a rate of 15^{o}C/min to 300^{o}C, where it was held for an additional 3 min. MS data from 90 to 600 mass-to-charge ratio (m/z) were collected starting 14 min after injection until the end of the run.

Co-expression

Plasticity of AtKO and OsKO2 was assessed by use of our previously described modular metabolic engineering system [19]. Utilizing this system we were able to co-express codon optimized, N-terminally modified CYPs with AtR1 on a PCDF-Duet vector [18, 20, 21] with a class 1 labdane related diterpene synthases such as OsKSL from pET based pDEST expression vectors with GGPP and CPS carried on a compatible pACYC-Duet (Novagen/EMD) derived plasmid, where pGG*e*C carries the AtCPS, *ent*-CPS from *A. thaliana* , pGG*s*C carries the OsCPS4 *syn*-CPS rom rice [22] and pGGnC carries the D621A mutant from *Abies grandis* that produces normal CPP [23]. CYP reactivity was assessed by co-transforming *E. coli* strain C41 (Lucigen, Middleton, WI) with the respective KSL and the corresponding pGGxC vector. This recombinant bacterium was grown in a culture of TB medium (20 mL) to a mid-log phase (OD_{600} ~0.8) at 37^{o}C, then at 16^{o}C for 1 h prior to induction with isopropylthiogalactoside (IPTG, 0.5 mM) and addition of aminolevulinic acid (75mg/ml) and riboflavin (2.0 mg/ml). Thereafter, it was grown for an additional 72 hours,

and the culture then extracted with an equal volume of hexanes. The extract was dried under a stream of N_2, and derivatized with diazomethane to examine potential acid formation and re-dissolved in hexane (100 µL) for analysis by gas chromatography with mass spectrometry detection (GC-MS) as previously described [21].

Diterpenoid production

The novel enzymatic products were obtained in sufficient amounts for NMR analysis by both increasing flux into isoprenoid metabolism and scaling up the culture volumes. Flux toward isoprenoid biosynthesis was increased by using the previously described pMBI which encodes for the bottom half of the mevalonate-dependent isoprenoid precursor pathway from yeast *Saccharomyces cerevisiae* [24]. This enables production of the isoprenoid precursors isopentenyl diphosphate and dimethylallyl diphosphate from mevalonate and feeding of 10 mM mevalonolactone significantly increases diterpenoid production, as previously described [25]. The resulting diterpenoids were extracted from 4 L of culture (media and cells) with an equal volume of hexanes, and the organic extract then dried by rotary evaporation. The resulting residue was resuspended in hexanes and fractionated over a 4 g-silica column by flash chromatography, using a Reveleris system (Grace, Deerfield, IL), with a hexane:acetone gradient. The hydroxylated diterpenoids eluted between 10-20 % acetone, and these fractions were combined and dried under N_2. The resulting residue was dissolved in acetonitrile and the hydroxylated diterpenoids purified by HPLC. This was carried out using an Agilent 1200 series instrument equipped with autosampler, fraction collector, and diode array UV detection, over a ZORBAX Eclipse XDB-C8 column (4.6 3 150 mm, 5 mm) at a 0.5 mL/min flow rate. The column was pre-equilibrated with 50% acetonitrile/distilled water, sample loaded, then the column washed with 50% acetonitrile/distilled water (0–2

min), and eluted with 50% to 100% acetonitrile (2–7 min), followed by a 100% acetonitrile wash (7–30 min). Following purification, each compound was dried under a gentle stream of N2, and then dissolved in 0.5 mL deuterated chloroform (CD3OD; Sigma-Aldrich), with this evaporation-resuspension process repeated two more times to completely remove the protonated acetonitrile solvent.

Chemical Structure Identification

NMR spectra for the diterpenoids were recorded at 25^{o}C on a Bruker Avance 700 spectrometer equipped with a cryogenic probe for 1H and 13C. Structural analysis was performed using 1 D 1H, 1 D DQF-COSY, HSQC, HMQC, HMBC, and NOESY experiment spectra acquired at 700 MHz and 13C (175 MHz) and DEPT135 spectra using standard experiments from the Bruker TopSpin v1.3 software. All samples were placed in NMR tubes purged with nitrogen gas for analyses, and chemical shifts were referenced using known methanol [13C 49.15 (7), 1H 3.31(5) ppm (m)] signals offset from tetramethylsilane. Correlations from the HMBC spectra were used to propose a partial structure, while COSY correlations between protonated carbons and HSQC spectra were used to complete the partial structure and assign proton chemical shifts. The configuration of the A and B rings (C1–C10) is predetermined by the configuration of the CPP intermediate, since chemical bonds in that portion of the molecule are not altered. Thus, nuclear Overhauser effect dipole-dipole signals observed could be used to assign configuration of the hydroxyl group.

Results

Substrate specificity screening

Based on the reported increased substrate range of AtKO [18] we were interested in a comparative investigation of substrate specificity of KOs from *Arabidopsis* and rice, which have evolved under different chemical backgrounds. To compare potential substrate selectivity of AtKO and OsKO2, we utilized our bacterial modular metabolic engineering system to screen against a range of olefins [19]. Specifically we co-expressed an N–terminally modified P450 [10, 18, 20] with AtR1 a reductase from *A. thaliana* [26], with corresponding upstream CPS and KS(L) genes from rice, castor bean and bacteria amongst others, to produce an array of labdane related diterpene (Figure 3, Table S1). Both AtKO and OsKO2 coexpressed with kaurene catalyzed turnover to the acid (Figure S1). AtKO was able to react with a much wider range of the tested labdane-related diterpene substrates than was OsKO2, which seemed to be largely specific for *ent*-kaurene, *ent*-isokaurene and trachylobane, all closely related structures of *ent*-stereochemistry. AtKO is more promiscuous, producing acids with substrates of *ent*- stereochemistry, although only alcohols with products of *syn* (9S, 10R) and normal (9R, 10R) stereochemistry (Figure 4).

Alternate substrate product identification

To determine the stereochemistry and/or regiochemistry of reactivity of the resultant products, we carried out structural characterization of a subset of products by NMR. To obtain sufficient compound, we increased the yield of our co-expression by utilizing pMBI

which expresses the downstream mevalonate pathway [25]. While OsKO2 only hydroxylates its substrates on C19, reflecting their close relationship to *ent*-kaurene (i.e., these are all compounds of 9*R*, 10R (*ent*) configuration), AtKO reacts with labdane-related diterpenes of varied stereochemistry, which further leads to altered hydroxylation regiochemistry. Specifically AtKO catalyzes C19 hydroxylation of *ent*-cassadiene and *ent*-sandaracopimaradiene, where kaurene is mostly turned over to kaurenoic acid, *ent*-cassadiene and *ent*-sandaracopimaradiene have a mixture of the acid and alcohol (Figures 5 and S2). AtKO catalyzes substrates of 9*S*, 10R (*syn*) stereochemistry. Catalysis of *syn*- pimara-7,15-diene results in 2α-Hydroxy-*syn*-pimara-7,15-diene and 3β-Hydroxy-*syn*-pimara-7,15-diene hydroxylation; but *syn*-aphidocolene is hydroxylated only at C3β position resulting in 3β-Hydroxy-*syn*-aphidocolene. The smaller hydroxyl peaks on the *syn*-aphidocolene spectra correspond to hydroxylation products of syn-pimara-7,15-diene (Figures 5and S3) [27] . The reaction with normal-pimara-7, 15-diene results in 2α-Hydroxy-normal-pimara-7,15-diene, a fraction of this is oxidized to a ketone. Normal-syn-pimara-8,15-diene is converted to 3β-Hydroxy-*syn*-pimara-7, 15-diene. While OsKO2 only hydroxylates its substrates on C19, reflecting their close relationship and similarity to *ent*-kaurene, AtKO reacts with different cyclic labdane-related diterpenes of varied stereochemistry, which further leads to altered hydroxylation regiochemistry (Figures 5 and S4).

Discussion

Biochemical promiscuity demonstrated by AtKO represents latent substrate capacity retention relative to OsKO2. Whereas AtKO will hydroxylate a surprisingly wide range of labdane-related diterpenes (11 of the 18 tested), OsKO2 only reacts with *ent*-kaurene and a few other very closely related compounds *ent*-isokaurene and *ent*-trachylobane (which

corresponds to a putative intermediate in kaurene cyclization) [28]. Moreover, while OsKO2 only hydroxylates its substrates on C19, reflecting their close structural similarity to *ent*-kaurene, AtKO reacts with labdane-related diterpenes of varied structure and stereochemistry, which further leads to altered hydroxylation regiochemistry. *Arabidopsis* produces GA as the only LRD. In contrast, rice has several LRD related phytoalexins. The wider range of LRDs in rice conceivably imposed selective pressure on OsKO2 to be more specific for kaurene and closely related substrates as was seen in *E. macrocarpa* to ensure restrictedness to phytohormone biosynthesis [17]. In the absence of LRDs in *Arabidopsis*, there was minimal selective pressure for specificity. AtKO promiscuity is analogous to that of enzymes involved in specialized metabolism can recognize more than one substrate but have different affinity towards them such as CYP720B4 which accepts a range of structurally related labdane olefins with varying affinities [29]. The inherent latent plasticity of AtKO is comparable to CYP76M8 which has the capacity to hydroxylate a range of diterpenes including those not found in rice [21].

We were able to further understand the biochemical plasticity of AtKO by reacting it with selected LRDs of *ent*, *syn* and normal stereochemistry. The multifunctional nature of the AtKO is limited to substrates of *ent*-stereochemistry, and only an initial hydroxylation occurs in substrates of *syn* and normal stereochemistry with an associated change in regiochemistry. In previous studies, AtKO converted atiserene to atiserenoic acid, whereas beyerene was converted only to the corresponding alcohol in the same regiochemical position. Comparison of the chemical structures with that of kaurene demonstrates that the this incomplete conversion may be a result of the alteration in the C/D region where changes in C11-14 plays a role in preventing complete hydroxylation to the acid and intermediates (alcohol and

aldehyde) as shown in Figure 5. This is similar to observation AtKO catalyzed conversion of beyerene to beyeren-19-ol where a change in the distal region of the C/D ring was sufficient to prevent additional reactivity [18]. Further feeding studies need to be carried out with 3α-Hydroxy-kaurene (reported to be a competitive inhibitor) as this would further highlight the role of the distal region in the multifunctional nature of KO [17]. For *syn* and normal substrates there is a change in regiochemistry with catalysis occurring at the C2α or the C3β position. This may be largely attributed to the C10 methyl group which is in a beta position for substrates of *syn* and normal compared to the alpha position in *ent*-substrates. Overlay structures utilizing C6, C10 and C19 as atom properties indicate variation in the C/D ring (Figure S5). The C3α position relative to the typical C19 position is in proximity of each other as demonstrated by three dimensional rendering [10]. CYPs are characterized by cofactor heme which is deep within the binding cavity, orientation of the substrate relative to the heme has been suggested to determine regioselectivity as is seen here [30].

In summary, our data are consistent with the hypothesis that enzymatic plasticity is shaped, at least in part, by evolutionary context of enzymes. In addition it provides insight into the substrate structure-activity relationship underlying AtKO catalysis of multiple substrates. The data suggest the minimum requirement of *ent*-stereochemistry and high structural similarity to kaurene for full multifunctional catalysis by AtKO; distal changes in the C/D ring disrupt catalysis resulting in increased formation of intermediates for substrates of *ent*-stereochemistry. Tri and tetra-cycles of *syn* and normal stereochemistry are oriented in a manner that presents the C2 and C3 for catalysis resulting in change in regio and stereo-chemical catalysis. The work described here demonstrates that multifunctional cytochrome P450, AtKO, which in addition to its ability to catalyze the native substrate (kaurene) has

latent substrate capacity, and is able to catalyze multiple substrates with varying degrees of regio and stereo specificity relative to its native substrate.

References

1. Hedden P, Thomas SG: **Gibberellin biosynthesis and its regulation**. *Biochem J* 2012, **444**(1):11-25.
2. Peters RJ: **Two rings in them all: the labdane-related diterpenoids**. *Nat Prod Rep* 2010, **27**(11):1521-1530.
3. Falara V, Pichersky E, Kanellis AK: **A copal-8-ol diphosphate synthase from the angiosperm Cistus creticus subsp. creticus is a putative key enzyme for the formation of pharmacologically active, oxygen-containing labdane-type diterpenes**. *Plant Physiol* 2010, **154**(1):301-310.
4. Mafu S, Hillwig ML, Peters RJ: **A novel labda-7,13e-dien-15-ol-producing bifunctional diterpene synthase from Selaginella moellendorffii**. *Chembiochem* 2011, **12**(13):1984-1987.
5. Prisic S, Xu M, Wilderman PR, Peters RJ: **Rice contains two disparate ent-copalyl diphosphate synthases with distinct metabolic functions**. *Plant Physiol* 2004, **136**(4):4228-4236.
6. Wilderman PR, Xu M, Jin Y, Coates RM, Peters RJ: **Identification of syn-pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis**. *Plant Physiol* 2004, **135**(4):2098-2105.
7. Wu Y, Zhou K, Toyomasu T, Sugawara C, Oku M, Abe S, Usui M, Mitsuhashi W, Chono M, Chandler PM *et al*: **Functional characterization of wheat copalyl diphosphate synthases sheds light on the early evolution of labdane-related diterpenoid metabolism in the cereals**. *Phytochemistry* 2012, **84**:40-46.
8. Cho EM, Okada A, Kenmoku H, Otomo K, Toyomasu T, Mitsuhashi W, Sassa T, Yajima A, Yabuta G, Mori K *et al*: **Molecular cloning and characterization of a cDNA encoding ent-cassa-12,15-diene synthase, a putative diterpenoid phytoalexin biosynthetic enzyme, from suspension-cultured rice cells treated with a chitin elicitor**. *Plant J* 2004, **37**(1):1-8.
9. Nemoto T, Cho EM, Okada A, Okada K, Otomo K, Kanno Y, Toyomasu T, Mitsuhashi W, Sassa T, Minami E *et al*: **Stemar-13-ene synthase, a diterpene cyclase involved in the biosynthesis of the phytoalexin oryzalexin S in rice**. *FEBS Lett* 2004, **571**(1-3):182-186.
10. Wang Q, Hillwig ML, Wu Y, Peters RJ: **CYP701A8: a rice ent-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism**. *Plant Physiol* 2012, **158**(3):1418-1425.
11. Dennis DT, West CA: **Biosynthesis of gibberellins. 3. The conversion of (-)-kaurene to (-)-kauren-19-oic acid in endosperm of Echinocystis macrocarpa Greene**. *J Biol Chem* 1967, **242**(14):3293-3300.

12. GRAEBE JE, DENNIS DT, UPPER CD, WEST CA: **BIOSYNTHESIS OF GIBBERELLINS. I. THE BIOSYNTHESIS OF (-)-KAUREN-19-OL, AND TRANS-GERANYLGERANIOL IN ENDOSPERM NUCELLUS OF ECHINOCYSTIS MACROCARPA GREENE**. *J Biol Chem* 1965, **240**:1847-1854.
13. Helliwell CA, Poole A, Peacock WJ, Dennis ES: **Arabidopsis ent-kaurene oxidase catalyzes three steps of gibberellin biosynthesis**. *Plant Physiol* 1999, **119**(2):507-510.
14. Sakamoto T, Miura K, Itoh H, Tatsumi T, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Agrawal GK, Takeda S, Abe K *et al*: **An overview of gibberellin metabolism enzyme genes and their related mutants in rice**. *Plant Physiol* 2004, **134**(4):1642-1653.
15. Ko KW, Lin F, Katsumata T, Sugai Y, Miyazaki S, Kawaide H, Okada K, Nojiri H, Yamane H: **Functional identification of a rice ent-kaurene oxidase, OsKO2, using the Pichia pastoris expression system**. *Biosci Biotechnol Biochem* 2008, **72**(12):3285-3288.
16. Peters RJ: **Uncovering the complex metabolic network underlying diterpenoid phytoalexin biosynthesis in rice and other cereal crop plants**. *Phytochemistry* 2006, **67**(21):2307-2317.
17. West CA: **Hydroxylases, monooxygenases, and cytochrome P-450**. New York: Academic Press; 1980.
18. Morrone D, Chen X, Coates RM, Peters RJ: **Characterization of the kaurene oxidase CYP701A3, a multifunctional cytochrome P450 from gibberellin biosynthesis**. *Biochem J* 2010, **431**(3):337-344.
19. Cyr A, Wilderman PR, Determan M, Peters RJ: **A modular approach for facile biosynthesis of labdane-related diterpenes**. *J Am Chem Soc* 2007, **129**(21):6684-6685.
20. Swaminathan S, Morrone D, Wang Q, Fulton DB, Peters RJ: **CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice**. *Plant Cell* 2009, **21**(10):3315-3325.
21. Wang Q, Hillwig ML, Okada K, Yamazaki K, Wu Y, Swaminathan S, Yamane H, Peters RJ: **Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster**. *J Biol Chem* 2012, **287**(9):6159-6168.
22. Xu M, Hillwig ML, Prisic S, Coates RM, Peters RJ: **Functional identification of rice syn-copalyl diphosphate synthase and its role in initiating biosynthesis of diterpenoid phytoalexin/allelopathic natural products**. *Plant J* 2004, **39**(3):309-318.
23. Peters RJ, Flory JE, Jetter R, Ravn MM, Lee HJ, Coates RM, Croteau RB: **Abietadiene synthase from grand fir (Abies grandis): characterization and mechanism of action of the "pseudomature" recombinant enzyme**. *Biochemistry* 2000, **39**(50):15592-15602.
24. Martin VJ, Pitera DJ, Withers ST, Newman JD, Keasling JD: **Engineering a mevalonate pathway in Escherichia coli for production of terpenoids**. *Nat Biotechnol* 2003, **21**(7):796-802.
25. Morrone D, Lowry L, Determan MK, Hershey DM, Xu M, Peters RJ: **Increasing diterpene yield with a modular metabolic engineering system in E. coli:**

comparison of MEV and MEP isoprenoid precursor pathway engineering. *Appl Microbiol Biotechnol* 2010, **85**(6):1893-1906.

26. Urban P, Mignotte C, Kazmaier M, Delorme F, Pompon D: **Cloning, yeast expression, and characterization of the coupling of two distantly related Arabidopsis thaliana NADPH-cytochrome P450 reductases with P450 CYP73A5**. *J Biol Chem* 1997, **272**(31):19176-19186.
27. Morrone D, Xu M, Fulton DB, Determan MK, Peters RJ: **Increasing complexity of a diterpene synthase reaction with a single residue switch**. *J Am Chem Soc* 2008, **130**(16):5400-5401.
28. J M, M B: **Diterepen Biosynthesis**. Comprehensive natural Products Chemistry: Pergamon Press; 2000.
29. Hamberger B, Ohnishi T, Séguin A, Bohlmann J: **Evolution of diterpene metabolism: Sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects**. *Plant Physiol* 2011, **157**(4):1677-1695.
30. Seifert A, Tatzel S, Schmid RD, Pleiss J: **Multiple molecular dynamics simulations of human p450 monooxygenase CYP2C9: the molecular basis of substrate binding and regioselectivity toward warfarin**. *Proteins* 2006, **64**(1):147-155.

Figures

Figure 1: Cyclization of geranyl geranyl diphosphate (GGPP) to ent, syn and normal stereochemistries of copalyl diphosphate (CPP)

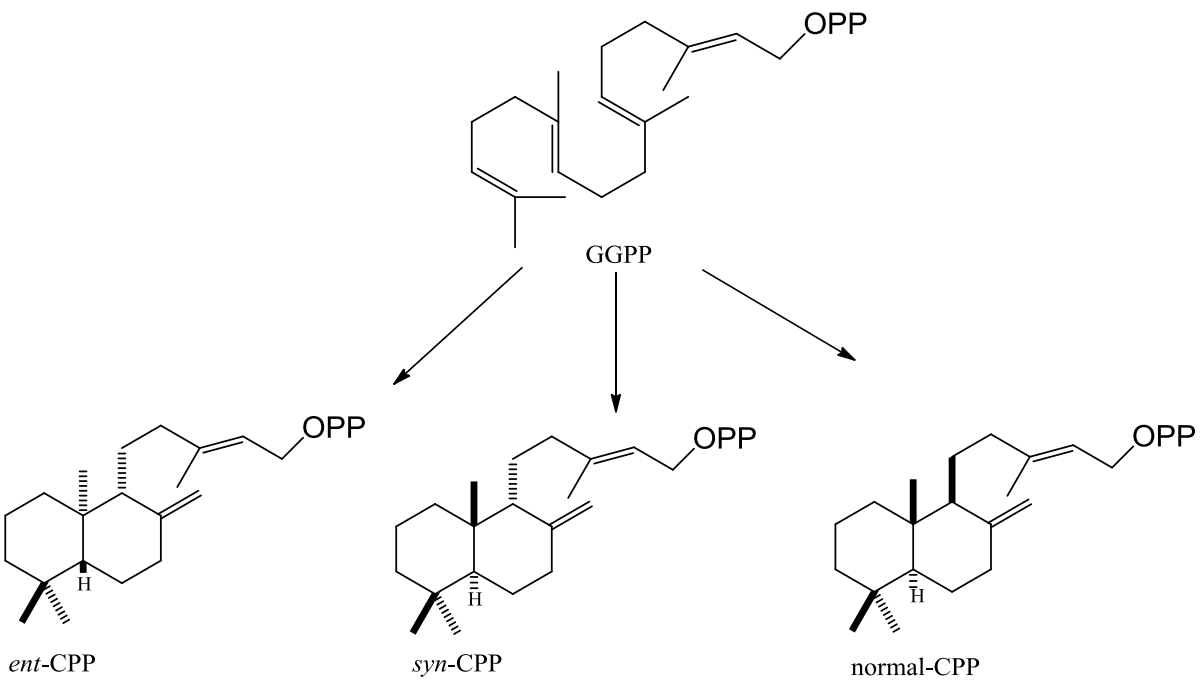
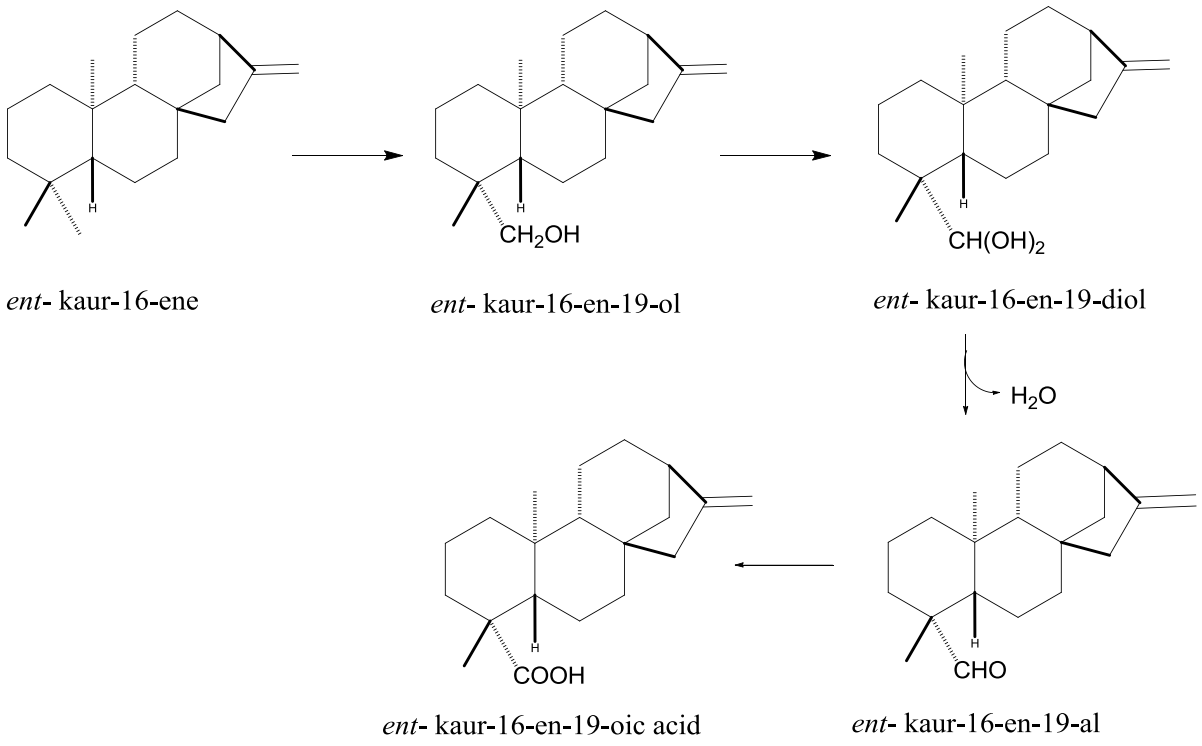


Figure 2: Conversion of *ent*-kaurene to *ent*-kaurenoic acid. (After Morrone et al. 2010)



ent- kaur-16-ene

ent- kaur-16-en-19-ol

ent- kaur-16-en-19-diol

ent- kaur-16-en-19-oic acid

ent- kaur-16-en-19-al

Figure 3: Range of labdane related diterpenes reacted with AtKO and OsKO

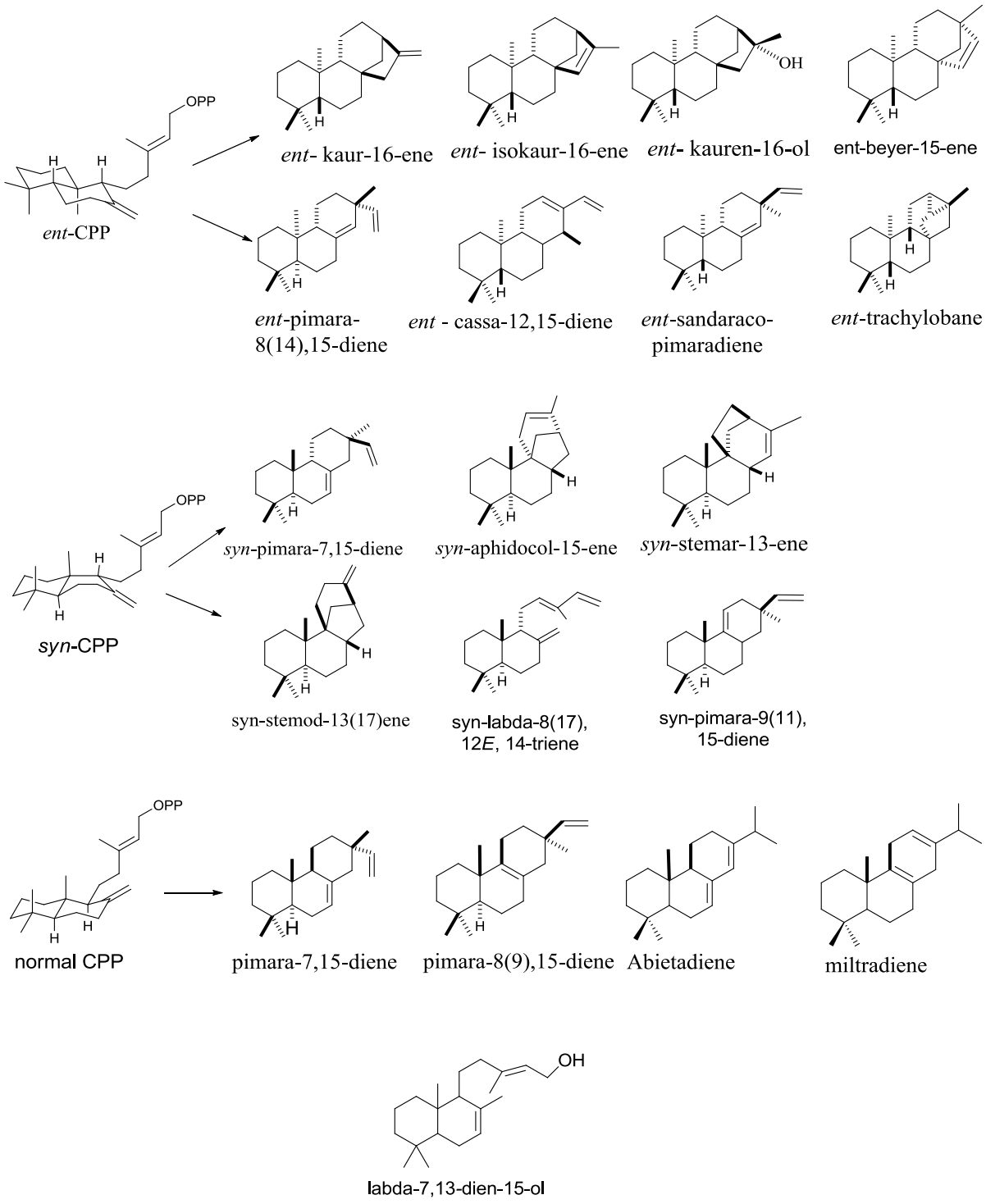
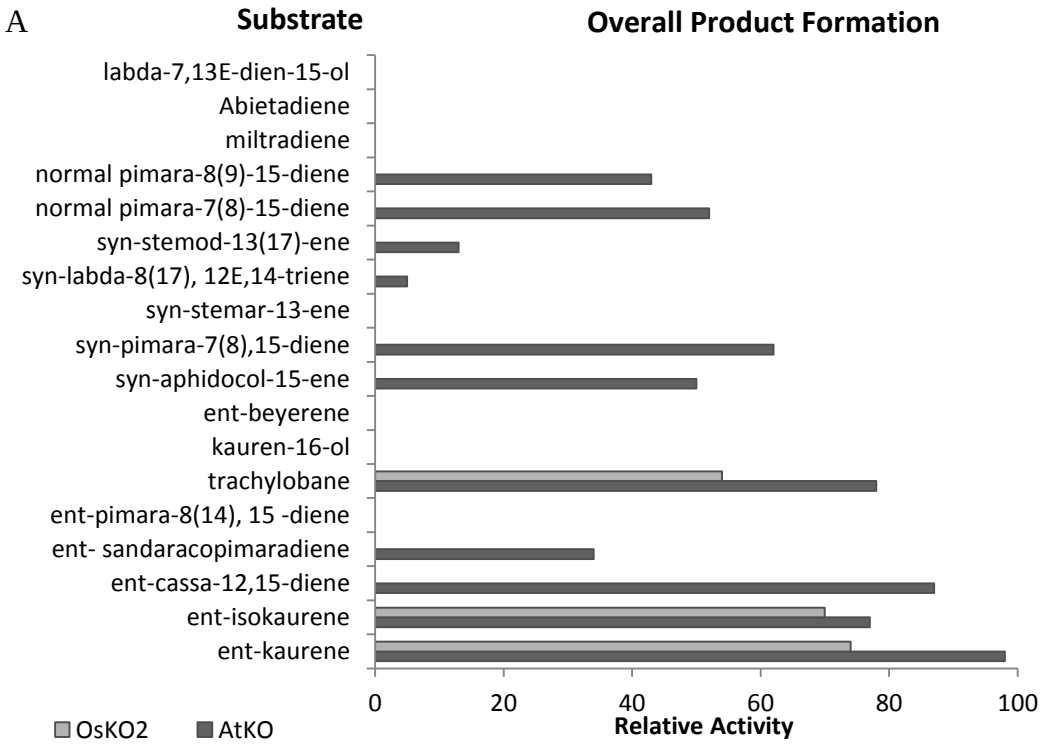
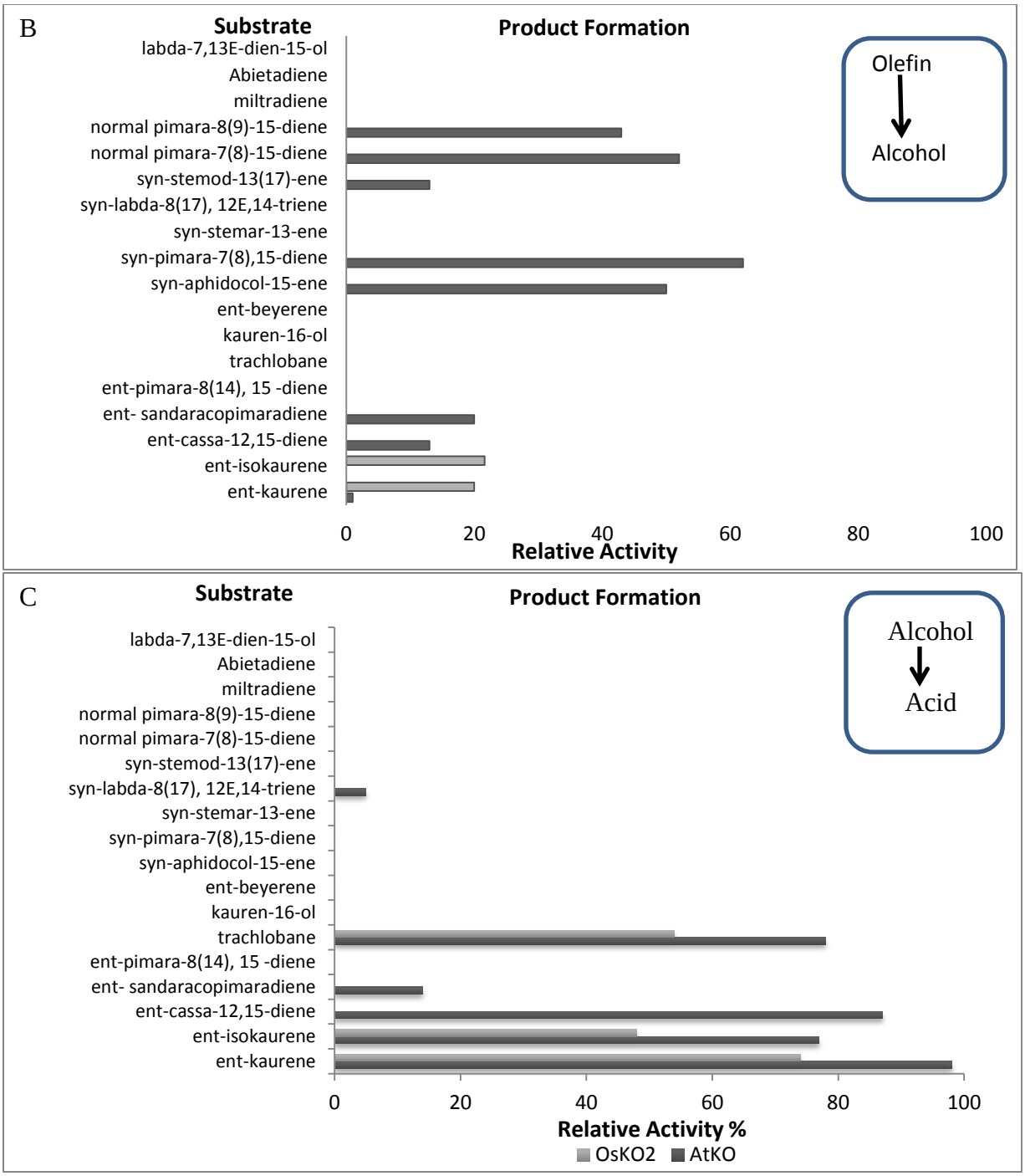


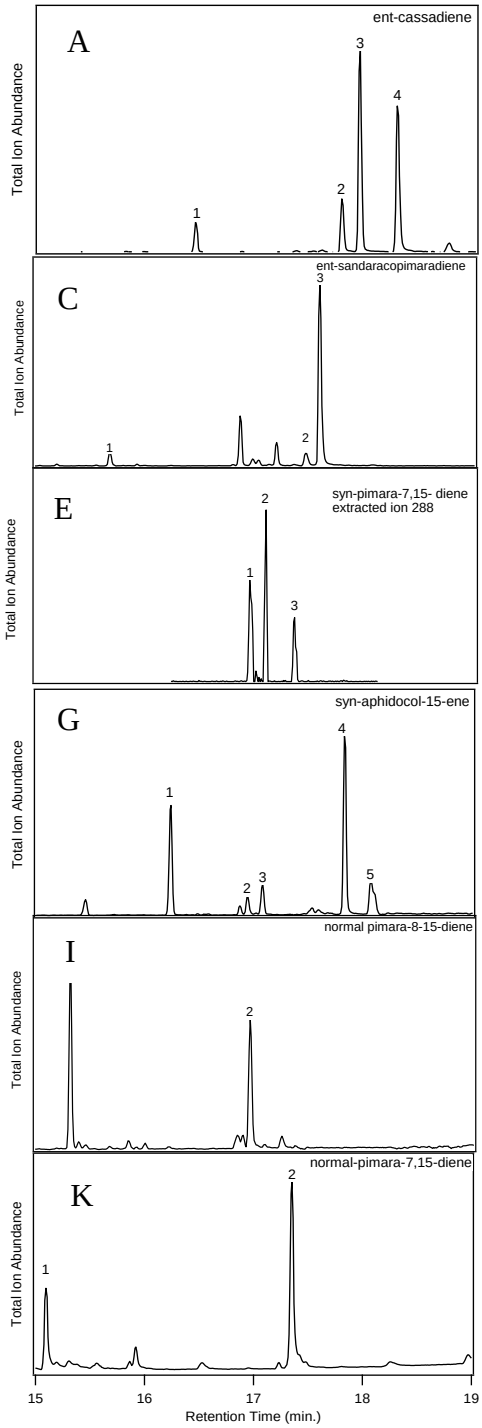
Figure 4: Comparative relative activity of AtKO and OsKO with olefins. Turnover of olefins in *E. coli* metabolic engineering system optimized to produce olefins and either AtKO or OsKO2 and AtR1 A) Total product turnover B) turnover to alcohol C) turnover to acid



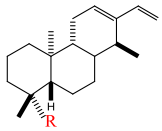

B
Substrate
Product Formation
labda-7,13E-dien-15-ol
Abietadiene
miltradiene
normal pimara-8(9)-15-diene
normal pimara-7(8)-15-diene
syn-stemod-13(17)-ene
syn-labda-8(17), 12E,14-triene
syn-stemar-13-ene
syn-pimara-7(8),15-diene
syn-aphidocol-15-ene
ent-beyerene
kauren-16-ol
trachlobane
ent-pimara-8(14), 15 -diene
ent- sandaracopimaradiene
ent-cassa-12,15-diene
ent-isokaurene
ent-kaurene
Olefin
Alcohol
0
20
40
60
80
100
Relative Activity
C
Substrate
Product Formation
labda-7,13E-dien-15-ol
Abietadiene
miltradiene
normal pimara-8(9)-15-diene
normal pimara-7(8)-15-diene
syn-stemod-13(17)-ene
syn-labda-8(17), 12E,14-triene
syn-stemar-13-ene
syn-pimara-7(8),15-diene
syn-aphidocol-15-ene
ent-beyerene
kauren-16-ol
trachlobane
ent-pimara-8(14), 15 -diene
ent- sandaracopimaradiene
ent-cassa-12,15-diene
ent-isokaurene
ent-kaurene
Alcohol
Acid
0
20
40
60
80
100
Relative Activity %
OsKO2
AtKO

.

Figure 5: AtKO catalyzes diterpene hydroxylations on multiple substrates of *ent, syn* and normal stereochemistry. Panel 1: GC chromatogram of extract from *E. coli* engineered for production of A) *ent*-cassadiene C) *ent*- sandaracopimaradiene E) *syn*-pimara-7,15-diene G) *syn*-aphidocol-15-ene, I) normal- pimara-7,15-diene K) normal-pimara-8,15-diene co-expressed with N-terminally modified AtKO and AtR1. Panel 2: Resultant product structure of corresponding peaks in B) *ent*- cassadiene D) *ent*- sandaracopimaradiene- F) *syn*-pimara-7,15-diene- Peak 3 is an unidentified alcohol peak H) syn-aphidocol-15-ene -Peak 2 and 3 correspond to 2β and 3α –Hydroxy-syn-pimara-7,15-diene J) normal- pimara-7,15-diene L) normal-pimara-8,15-diene. Relevant mass spectra are shown in Figure S2-4



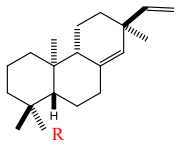
B



R
1- H
2- CHO
3- COOH
4- CH_2OH

ent - cassadiene

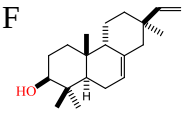
D



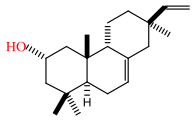
R
1- H
2 - COOH
3 - CH_2OH

ent-sandaracopimaradiene

F

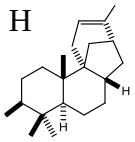


Peak 1
3β-Hydroxy-*syn*-pimara-7,15-diene

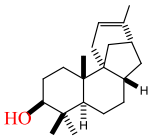


Peak 2
2α-Hydroxy-*syn*-pimara-7,15-diene

H

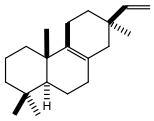


Peak 1
syn-aphidocol-15-ene

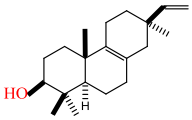


Peak 4
3β–*syn*-aphidocol-15-ene

J

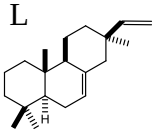


Peak 1
normal pimara-8(9),15-diene

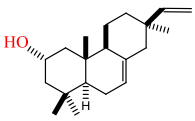


Peak 2
3β-Hydroxy-normal pimara-8(9),15-diene

L



Peak 1
normal pimara-7,15-diene



Peak 2
2α-Hydroxy-normal pimara-7,15-diene

Supplementary Information

Table 1: Combinatorial gene expression of GGPPs, CPS and KSL genes to produce olefins in metabolic engineering system

GGPPs and CPS in pACYC Duet Vector	**Class 1 in pDEST 14/15**	**Resultant Olefin**
pGGeC	AtKS	*ent*-Kaurene
(GGPPs and ent-CPS)	OsKSL5	*ent*-pimara-8(14), 15-diene
	OsKSL6	*ent*-iso- kaurene
	OsKSL7	*ent*-cassa-12,15-diene
	OsKSL10	*ent*-sandaracopimaradiene
	RcKSL2	trachlobane
	PpCPSKS	ent kaurene
		Kauren-16-ol
	TaKSL5	ent-beyerene
pGGsC	OsKSL4 T696I	*syn*-aphidocol-15-ene
(GGPPs and syn-CPS)	OsKSL4	*syn*-pimara-7,15-diene
	OsKSL8	*syn*-stemar-13-ene
	OsKSL10	syn-labda-8(17), 12*E*, 14-triene
	OsKSL11	*syn*-stemodene-13 (17)-ene
	TaKSL4	syn-pimara-9(11), 15-diene
pGGnC	AfDTS	normal pimara-7(8)-15-diene
(GGPPs and normal-CPS)	SaDTS	normal pimara-8(9)-15-diene
	DsKSL	miltradiene
	AgAS	abietadiene
pGG (GGPPs)	SmCPSKSL1	Labda-7,13E-dien-15-ol

Abbreviations: GGPPs- geranyl geranyl diphosphate synthase; CPS-copalol diphosphate synthase; KSL-kaurene synthase like; At- *Arabidopsis thaliana*; Ag- *Abies grandies*; Af-*Aspergillus fumigada*; Ds-*Salvia miltorhizzia* (Danshen) ; Pp-*Physcomitrella patens*; Os-*Oryza sativa*; Sa-*Salinispora arenicola*; Rc-Ricinus comunis; Sm- *Selaginella moellendorffii*; *Ta-Triticum aestavia*

Figure S1: KO activity of AtKO and OsKO2. GC chromatogram of extract from *E. coli* engineered for production of ent-kaurene co-expressed with N-terminally modified AtKO or OsKO2 and AtR1. Peaks correspond to the successive oxidation: 272 –residual diterpene; 288-kaurenol; 286- kaurenal and 316 – derived methyl ester of ent-kaurenoic acid

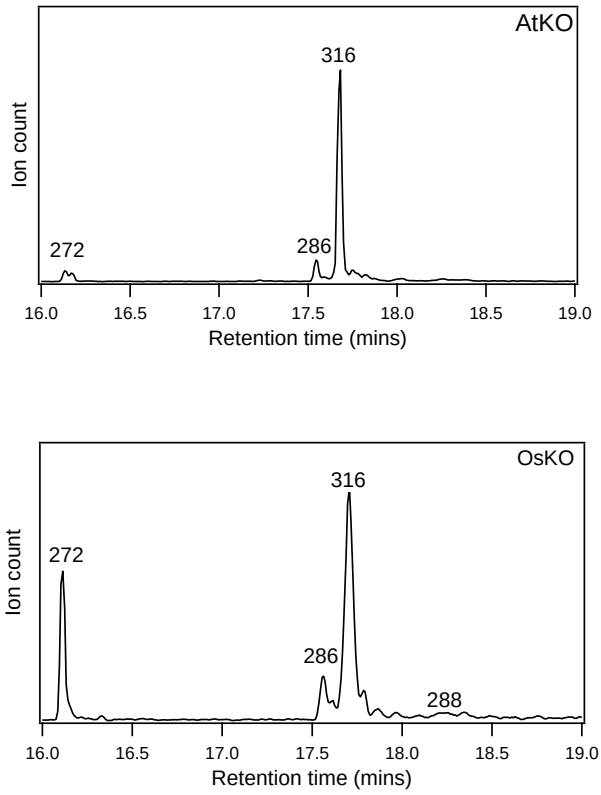


Figure S2: GC chromatogram of extract from *E. coli* engineered for production of A) *ent*-cassadiene F) *ent*-sandaracopimaradiene co-expressed with N-terminally modified AtKO and AtR1. Mass spectra for respective peaks in the chromatogram B) Peak 1-*ent*-cassadiene C) *ent*-cassadienal D) Peak 3- *ent*-cassadienoic acid E) Peak 4-19α-hydroxy-*ent* cassadiene G) Peak 1-ent-sandaracopimaradiene H) Peak 2-*ent* sandaracopimaradienoic acid I) Peak 5-19α-hydroxy- ent-sandaracopimaradiene

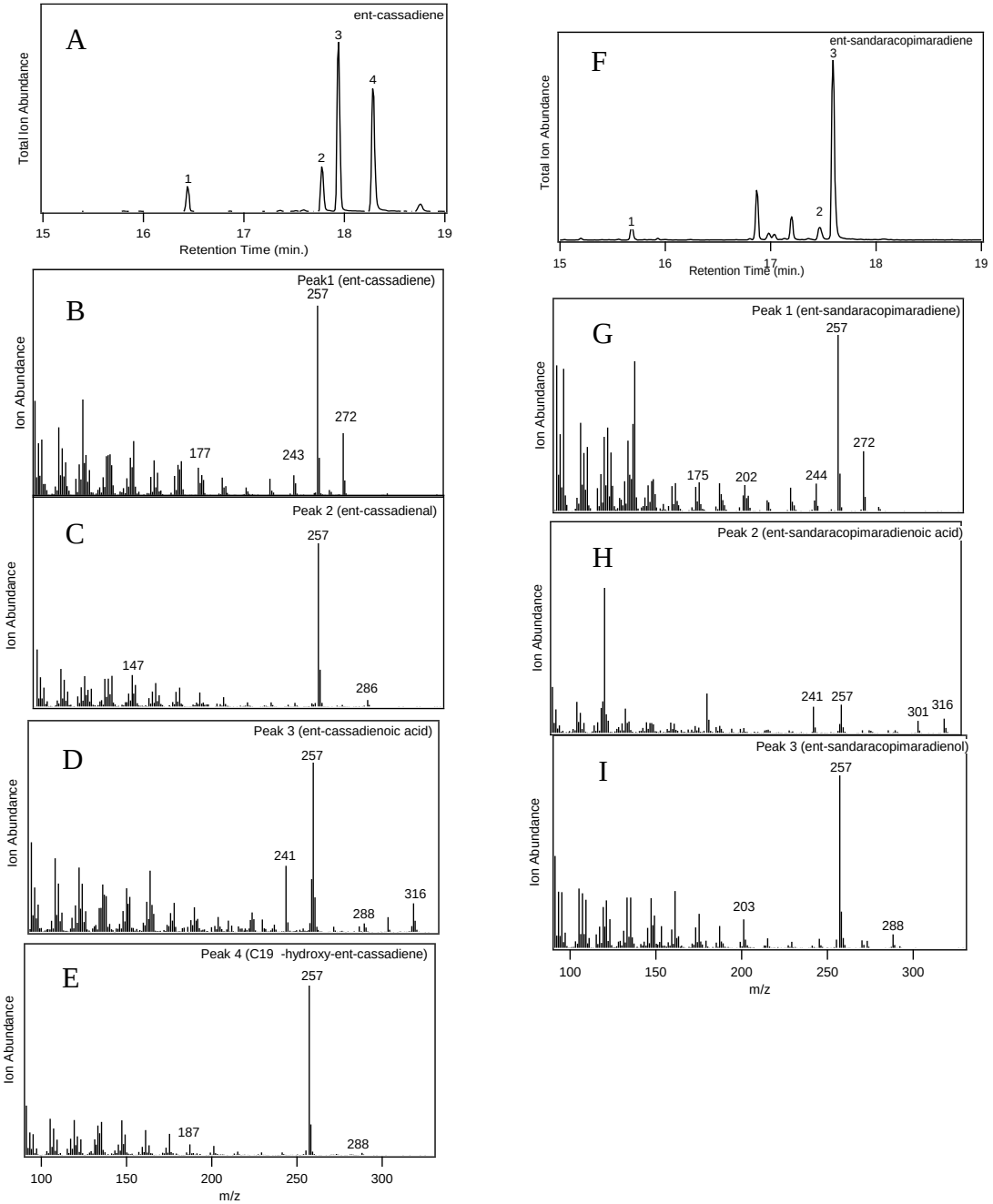


Figure S3: GC chromatogram of extract from *E. coli* engineered for production of A) *syn*-pimara-7,15-diene E) syn-aphidocol-15-ene co-expressed with N-terminally modified AtKO and AtR1. Mass spectra for respective peaks in the chromatogram B) Peak 1-2α-hydroxy-syn-pimara-7,15-diene C) Peak 2-3β-hydroxy-syn-pimara-7,15-diene D) Peak 3- unidentified compound F) Peak 1-syn-aphicol-15-ene G) Peak 4-3β-hydroxy-syn-aphidocol-15-ene J) Peak 5-unidentified product

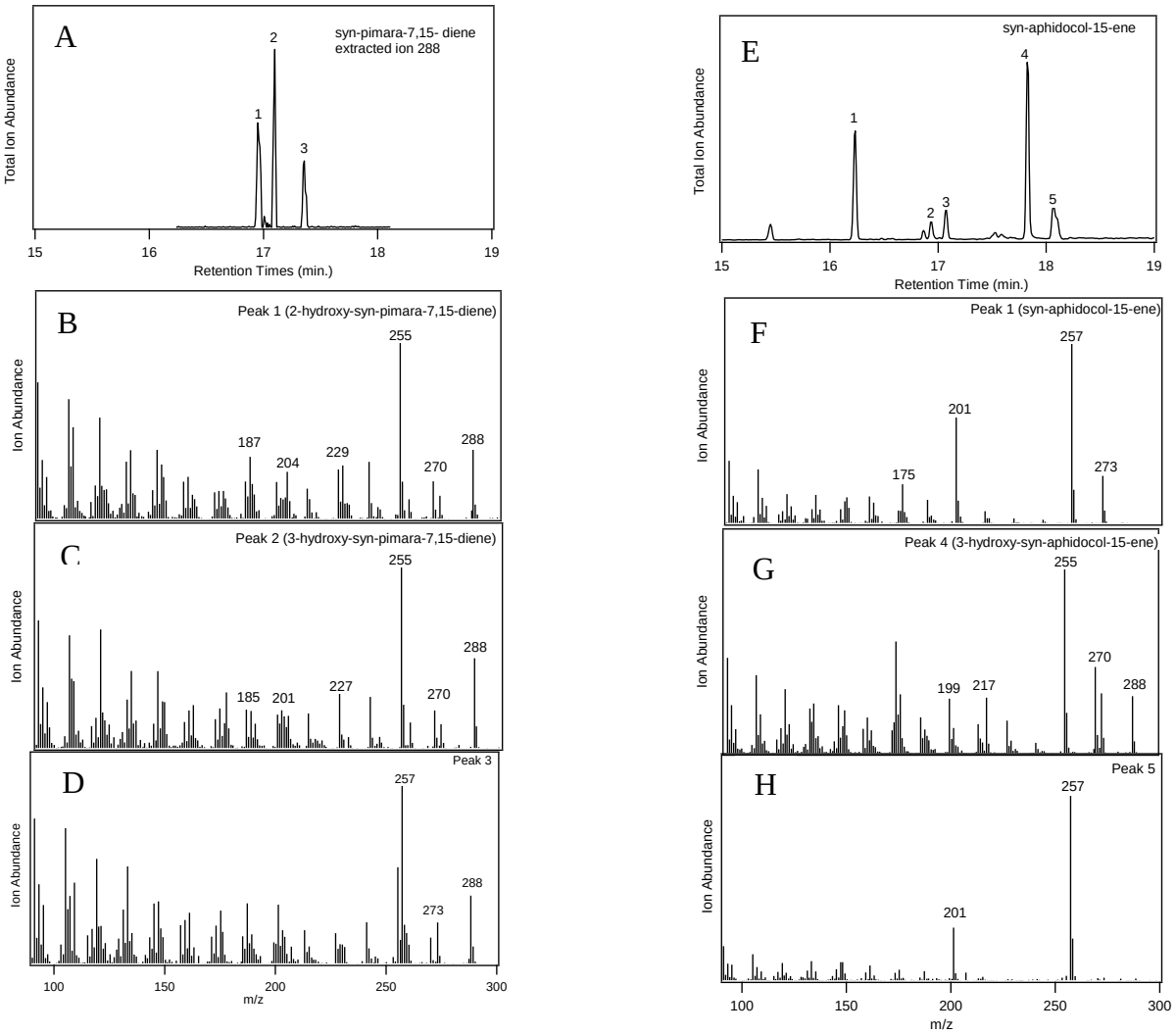


Figure S4: GC chromatogram of extract from *E. coli* engineered for production of A) normal-pimara-7,15-diene D) normal-pimara-8,15-diene co-expressed with N-terminally modified AtKO and AtR1. Mass spectra for respective peaks in the chromatogram B) Peak normal-pimara-7,15-diene C) Peak 2- 2α-hydroxy-normal-pimara-7,15-diene E) Peak 1-normal-pimara-8,15-diene F) Peak 2-3β-hydroxy-syn-pimara-8,15-diene

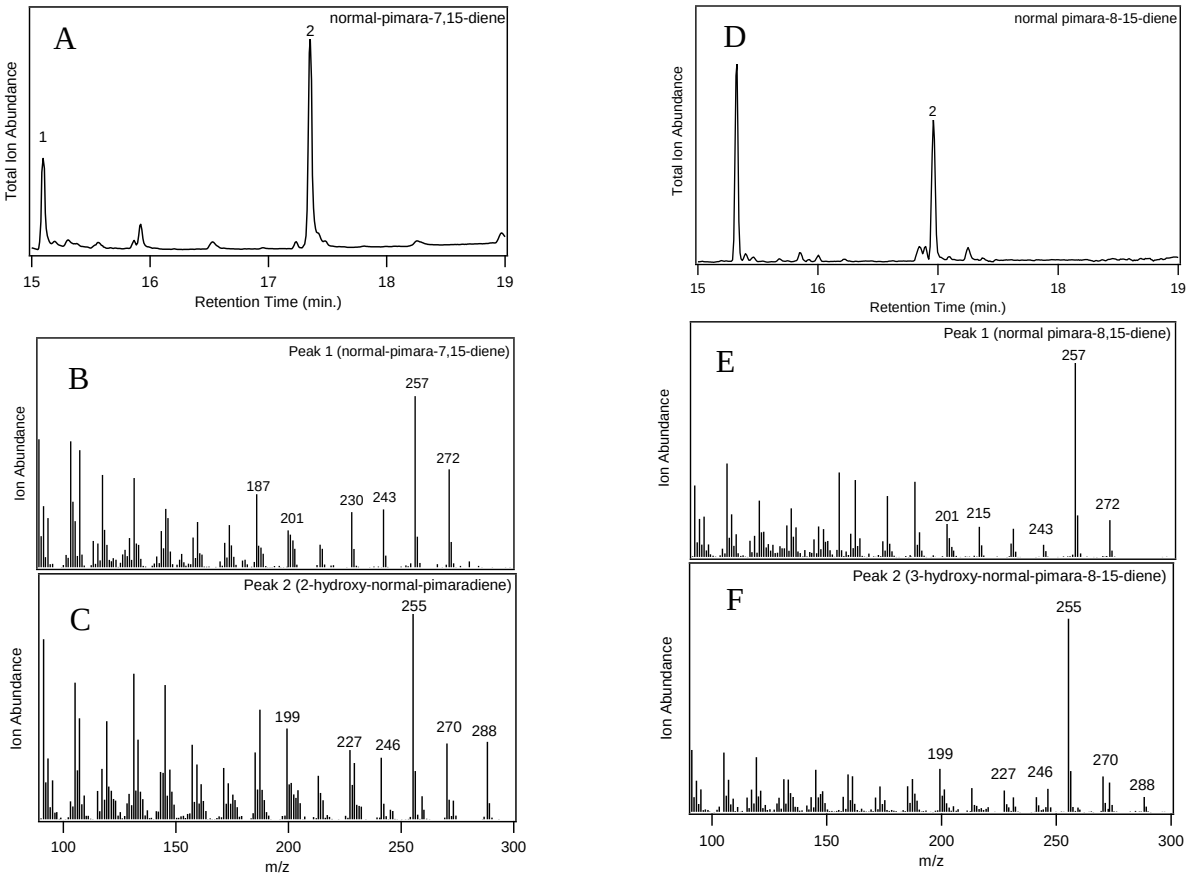
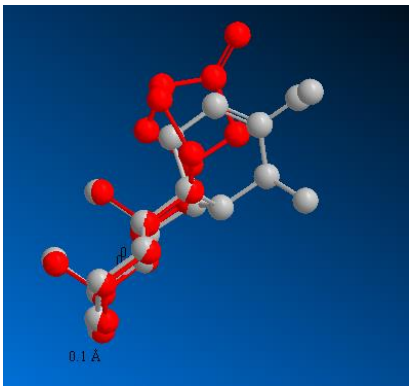
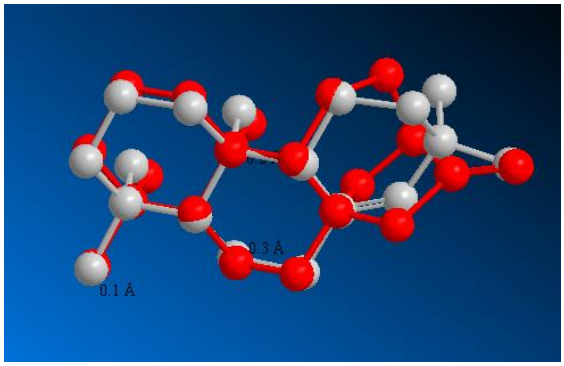


Figure S5: A three dimensional overlay of kaurene with A) *ent*-cassadiene B) *ent*-sandaracopimaradiene C) *syn*-pimara-7-15-diene D) *syn*-aphidocol-15-ene E) normal-pimara-7,15-diene utilizing C6, C10 and C19 as atom pairs

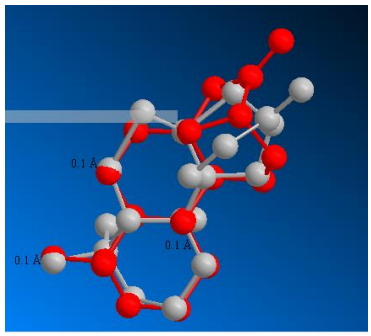


A) *ent*-cassadiene

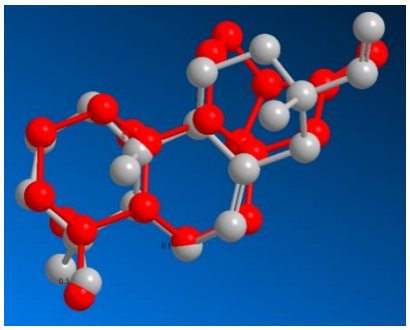


B) *ent*-sandaracopimaradiene

C) *syn*-pimara-7, 15-diene



D) *syn*-aphidocol-15-ene



E) normal-pimara-7,15-diene

Figure S6: A) AtKO products ent stereochemistry B) HMBC correlations and NOESY Nuclear Overhauser Effect dipole-dipole correlations used to assign configurations

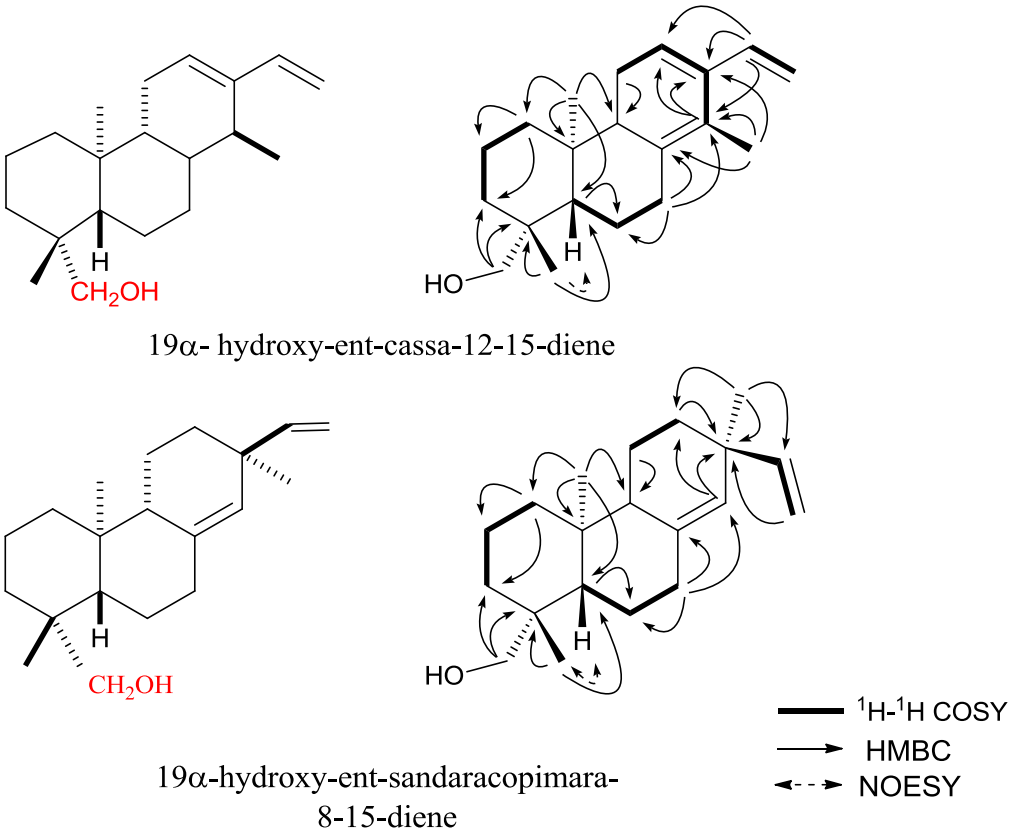


Figure S7: A) AtKO products syn stereochemistry B) HMBC correlations and NOESY Nuclear Overhauser Effect dipole-dipole correlations used to assign configurations

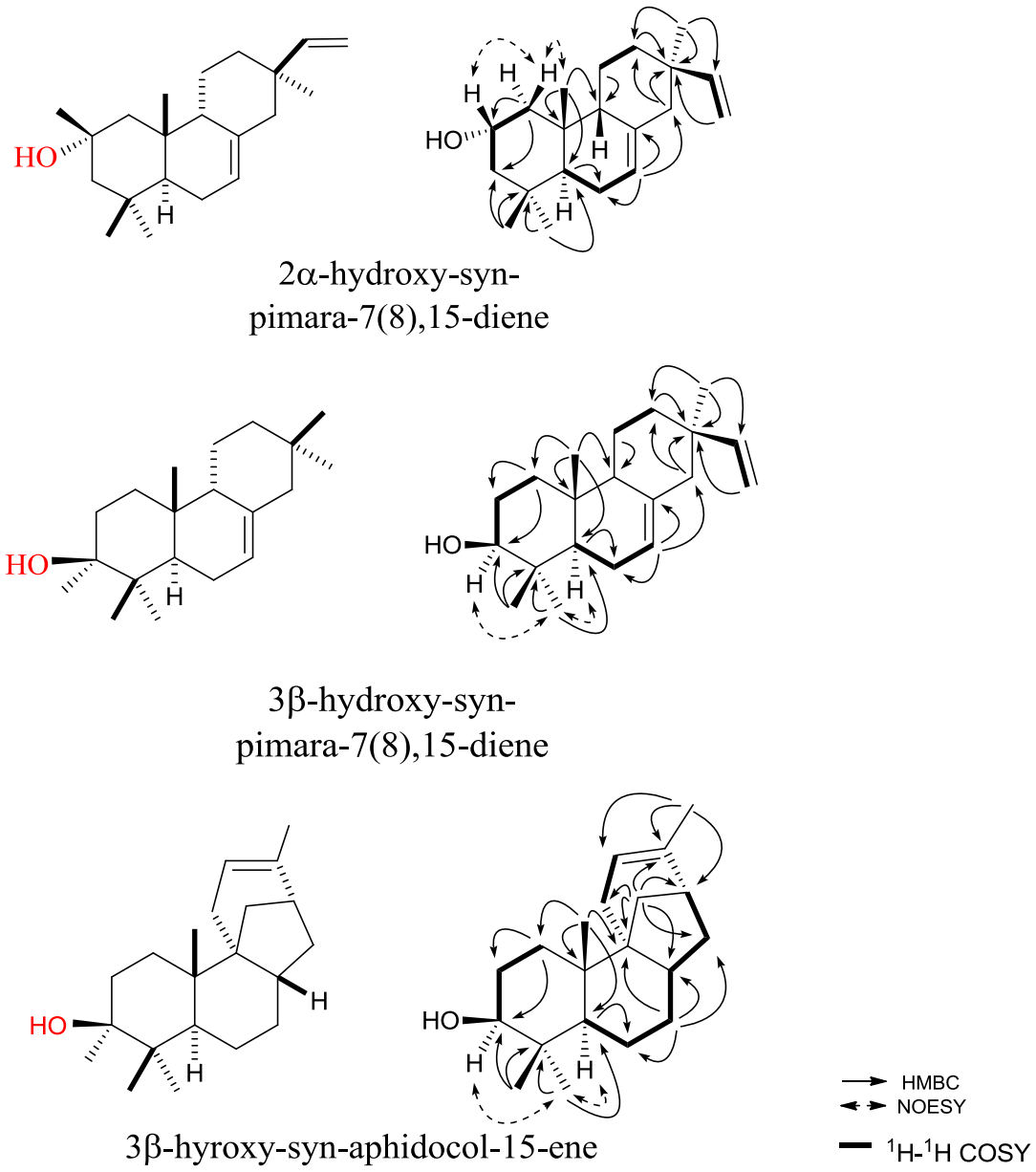
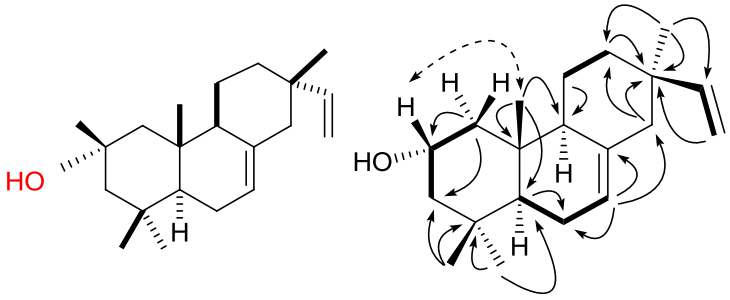
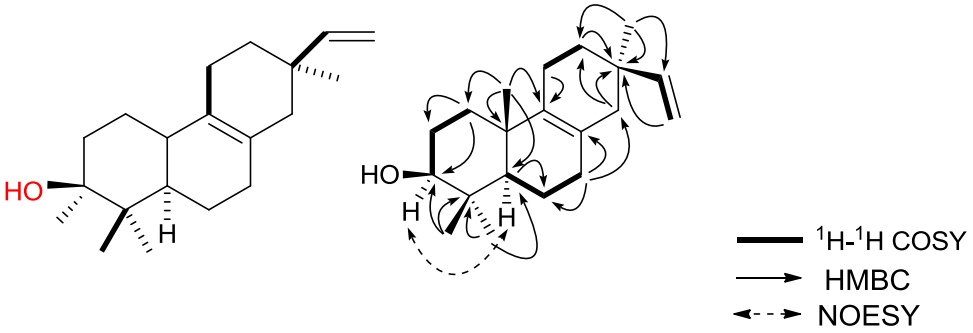


Figure S8: A) AtKO products normal stereochemistry B) HMBC correlations and NOESY Nuclear Overhauser Effect dipole-dipole correlations used to assign configurations



2α-hydroxy-
normal pimara-7(8),15-diene



3β-hydroxy-syn pimara-8,15-diene

Table S2: H' and ^{13}C NMR assignments for 19α-hydroxy-cassa-12,15-diene

Position	**19α-hydroxy-cassa-12,15-diene**	
	δ_H	δ_C
1 a	1.63 (1H, m)	39.4
b	0.96 (1H, d)	
2 a	1.47 (1H, m)	18.6
b	1.41 (1H, m)	
3 a	1.80 (1H, m)	35.7
b	0.95 (1H, m)	
4		38.7
5	1.03 (1H, m)	56.5
6 a	1.71 (1H, m)	21.8
b	1.33 (1H, m)	
7 a	1.63 (1H, m)	31.6
b	1.32 (1H, m)	
8	1.51 (1H, m)	35.3
9	1.30 (1H, m)	44.6
10		37.2
11 a	2.09 (1H, m)	25.6
11 b	1.90 (1H, ddd, *J* = 19.9, 9.3, 3.4 Hz)	
12	5.58 (1H, t, *J* = 3.8 Hz)	128.7
13		142.1
14	2.37 (1H, m)	32.2
15	6.20 (1H, dd, *J* = 17.6, 10.8 Hz)	138.9
16 a	5.05 (1H, d, *J* = 17.6 Hz)	109.7
b	4.88 (1H, d, *J* = 10.8 Hz)	
17	0.90 (3H, s)	14.7
18	0.96 (3H, s)	27.1
19	3.78 (1H, d, *J* = 10.9 Hz)	65.8
	3.44 (1H, d, *J* = 10.9 Hz)	
20	0.77 (3H, s)	15.1

Table S3: H' and ^{13}C NMR assignments for 19α-hydroxy-sandaracopimaradiene

Position	**19α-hydroxy-sandaracopimaradiene**	
	δ_H	δ_C
1 a	1.66 (1H, m)	39.9
b	0.98 (1H, m)	
2 a	1.42 (1H, m)	19.3
b	1.39 (1H, m)	
3 a	1.77 (1H, br.d, *J* = 13.7 Hz)	36.0
b	0.89 (1H, m)	
4		39.2
5	1.13 (1H, dd, *J* = 12.9, 1.9 Hz)	56.4
6 a	1.60 (1H, m)	23.1
b	1.21 (1H, m*)*	
7 a	2.17 (1H, ddd, *J* = 14.1, 4.3, 1.9 Hz)	36.9
b	1.93 (1H, m)	
8		137.4
9	1.63 (1H, m)	51.3
10		38.8
11 a	1.52 (1H, m)	19.5
b	1.39 (1H, m)	
12	1.37 (1H, m)	35.1
	1.27 (1H, dd, *J* = 12.5, 3.7 Hz)	
13		38.0
14	5.13 (1H, m)	129.4
15	5.69 (1H, dd, *J* = 17.5, 10.6 Hz)	149.6
16 a	4.82 (1H, d, *J* = 17.5 Hz)	110.6
b	4.80 (1H, d, *J* = 10.6 Hz)	
17	0.95 (3H, s)	26.6
18	0.90 (3H, s)	27.6
19	3.75 (1H, d, *J* = 10.8 Hz)	65.8
	3.35 (1H, d, *J* = 10.8 Hz)	
20	0.68 (3H, s)	16.7

Table S4: H' and ^{13}C NMR assignments for 2α-hydroxy-syn-pimara-7,15-diene

Position	**2α-hydroxy-syn-pimara-7,15-diene**	
	δ_H	δ_C
1 a	1.46 (1H, m)	46.8
b	1.27 (1H, t, *J* = 11.3 Hz)	
2	3.84 (1H, tt, *J* = 11.3, 3.3 Hz)	65.9
3 a	1.68 (1H, dt, *J* = 12.2, 3.3)	52.5
b	1.04 (1H, dd, *J* = 12.2, 11.3 Hz)	
4		35.2
5	1.12 (1H, dd, *J* =12.0, 4.5 Hz)	43.5
6 a	1.94 (1H, m)	24.21
b	1.76 (1H, m*)*	
7	5.24 (1H, br.d, *J* = 5.3 Hz)	120.3
8		137.1
9	1.31 (1H, m)	53.7
10		37.6
11 a	1.63 (1H, m)	25.7
11 b	1.20 (1H, m)	
12	1.44 (1H, m)	38.2
	1.37 (1H, td, *J* = 13.1, 4.2 Hz)	
13		39.4
14 a	1.92 (1H, m)	48.4
b	1.75 (1H, *J* = 12.3, 2.2 Hz)	
15	5.74 (1H, dd, *J* = 17.5, 10.7 Hz)	150.9
16 a	4.85 (1H, d, *J* = 17.5 Hz)	109.8
b	4.78 (1H, d, *J* = 10.7 Hz)	
17	0.81 (3H, s)	22.4
18	0.873 (3H, s)	24.18
19	0.870 (3H, s)	34.1
20	0.90 (3H, s)	23.5

Table S5: H' and ^{13}C NMR assignments for 3β-hydroxy-syn-pimara-7,15-diene

Position	**3β-hydroxy-syn-pimara-7,15-diene**	
	δ_H	δ_C
1 a	1.50 (1H, m)	34.8
b	1.16 (1H, m)	
2	1.56 (2H, m)	28.0
3	3.12 (1H, t, *J* = 7.6 Hz)	80.2
4		39.4
5	1.13 (1H, m)	43.8
6 a	1.93 (1H, m)	24.0
b	1.88 (1H, m)	
7	5.24 (1H, br.d, *J* = 4.8 Hz)	120.1
8		137.3
9	1.28 (1H, m)	53.6
10		35.6
11 a	1.61 (1H, m)	25.7
11 b	1.14 (1H, m)	
12	1.42 (1H, m)	38.3
	1.36 (1H, *J* = 13.1, 4.0 Hz)	
13		39.2
14 a	1.92 (1H, m)	48.4
b	1.75 (1H, *J* = 12.2, 2.2 Hz)	
15	5.74 (1H, dd, *J* = 17.5, 10.8 Hz)	150.9
16 a	4.84 (1H, d, *J* = 17.5 Hz)	109.8
b	4.78 (1H, d, *J* = 10.8 Hz)	
17	0.802 (3H, s)	22.4
18	0.799 (3H, s)	16.2
19	0.93 (3H, s)	28.9
20	0.85 (3H, s)	22.8

Table S6: H' and ^{13}C NMR assignments for 3β-hydroxy-syn-aphidocol-15-ene

Position	**3β-hydroxy-syn-aphidocol-15-ene**	
	δ_H	δ_C
1 a	1.68 (1H, td, *J* = 12.4, 5.5 Hz)	32.4
b	1.14 (1H, m)	
2	1.56 (2H, m)	28.6
3	3.08 (1H, dd, *J* = 10.1, 6.0 Hz)	79.9
4		39.5
5	1.21 (1H, m)	45.1
6 a	1.54 (1H, m)	23.3
b	1.28 (1H, m)	
7 a	1.58 (1H, m)	26.1
b	1.36 (1H, qd, *J* = 12.6, 4.4 Hz)	
8	2.00 (1H, m)	42.4
9		48.8
10		40.5
11 a	2.19 (1H, m)	28.5
b	2.08 (1H, m)	
12	4.90 (1H, br.s)	118.4
13		148.0
14	2.00 (1H, m)	39.6
15	1.44 (1H, dd, *J* = 10.3, 5.4 Hz)	36.0
	1.17 (1H, m)	
16 a	1.87 (1H, td, *J* = 11.3, 7.5 Hz)	39.0
b	0.93 (1H, d, *J* = Hz)	
17	1.56 (3H, s)	22.3
18	0.72 (3H, s)	15.8
19	0.90 (3H, s)	29.2
20	0.92 (3H, s)	16.3

Table S7: H' and ^{13}C NMR assignments for 2α- hydroxyl-normal-pimara-7,15-diene

Position	2α-hydroxyl-normal-pimara-7,15-diene	
	δ_H	δ_C
1 a	2.05 (1H, dt, *J* = 11.9, 2.8 Hz)	49.6
b	0.88 (1H, m)	
2	3.77 (1H, m)	65.8
3 a	1.68 (1H, dt, *J* = 12.0, 2.9 Hz)	51.8
b	1.06 (1H, t, *J* = 12.0 Hz)	
4		35.3
5	1.03 (1H, dd, *J* = 12.0, 4.0 Hz)	50.3
6 a	1.87 (1H, m)	23.9
b	1.78 (1H, m)	
7	5.30 (1H, br.d, *J* = 3.1 Hz)	122.2
8		135.9
9	1.63 (1H, br.d, *J* = 9.8 Hz)	52.6
10		38.0
11 a	1.48 (1H, m)	20.8
11 b	1.28 (1H, m)	
12	1.41 (1H, m)	36.6
	1.29 (1H, m)	
13		37.4
14 a	1.89 (1H, m)	46.6
b	1.84 (1H, dd, *J* = 13.9, 2.5 Hz)	
15	5.72 (1H, dd, *J* =17.5, 10.8 Hz)	150.9
16 a	4.85 (1H, d, *J* = 17.5 Hz)	109.9
b	4.79 (1H, d, *J* = 10.8 Hz)	
17	0.78 (3H, s)	22.1
18	0.88 (3H, s)	23.8
19	0.85 (3H, s)	34.2
20	0.82 (3H, s)	16.4

Table S8: H' and ^{13}C NMR assignments for 3β-hydroxy- normal-pimara-8,15-diene

Position	**3β-hydroxy-normal-pimara-8,15-diene**	
	δ_H	δ_C
1 a	1.66 (1H, m)	35.3
b	1.06 (1H, td, *J* = 12.9, 3.5 Hz)	
2 a	1.59 (1H, m)	28.3
b	1.51 (1H, m)	
3	3.15 (1H, dd, *J* = 11.3, 3.1 Hz)	79.6
4		39.4
5	1.02 (1H, d, *J* = 12.2 Hz)	51.6
6 a	1.60 (1H, m)	19.2
b	1.40 (1H, m*)*	
7	1.86 (2H, m)	33.2
8		125.2
9		137.0
10		37.9
11	1.80 (2H, m)	21.8
12 a	1.41 (1H, m)	35.5
b	1.23 (1H, m)	
13		35.7
14 a	1.73 (1H, d, *J* = 17.3 Hz)	42.4
b	1.64 (1H, d, *J* = 17.3 Hz)	
15	5.65 (1H, dd, *J* = 17.3, 10.7 Hz)	146.8
16 a	4.81 (1H, d, *J* = 10.7 Hz)	111.3
b	4.76 (1H, d, *J* = 17.3 Hz)	
17	0.884 (3H, s)	28.6
18	0.73 (3H, s)	16.1
19	0.92 (3H, s)	28.5
20	0.877 (3H, s)	20.0

CHAPTER IV: INVESTIGATING CYTOCHROMES P450 INVOLVED IN GA BIOSYNTHESIS FROM SELAGINELLA *MOELLENDORFFII*

Abstract

Kaurene oxidases (KO) are multifunctional cytochromes P450 that catalyze oxidation of *ent*-kaur-16-ene to *ent*-kaur-16-en-19-oic acid, nominally as an early step in the complex process of gibberellin phytohormone biosynthesis. Recent studies demonstrate that the lycophyte *Selaginella moellendorffii* produces and responds to gibberellin, whereas the bryophyte *Physcomitrealla patens* does not. Both plants produce *ent*-kaurene, *P. patens* contains a functional kaurene oxidase homolog and homologs have been identified in *S. moellendorffii,* albeit only distantly related to those known to operate in angiosperms (as demonstrated by their classification into separate sub-families of the relevant CYP701 family of P450s). However, only *S. moellendorffii* contains the subsequently acting enzymes necessary for gibberellin biosynthesis. This suggests that the complex process of gibberellin biosynthesis may have evolved via sequential evolution; first the formation of *ent*-kaurenoic acid, with its own physiological relevance, followed by the remainder of the pathway. Here, we describe the work in progress in characterizing the putative cytochromes p450 involved in GA biosynthesis.

Introduction

Gibberellic acid (GA) is a phytohormone important in regulation of growth and development of plants. Its complex biosynthesis is divided into three stages comprising of terpene synthases which catalyze the reaction of geranyl geranyl diphosphate to *ent*- kaurene, two functionally independent cytochrome P45O groups which convert *ent*-kaurene to $GA_{12.}$

The formation of bioactive GAs in plants is catalyzed by 2 oxoglutarate dependent dioxygenases (2ODDs) in plants.

Kaurene oxidases (KO) are multifunctional cytochromes P450, of the CYP701 family that catalyze oxidation of *ent*-kaur-16-ene to *ent*-kaur-16-en-19-oic acid through kaurenol and kaurenal as intermediates [1-4]. Functional KO's have been identified and reported in angiosperms notably arabidopsis, pea, cucumber, rice and lower non vascular land plant *Physcomitrella patens* belonging to the CYP701A and CYP701B subfamilies respectively ([1, 3-5]. Putative homologs of the CYP701 cytochrome P450s have been identified in *Selaginella moellendorffii* suggesting evolutionary enzymatic conservation of this stage of the GA pathway as genes of similar function are grouped within the same family [6].

The lycophyte, *S. moellendorffii* is well positioned in the evolutionary phylogenetic tree of land plants to aid in the understanding of comparative evolution of vascular plants. The lycophytes emerged after the bryophytes about 150-200 million years earlier than the angiosperms. It is important in studying the evolution of plant adaptations that facilitated land colonization by green plants [7]. *S. moellendorffii* can offer insight into the biochemical development of biosynthetic and regulatory hormone pathways such as that of GA.

It has been demonstrated that *S. moellendorffii* has a functional GA signaling system whereas that of the moss, *P. patens,* did not exhibit functional conservation [8]. *P. patens* contain functional terpene synthase and KO enzymes to produce *ent*-kaurenoic acid. Kaurenoic acid has a physiological role in *P. patens* as protonematal defects of bifunctional CPS/KS mutants could be rescued by application of kaurene or kaurenoic acid [3, 9]. However, *P. patens* does not however contain the downstream enzymes [8, 10]. *S.*

moellendorffii genome contains putative homologs of KO, KAO and the downstream enzymes for GA biosynthesis of which only 2ODDs have been functionally characterized.

The CYP88 family is involved in the conversion of *ent*-kaurenoic acid to GA_{12} in the GA biosynthetic pathway. This family of CYPs first appears in the lycophyte and is absent in *P.patens*. Three putative CYP88 genes in two different subfamilies have been identified in the *Selaginella* genome.

The catalytic cycle of cytochromes P450 is dependent upon the transfer of two electrons from NADPH to P450 mediated through NADPH-cytochrome p450 oxidoreductase (CPR) [11]. There is a high conservation amongst amino acids of CPRs from different species and this is reflected in the ability for a CPR from a different species to partially complement for P450 activity of an unrelated species (AtR1 in yeast and used for different species). Plant genomes contain multiple CPR isoforms with up to 3 being identified in higher plants (*Oryza sativa*, *Populus tricocarpa),* while the moss, *P. patens* contains four and the lycophyte, *S. moellendorffii* has three. The presence of different paralogs in plants suggests the need for specific interactions between some P450s and its CPR.

In this report we describe the work that has been carried out to characterize the putative cytochromes p450 of the GA biosynthetic pathway.

Experimental Procedures

General Procedures

Molecular biology reagents were obtained from Invitrogen and all other chemicals were from Fisher Scientific. Sequence alignments were carried out using the Align program from the Vector NTI software package (Invitrogen), using standard parameters.

The sequence for the genes encoding KO, KAO and CPR were obtained from the information available at the Joint Genome Institute (http://www.jgi.doe.gov) and Phytozyme. Fully codon optimized synthetic versions of SmKO1, SmKAO and SmCPR2 were obtained from Genscript. An additional synthetic version for SmKO1 was obtained from DNA 2.0. Modified version of the synthetic versions were expressed in the metabolic engineering system as previously described [2, 12-17]

Cloning

The SmKO1, SmKAO1, SmKAO3, CPR1 and CPR3 genes were cloned from *S. moellendorffi* plant utilizing primers shown in Table 1. The native and synthetic genes was amplified using PCR, purified using DNA gel extraction kit (Millipore) and cloned into the Gateway system vector pENTR/SD/D-TOPO using a topoisomerase mediated procedure. The CYPs were modified by 5' primer performs a deletion and introduces a ten amino acid leader sequence [2, 12, 14-19].

Functional analysis

Expression in *E.coli* - For functional analysis, the SmKO gene was co-expressed in C41 cells with the upstream genes: pGGeC that contains a GGPP synthase and CPS genes; pDEST14/AtKS which is a kaurene synthase from *A. thaliana*: PCDF sSmKO/AtR1, that is a cytochromes P450 reductase from *A. thaliana*. PCDF AtKO/AtR1 was expressed under similar expression conditions. pJ821/SmKO was coexpressed with pGGec and PCDF/AtKS-AtR1, or pGG/FfKS PCDF-SmCPR1 or PCDF-SmCPR2. pJ821/SmKO was also subcloned into pENTR and pDEST 17 and PCDF DEST-AtR1 destination vectors. Co-expression combinations were as follows pGG/FfKS with PCDF/SmKO-AtR1 or pGG/GfKS,

PCDF/SmCPR1 or SmCPR2 and pDEST 17/SmKO. A summary of expression of SmKO with alternative SmCPRs is indicated in Table 2. SmKAO genes were modified as previously described and cloned into PET Duet/ SmKAO-AtR1 and coexpressed with PCDF/AtKO-AtR1 and pGG/GfKS. The strains were grown in 50 mL cultures of TB media to a mid-log phase (OD_{600}=0.6) at 37^{o}C. They were then transferred and incubated at 16^{o}C for I hour after which IPTG was added to a final concentration of 0.5 mM, 0.2 mg/ml riboflavin and 1mM aminolevulinic acid. Thereafter, they were grown for approximately 72 hours and extracted with 50 mL of hexanes. The separated hexanes were dried down under nitrogen. The sample was methylated with diazomethane and redissolved in 200 μL of hexane for analysis by gas chromatography – mass spectrometry. Kaurenoic acid and GA_{12} were identified by comparison of retention time and mass spectra with that of authentic standards.

Expression in Yeast

For expression in *Saccharamyces cerevisiae* WAT11, native SmKO1 were cloned into pYES-DEST52 and transformed into the WAT11 strain [20, 21] using the lithium acetate protocol followed by selection on SC-Ura media. *S. cerevisiae* containing the desired plasmid was grown using the low density protocol and induced using 2% galactose [20]. A 50 ml culture was grown for 5 hours at 28^{o}C, after which the culture was supplemented with 50 μM kaurene (in THF and methanol). After 24 h, media was extracted with an equal volume of hexane and analyzed by GC-MS analysis as described above.

For *Pichia pastoris* expression native SmKO genes were cloned into pPICZA plasmid (Invitrogen Carlsbad) and pEP Strep plasmid – pPIC HOLI backbone with a DEST cassette inserted (Gershenzon laboratory). The pPICZA were transformed into X-33 and pEPStrep

into GS115. Positive transformants were expressed as described in the manual. In vitro assays were carried out containing 500 µL, ATR1, NADPH 0.1 mM and regeneration system, 50 uM kaurene. Samples were incubated overnight at 28^{o}C, extracted with hexanes and analyzed by GC-MS.

Results

Bioinformatic analysis and cloning

SmKO – Four putative KO homolog, SmKO1/CYP701C1 (JGI ID#-74427) and SmKO2/CYP701C2 (JGI I#D-421373) and their allelic variants, were identified in the Joint Genome Institute website (http://www.jgi.doe.gov). Sequence analysis of the SmKO1 homolog in the genome indicates that SmKO1/CYP701C1 gene encodes a protein that contains 500 amino acids in length. SmKO2/CYP701C2 has a missing portion of the conserved cytochromes P450 sequence FxxGxxxCxG on the C terminal end. It also has additional amino acids and missing amino acids in conserved regions as compared to the other KO producing species. The predicted sequences of SmKO1 and SmKO2 are 90 % identical to each other.

Evolutionary relatedness of the putative SmKO1/CYP701C1 homolog to characterized functional KO's is demonstrated in the phylogentic tree in Figure 1. SmKO1 is most closely related to PpKO with 49 % identity with PpKO and less closely related to the angiosperm KO reflecting between 43 and 44 % identity. This is lower than the identity observed amongst the *A. thaliana, C. maxima, P. sativa* and *O. sativa* which varies between 53 and 64 %. These relationships are reflected in assignment of the *S. moellendorffii* KO into the CYP701C subfamily, *P. patens* is in the CYP701B whereas that of higher plant P450's

are in the CYP701A subfamily indicating evolutionary divergence. Cytochrome P450 nomenclature relies on evolutionary relationships with classification into the same family being indicative of belonging to a similar stage in the biosynthetic pathway [6]. The *S. moellendorffii* and *P. patens* KO are only distantly related to those known to operate in angiosperms as demonstrated by their classification into separate sub-families of the relevant CYP701 family of P450s (Figure 1).

SmKO cloning-SmKO1/CYP701C1 (JGI ID#-74427) and its allelic variant were cloned from *S. moellendorffii.* There are a few differences noted from the predicted sequence including a 9 nucleotide insertion at the N-terminal end and 6 nucleotide deletions in the middle of the gene. This is a reoccurring theme which was also identified for SmCPSKSL1 [22]. The ORF initiates with a valine instead of a methionine which is an anomaly for eukaryotes or may suggest the presence of an upstream start codon (Figure 2). However, the sequence was cloned using UTR primers and no upstream start codons could be identified. This was not a major concern as the modification position for the leader sequence was within the identified reading frame. Also the C-terminal end is much longer than that of KO sequences including that of *P. patens.* Truncations of the C-terminus relative to characterized P450s were constructed as well. Of interest SmKO1-1 and SmKO1-2 are 98 % identical to each other but differ at AtKO E316 where SmKO1-1 encodes a glutamate wheras SmKO1-2 encodes glycine. A glutamate residue is conserved for all KOs involved in primary metabolism, CYP701A8 which is involved in specialized metabolism has an alanine at this point (Figure 2). Expression of the mutant AtKO E316A shows altered regiochemistry to C3α producing 3α-Hydroxy kaurene: *ent*-kaurenoic acid in a 1: 9 ratio.

SmKAO- Three putative genes of the CYP88 family are identified in the Selaginella genome. Sequence alignments indicate that SmKAO1 is 79 % to SmKAO2 and 70 % to SmKAO3. SmKAO2 and SmKAO3 are 70 % identical to each other. The SmKAO genes are 60 to 63 % identical to those of higher plants. We cloned the SmKAO1 and SmKAO3 genes and the sequences were as in the predicted genome. The percentage identity in this family is higher than that of the traditional definition (Figure 3).

SmCPR- Three putative SmCPR are identified in the *Selaginella* genome. Phylogenetic analysis of the predicted sequences reveals that they are diverged from CPRs of higher plants and *P. patens*. SmCPRa is 69 % identical to SmCPRb and SmCPRc; SmCPRb and SmCPRc are 78 % identical to each other. The sequences are highly conserved as is seen for higher plants where it ranges from 63 % for the *Arabidopsis* paralogs to 90 % for the poplar paralogs. Cloned SmCPR1 was similar to predicted sequence. SmCPR3 has a stop codon after 1200 nucleotides from 4 of the clones analyzed. This could potentially be a pseudo gene as after the stop codon introns are present. 3' RACE needs to be carried out to confirm this portion of the sequence (Figure 3).

Expression Experiments

SmKO - The N terminal region of the *Arabidopsis* KO gene encodes for a transmembrane domain sequence that localizes it to the plastid envelope [23]. Deletion and modification of the sequence enables solubilization of the protein whilst maintaining the correct folding [2, 14-18, 24]. Thus, synthetic SmKO1/CYP701C1, was truncated and modified to enable dissociation of the protein from the membrane during expression in the *E. coli* based metabolic engineering system (Figure 2). AtR1 is a NADPH cytochromes P450 monooxygenase reductase from *A. thaliana* that is required for the reconstitution of P450

activity. Expression of synthetic recombinant SmKO1/CYP701C1 from GeneScript in the presence of upstream GGPP synthase, CPS and AtKS in the *E. coli* metabolic engineering system [12] did not lead to turnover of *ent*-kaurenoic acid (Figure 3). We also synthesized a gene from DNA2.0, as a previously inactive gene in rice was found to be functional. The genes are a pJ821 plasmid which when coexpressed in our metabolic engineering system reduces the flux to kaurene. Subcloning the gene into PCDF Dest/ AtR1 improved turnover to kaurene but it was still nonfunctional. Potential activity of SmKO rice diterpenes was also assessed but no substrate turnover was observed.

SmKAO - the SmKAO genes were modified as previously described. They were coexpressed in *E.coli* system with upstream genes that produce kaurenoic acid but did not exhibit any functionality.

SmCPR- Cytochrome C assays were carried out for SmCPR1 and SmCPR2 proteins and they were functional. Testing for functionality of CPR by coexpression with AtKO in the metabolic engineering system revealed that AtKO is capable of complete conversion of kaurene to kaurenoic acid in the presence or absence of a CPR. We would need to utilize an alternative P450 to assess product turnover in relation to SmCPR.

Discussion

SmKO - Selaginella KO is assigned to the CYP701C family whereas those from the higher plants are within the CYP701A subfamily indicating some evolutionary divergence. P450 nomenclature relies on evolutionary relationships. In most cases the classification system will follow the arbitrary rule of amino acid similarity with members within a family having 40 % amino acid identity and those within a subfamily having 55 % identity.

However, classification in the same family indicates functional equivalence or belonging to a similar stage of the biosynthetic pathway in this case GA metabolism[6].

Functional expression of SmKO is proving to be challenging. It is feasible that this is this is not the full length clone. 5' Race cloning to confirm the N-terminal end has not yielded results as yet. SmKO2 was initially disregarded because it did not have the conserved sequence, considering the inaccuracy of some of the genes it is worth attempting to isolate the gene.

As we were unable to obtain functional expression, one of the variables considered is CPR. CPRs have highly conserved sequences amongst species and even kingdoms and are able to complement for unrelated species. A common example is the use of *S. cerevisiae* WAT11 which expresses AtR1 has been used to characterize numerous P450s including ferulate 5 hydroxylase (F5H) involved in lignin biosynthesis from Selaginella [21, 25]. Even the KO from *P. patens*, was functionally characterized using AtR1 in *Pichia pastoris* [3]. However reductases can improve product outcome and/or alter product [26-28]. We coexpressed SmKO with AtR1, SmCPR1 and SmCPR2, while this did not yield any functionality it was an important variable to consider. An alternative approach to assess functionality is to overexpress SmKO in *A. thaliana ga3*-1 lines which is a gibberellin responsive dwarf deficient in kaurene oxidase activity[1].

SmKAO – the CYP88 family and subsequent steps in the GA pathway are first seen lycophyte and are absent in *P. patens*. It will be interesting to see the function of the CYP88E1 and CYP88F1. The CYP88D subfamily involved in glycchrizin exhibits divergence from the GA pathway [29]. Unfortunately we are unable to obtain functionality

from our analyses. Since we isolated genes from the plant we are relatively confident in our sequences. We will continue to seek avenues to determine functionality of CYPs.

In summary we have identified and isolated putative CYPs involved in GA biosynthesis but are facing challenges in functional expression.

References

1. Helliwell CA, Poole A, Peacock WJ, Dennis ES: **Arabidopsis ent-kaurene oxidase catalyzes three steps of gibberellin biosynthesis**. *Plant Physiol* 1999, **119**(2):507-510.

2. Morrone D, Chen X, Coates RM, Peters RJ: **Characterization of the kaurene oxidase CYP701A3, a multifunctional cytochrome P450 from gibberellin biosynthesis**. *Biochem J* 2010, **431**(3):337-344.

3. Miyazaki S, Katsumata T, Natsume M, Kawaide H: **The CYP701B1 of Physcomitrella patens is an ent-kaurene oxidase that resists inhibition by uniconazole-P**. *FEBS Lett* 2011, **585**(12):1879-1883.

4. Davidson SE, Smith JJ, Helliwell CA, Poole AT, Reid JB: **The pea gene LH encodes ent-kaurene oxidase**. *Plant Physiol* 2004, **134**(3):1123-1134.

5. Ko KW, Lin F, Katsumata T, Sugai Y, Miyazaki S, Kawaide H, Okada K, Nojiri H, Yamane H: **Functional identification of a rice ent-kaurene oxidase, OsKO2, using the Pichia pastoris expression system**. *Biosci Biotechnol Biochem* 2008, **72**(12):3285-3288.

6. Nelson DR, Schuler MA, Paquette SM, Werck-Reichhart D, Bak S: **Comparative genomics of rice and Arabidopsis. Analysis of 727 cytochrome P450 genes and pseudogenes from a monocot and a dicot**. *Plant Physiol* 2004, **135**(2):756-772.

7. Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, dePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA *et al*: **The Selaginella genome identifies genetic changes associated with the evolution of vascular plants**. *Science* 2011, **332**(6032):960-963.

8. Hirano K, Nakajima M, Asano K, Nishiyama T, Sakakibara H, Kojima M, Katoh E, Xiang H, Tanahashi T, Hasebe M *et al*: **The GID1-mediated gibberellin perception mechanism is conserved in the Lycophyte Selaginella moellendorffii but not in the Bryophyte Physcomitrella patens**. *Plant Cell* 2007, **19**(10):3058-3079.

9. Hayashi K, Horie K, Hiwatashi Y, Kawaide H, Yamaguchi S, Hanada A, Nakashima T, Nakajima M, Mander LN, Yamane H *et al*: **Endogenous diterpenes derived from ent-kaurene, a common gibberellin precursor, regulate protonema differentiation of the moss Physcomitrella patens**. *Plant Physiol* 2010, **153**(3):1085-1097.

10. Sun TP: **The molecular mechanism and evolution of the GA-GID1-DELLA signaling module in plants**. *Curr Biol* 2011, **21**(9):R338-345.

11. Jensen K, Møller BL: **Plant NADPH-cytochrome P450 oxidoreductases**. *Phytochemistry* 2010, **71**(2-3):132-141.

12. Cyr A, Wilderman PR, Determan M, Peters RJ: **A modular approach for facile biosynthesis of labdane-related diterpenes**. *J Am Chem Soc* 2007, **129**(21):6684-6685.

13. Swaminathan S, Morrone D, Wang Q, Fulton DB, Peters RJ: **CYP76M7 is an ent-cassadiene C11α-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice**. *Plant Cell* 2009, **21**:3315-3325.

14. Wang Q, Hillwig ML, Peters RJ: **CYP99A3: functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice**. *Plant J* 2011, **65**(1):87-95.

15. Wang Q, Hillwig ML, Okada K, Yamazaki K, Wu Y, Swaminathan S, Yamane H, Peters RJ: **Characterization of CYP76M5-8 indicates metabolic plasticity within a plant biosynthetic gene cluster**. *J Biol Chem* 2012, **287**(9):6159-6168.

16. Wang Q, Hillwig ML, Wu Y, Peters RJ: **CYP701A8: a rice ent-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism**. *Plant Physiol* 2012, **158**(3):1418-1425.

17. Wu Y, Hillwig ML, Wang Q, Peters RJ: **Parsing a multifunctional biosynthetic gene cluster from rice: Biochemical characterization of CYP71Z6 & 7**. *FEBS Lett* 2011, **585**(21):3446-3451.

18. Swaminathan S, Morrone D, Wang Q, Fulton DB, Peters RJ: **CYP76M7 is an ent-cassadiene C11alpha-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in rice**. *Plant Cell* 2009, **21**(10):3315-3325.

19. Wang Q, Hillwig ML, Peters RJ: **CYP99A3: Functional identification of a diterpene oxidase from the momilactone biosynthetic gene cluster in rice**. *Plant J* 2011, **65**(1):87-95.

20. Pompon D, Louerat B, Bronine A, Urban P: **Yeast expression of animal and plant P450s in optimized redox environments**. *Methods Enzymol* 1996, **272**:51-64.

21. Urban P, Mignotte C, Kazmaier M, Delorme F, Pompon D: **Cloning, yeast expression, and characterization of the coupling of two distantly related Arabidopsis thaliana NADPH-cytochrome P450 reductases with P450 CYP73A5**. *J Biol Chem* 1997, **272**(31):19176-19186.

22. Mafu S, Hillwig ML, Peters RJ: **A novel labda-7,13e-dien-15-ol-producing bifunctional diterpene synthase from Selaginella moellendorffii**. *Chembiochem* 2011, **12**(13):1984-1987.

23. Helliwell CA, Sullivan JA, Mould RM, Gray JC, Peacock WJ, Dennis ES: **A plastid envelope location of Arabidopsis ent-kaurene oxidase links the plastid and endoplasmic reticulum steps of the gibberellin biosynthesis pathway**. *Plant J* 2001, **28**(2):201-208.

24. Scott EE, Spatzenegger M, Halpert JR: **A truncation of 2B subfamily cytochromes P450 yields increased expression levels, increased solubility, and decreased aggregation while retaining function**. *Arch Biochem Biophys* 2001, **395**(1):57-68.

25. Weng JK, Akiyama T, Bonawitz ND, Li X, Ralph J, Chapple C: **Convergent evolution of syringyl lignin biosynthesis via distinct pathways in the lycophyte Selaginella and flowering plants**. *Plant Cell* 2010, **22**(4):1033-1045.

26. Lah L, Podobnik B, Novak M, Korošec B, Berne S, Vogelsang M, Kraševec N, Zupanec N, Stojan J, Bohlmann J *et al*: **The versatility of the fungal cytochrome P450 monooxygenase system is instrumental in xenobiotic detoxification**. *Mol Microbiol* 2011, **81**(5):1374-1389.

27. Park S, Kim YS, Rupasinghe SG, Schuler MA, Back K: **Rice P450 reductases differentially affect P450-mediated metabolism in bacterial expression systems**. *Bioprocess Biosyst Eng* 2013, **36**(3):325-331.

28. Ro DK, Ehlting J, Douglas CJ: **Cloning, functional expression, and subcellular localization of multiple NADPH-cytochrome P450 reductases from hybrid poplar**. *Plant Physiol* 2002, **130**(4):1837-1851.

29. Seki H, Ohyama K, Sawai S, Mizutani M, Ohnishi T, Sudo H, Akashi T, Aoki T, Saito K, Muranaka T: **Licorice beta-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin**. *Proc Natl Acad Sci U S A* 2008, **105**(37):14204-14209.

Figures

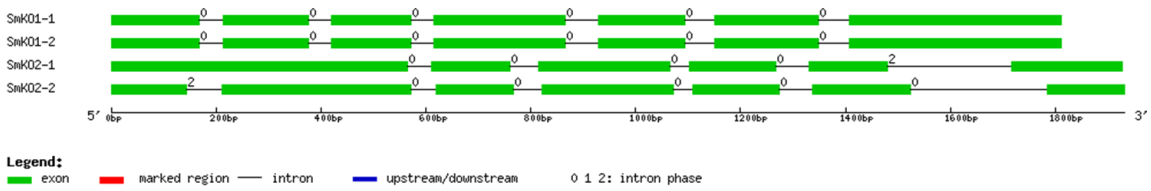
Table 1: Primers used for cloning

Primer Name	**Primer**
SmKO1 5'UTR F SmKO1 3' UTR R	TGCAGGGTCTTTTCTTCCTCCTG GACCACCACGAAGAACACGAAGA
SmKAO1 5' SmKAO1 3'	caccATGAATCTCAAGTGGGCG TTATCTACGTCTGCGAAC
SmKAO3 5' SmKAO3 3'	caccATGGCAGTGGCCTTGATC CTAGACAAGTGCTCGCTTCG
SmCPR1 5' FL SmCPR1 5' ATG SmCPR1 3'	caccGGTCTCGAGTCACTCCCGGTTCTCCTGTCCG GGATCTCCCAAGCCAATATGATTTTATCC
SmCPR3 5' SmCPR3 3'	caccATGGCTTCCGCGACTATTCCCCCG CTACGAACGAACCTCTTGTTTCAC
RT.Sm6PGD.F RT.Sm6PGD.R	GCTCATGGATCCCGAGTTT ATCCACTCGCTCGTAGGTGT

Figures

Figure 1: A) Intron/exon organization of four putative full length SmKO genes B) Phylogenetic Tree

A)



B)

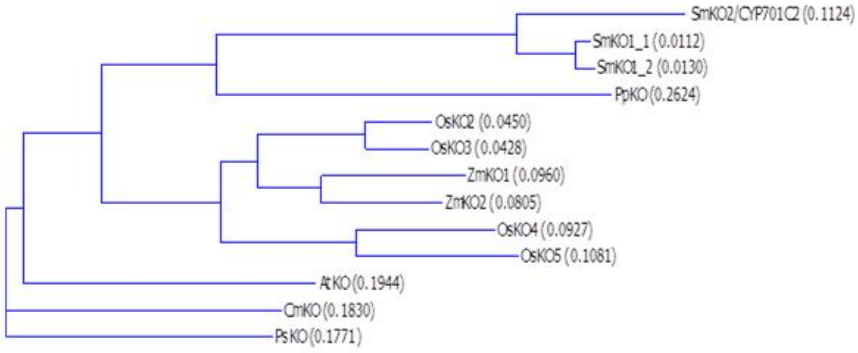
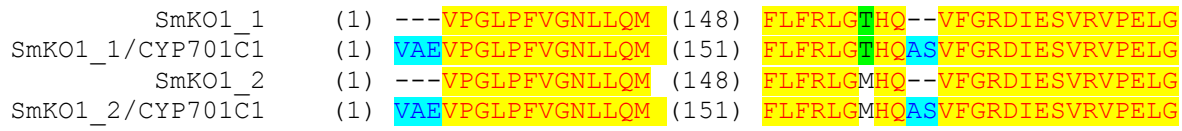
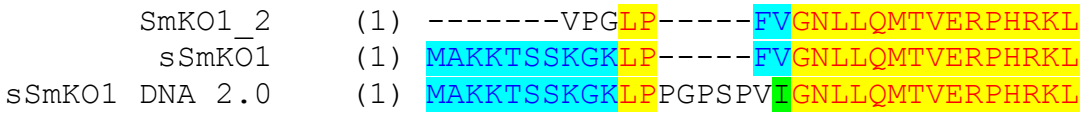


Figure 2: SmKO alignments A) SmKO 1_1 &2 cDNA sequences aligned against predicted sequences SmKO1_1 & 2/CYP701C predicted sequences B) sSmKO1 modified according to CYP76M5-8 (Swaminathan et al. 2009) and sSmKO1_2DNA2.0 modified according to AtKO (Morrone et. al 2010) C) Conserved glutamate at AtKO 316 for KO involved in primary metabolism and alanine for OsKO4 and OsKO5 which are inducible and involved in phytoalexins biosynthesis

A.



B.



C.

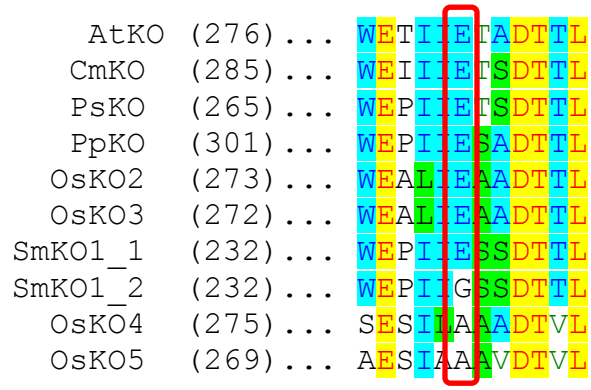
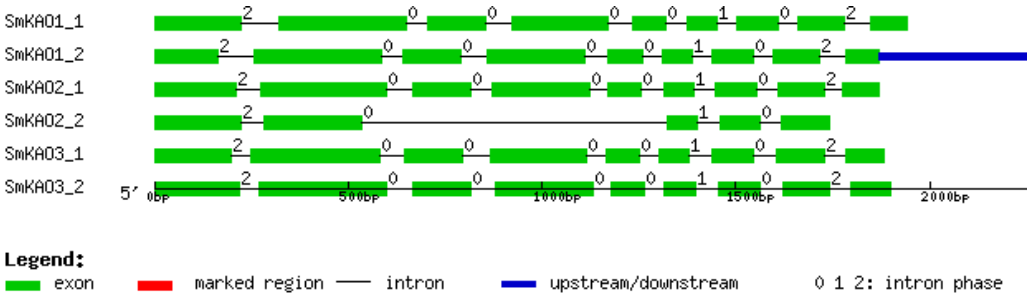


Figure 3: A) Intron/exon organization of putative full length SmKAO genes B) Phylogenetic association

A.



B

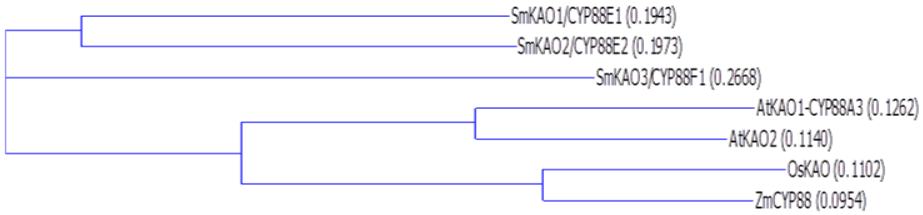
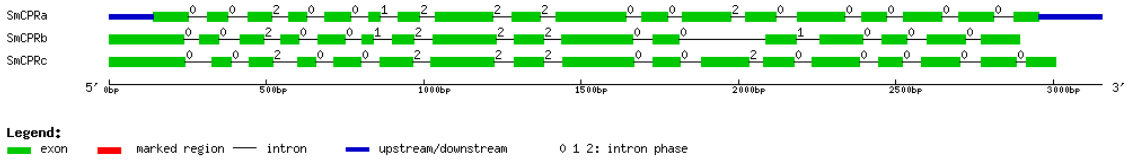
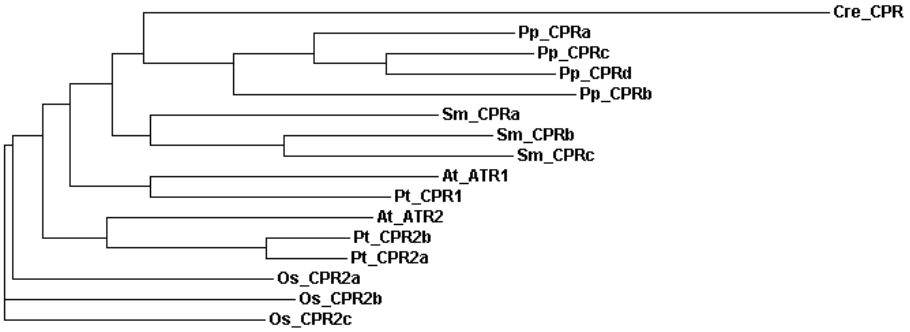


Figure 4: A) Intron/exon organization of putative full length SmCPR genes B) Phylogenetic association

A.



B.



CHAPTER V: CONCLUSION AND FUTURE DIRECTIONS

The dissertation is a report of a novel diterpene synthase that terminates deprotonation at C7 instead of C8 methyl which is typical for class II reactions. The class I site carries out replacement of a diphosphate with a hydroxyl instead of cyclization as with most class 1 enzymes. The bifunctional enzyme produces labda-7, 13*E*-dien-15-ol a widely reported compound, but this was the first report of its synthase [1].

We isolated SmCPSKSL1 when looking for a kaurene synthase to determine when the shift from bifunctional to monofunctional TPS occurred in GA biosynthetic pathway in land plants. To date there is no published work upstream TPS genes in *S. moellendorffii* involved in GA biosynthesis. Recent work in the laboratory shows that there is conserved TT motif upstream of the DDxxD class 1 motif in kaurene synthase genes. There are 18 putative TPS in *S. moellendorffii*, of which four have been identified. The motif can help to narrow down which genes to focus on. From the predicted sequences there is a single bifunctional gene (JGI ID – 412139) with the identified TT motif which we have managed to isolate. There are 2 KSLs (JGI ID 450730 & 413764) with the associated TT motif which we can renew our efforts isolate the genes.

Even though our primary interest is kaurene synthase genes the number of TPS in the genome as well as the characterized TPS gene indicates great chemical diversity within *Selaginella*. Functional characterization of these TPS genes will improve our understanding of this chemodiversity as well as the evolution of terpene synthases. An example of this discovery of SmMTPS genes that have been acquired from microbes through horizontal gene transfer, followed by gene duplication and neofunctionalization resulting in a new class of TPS in land plants which is evolutionarily distinct from that of higher plants [2, 3].

The remainder of the dissertation is an investigation on the evolution of KOs in higher plants and land plants. We demonstrated the evolutionary based enzyme promiscuity of AtKO relative to that of OsKO2. AtKO has an ability to recognize multiple LRDs whereas rice because of an LRD rich chemical background has evolved to be more specific for kaurene. Substrates of *ent*-stereochemistry are catalyzed at C19, but due to differences in the C/D region conversion to the acid is limited and the intermediate alcohol an aldehyde are also produced especially for the pimaranes. Substrates of *syn* and normal stereochemistry are catalyzed on C2 or C3 instead of C19 (as in kaurene) suggesting a change in orientation of the substrate towards the catalytic heme center. To complete the project we would like to obtain relative kinetic constants to compare affinity for the different substrates in comparison to kaurene [4, 5]. The project illustrates the versatility of our *E.coli* based metabolic engineering system enabling functional characterization, multi substrate comparison as well as upscaling for increased production of compounds for structural analysis by NMR. It would also be interesting to view the physiological and metabolite effects of expression of AtKO in rice OsKO2 knockout line. The changes would be expected to be subtle if any. In line with this project, preliminary work on enzymatic determinants underlying the change in regiospecificity has been carried out. Mutations in SRS4 of AtKO (E316A) showed a change in regiospecificity of AtKO (unpublished work). The mutant is able to hydroxylate at the C3α position. Previous studies in the CYP4 family suggest the glutamate is responsible for covalent linkage to the heme to assist in oxidation. We envision a similar scenario here and are in the process of investigating possible covalent ester linkage to the heme.

To further understand the evolution of the GA pathway in land plants we are investigating the function of putative P450s involved in GA metabolism. *S. moellendorffii*

has a functional GA system and to date only the downstream enzymes GA3 and GA20 oxidases have been functionally characterized. Functional characterization of the putative P450s is important to demonstrate biochemical evolution of the pathway in land plants.

We have synthesized the putative P450s and expressed them in our metabolic engineering system with little success. It is worthwhile to intensify our cloning efforts to ensure that our terminal ends for SmKO1 and SmKO2 are indeed accurate. Yeast expression will be revisited using pYEDP60 plasmid for both SmKO and SmKAO genes. Recent work done in our laboratory shows improved expression when using ligation as compared to LR reaction. As all my constructs were performed through LR cloning, I will construct my CYPs using the traditional ligation reaction to see if I can obtain expression. As an alternative to microbial expression, I will overexpress SmKO in *A. thaliana ga3-1* a line deficient in KO gene [6].

In an attempt to attain expression we isolated reductases from *S. moellendorffii*. The reductases from *P. patens* and *S. moellendorffi* are evolutionarily diverged from that of other land plants. For this reason it would be worthwhile to characterize the genes, to assess whether the reduction for cytochrome C (an artificial substrate for CPR) by SmCPRa and SmCPRc NADPH or NADH dependent.

In summary we isolated and functionally characterized a novel bifunctional diterpene synthase that produces labda-7, 13 *E*-dien-15-ol a widely reported compound, but this was the first report of its synthase. We explored substrate specificity and product outcome of evolutionary related enzymes from higher plants (Arabidopsis and rice) utilizing an *E. coli* based metabolic engineering system to express recombinant cytochromes p450 against a wide range of substrates.

References

1. Mafu S, Hillwig ML, Peters RJ: **A novel labda-7,13e-dien-15-ol-producing bifunctional diterpene synthase from Selaginella moellendorffii**. *Chembiochem* 2011, **12**(13):1984-1987.
2. Li G, Köllner TG, Yin Y, Jiang Y, Chen H, Xu Y, Gershenzon J, Pichersky E, Chen F: **Nonseed plant Selaginella moellendorffi [corrected] has both seed plant and microbial types of terpene synthases**. *Proc Natl Acad Sci U S A* 2012, **109**(36):14711-14715.
3. Weng JK, Noel JP: **Chemodiversity in Selaginella: a reference system for parallel and convergent metabolic evolution in terrestrial plants**. *Front Plant Sci* 2013, **4**:119.
4. Nevarez DM, Mengistu YA, Nawarathne IN, Walker KD: **An N-aroyltransferase of the BAHD superfamily has broad aroyl CoA specificity in vitro with analogues of N-dearoylpaclitaxel**. *J Am Chem Soc* 2009, **131**(16):5994-6002.
5. Pi N, Leary JA: **Determination of enzyme/substrate specificity constants using a multiple substrate ESI-MS assay**. *J Am Soc Mass Spectrom* 2004, **15**(2):233-243.
6. Helliwell CA, Sheldon CC, Olive MR, Walker AR, Zeevaart JA, Peacock WJ, Dennis ES: **Cloning of the Arabidopsis ent-kaurene oxidase gene GA3**. *Proc Natl Acad Sci U S A* 1998, **95**(15):9019-9024.

ACKNOWLEDGEMENTS

Working on the PhD has been a wonderfully fulfilling and overwhelming journey which was made possible by a combination of excellent mentors and supportive family and friends.

Firstly, I would like to express my sincere gratitude to Reuben who guided me through my journey in a very supportive and patient manner. I appreciate all his contributions in terms of selfless time, alternate stimulating ideas and exposing me to wide range of projects. Also he has cultivated a wonderful working environment which encourages one to give your best always.

In addition, I would like to thank my POS committee Dr. Alan Myers, Dr. Mark Hargrove, Dr. Basil Nikolau and Dr. David Hannapel for their comments, analysis and invaluable feedback on my work.

I would also like to thank Meimei Xu for her mentorship, advice, ensuring smooth operation of the lab and endless supply of chocolate. The Peters' group past and present members have been a source of advice and discussion, collaboration and a source of friendship. I am especially grateful to Yisheng Wu who made the lab experience that much better.

My time in Ames was enriched by the many groups of people who became a part of my life. This dissertation would not have been possible without the support of my friends who looked after my daughter so that I could attend class or prepare for examinations. Special thanks go to the Peters family (Namibia),the Lichoros, Rodah and Alex , Viwe and Robson.

I would like to thank my family for their love and encouragement. Most of all I would like to thank Nandi for being the most awesome daughter, for working on weekends, putting a smile on my face and tolerating me when at my worst, continually reminding me that there is more to life. Above all I give thanks and praise to Almighty God.