

RICE UNIVERSITY

Investigation of Triterpene Biosynthesis in *Arabidopsis thaliana*

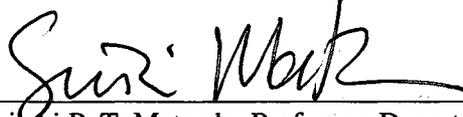
by

Mariya D. Kolesnikova

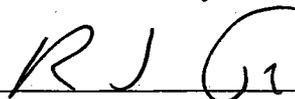
A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

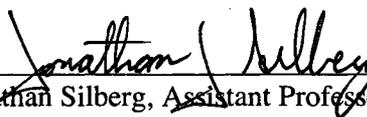
APPROVED, THESIS COMMITTEE:



Seitchi P. T. Matsuda, Professor, Department
Chair
Department of Chemistry
Department of Biochemistry and Cell Biology



Ronald J. Parry, Professor
Department of Chemistry
Department of Biochemistry and Cell Biology



Jonathan Silberg, Assistant Professor
Department of Biochemistry and Cell Biology

HOUSTON, TEXAS

May 2008

UMI Number: 3362344

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

UMI[®]

UMI Microform 3362344
Copyright 2009 by ProQuest LLC
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

Investigation of Triterpene Biosynthesis in *Arabidopsis thaliana*

By

Mariya D. Kolesnikova

This thesis describes functional characterization of three oxidosqualene cyclase genes (*At1g78955*, *At3g45130*, and *At4g15340*) from the model plant *Arabidopsis thaliana* that encode enzymes with novel catalytic functions. Oxidosqualene cyclases are a family of membrane proteins that convert the acyclic substrate oxidosqualene into polycyclic products with many chiral centers. The complex mechanistic pathways and relevant catalytic motifs can be elucidated through judicious applications of mutagenesis, heterologous expression in combination with a genome mining approach, and protein modeling. Functional characterization of oxidosqualene cyclases allows improved understanding on how these proteins guide catalytic reactions and how protein-substrate interactions affect the reaction outcome, as well as identification of triterpenes with novel structures and stereochemistry.

This work describes characterization of *Arabidopsis* oxidosqualene cyclases, including the first plant lanosterol synthase (LSS1), an enzyme with novel catalytic motifs different from those previously observed in animal, fungal, and trypanosomal lanosterol synthases, establishing that plant lanosterol synthases comprise a catalytically distinct class of lanosterol synthases. Phylogenetic analysis reveals that lanosterol synthases are broadly distributed in eudicots but evolved independently from those in animals and fungi. Discovery of plant lanosterol synthase also suggests lanosterol as

precursor for plant 4,4-dimethyl- Δ^8 sterols. Additional mutagenesis experiments on *Arabidopsis* lanosterol synthase (Asn477His and Val481Ile) allowed for introduction of cycloartenol activity in a lanosterol synthase background, providing the best example of engineered biosynthesis of cyclopropyl structures known to date.

This thesis also describes the first enzyme (camelliol C synthase, CAMS1) that efficiently blocks B-ring formation to make a monocyclic triterpene camelliol C. Phylogenetic analysis reveals that this cyclase has evolved from enzymes that generate pentacycles, and sequence comparison between oxidosqualene cyclases with different catalytic functions allowed for identification of key residues that increases steric bulk in the active site to promote monocycle formation.

Finally, this thesis describes an enzyme arabidiol synthase (PEN1) that produces the tricyclic triterpene diol arabidiol. Analysis of the arabidiol structure and characterization of numerous minor products of arabidiol synthase, including several novel compounds, resulted in formulation of a general rule for water addition in triterpene biosynthesis and an explanation for the domination of deprotonation over water addition in triterpene biosynthesis.

Acknowledgments

First of all I would like to thank Professor Seiichi P.T. Matsuda, my advisor, for everything he has done for me during these years in graduate school. For accepting me into his Lab, for allowing to work on interesting projects, for giving me freedom in research and supporting my steps aside from main research directions, for promoting a self-managing environment, for giving useful criticisms when I deserved it, for giving second chances, for supporting, motivating, encouraging, and providing ideas. I deeply appreciate it.

I want to thank Professor Ronald Parry for being my on Thesis Committee, for his advice throughout the years and for providing very useful lectures on secondary metabolism. I would also like to thank Professor Jonathan Silberg, who also kindly agreed to be on my Thesis Committee and gave valuable advice regarding my thesis. I want to thank Professor Bonnie Bartel for providing valuable insights in my work.

Of course I want to thank Dr. William K. Wilson for his incredible help with NMR, with manuscripts, for giving valuable ideas, for being a very genuine and kind person and a very good friend. Thank you, Bill, without you everything would be different. I also want to thank Professor Quanbo Xiong, for teaching me my baby steps in the Lab, for being very supportive and being a great friend during all these years.

I want to thank Caroline (Carrie) McNeil for so much! For being my best friend, for providing constant support, for being here for me all the time, for making all these years so much easier. I want to thank Carrie for her great help with my thesis. I cannot say enough to thank you, my dear friend.

I want to acknowledge two undergraduate students who worked with me on the projects, who also were very good friends and colleagues, who made the best team anyone could ever wish for, David A. Lynch and Allie C. Obermeyer.

I want to thank my current and former colleagues Dr. Gia Fazio for her friendship and help during my first year, Dr. Renee LeClair for providing me a “lab boot camp”, and making me stronger, Yulia Ivanova for bringing part of homeland into the lab, Dr. Hui Shan for being a nice and kind colleague, Dr. Dereth Phillips for teaching me plant work and her positive mood no matter what, Dr. Jeanne Rasbery for teaching me mRNA work and sharing stories. I want to thank Carl Onak for mother liquor and unforgettable stories about his grandparents. I thank my other colleagues Alyssa Baevich, Pietro Morlacchi, Dorianne Castillo, and the other undergraduate students for creating a nice working environment.

I also want to thank my parents Kolesnikov Dmitriy and Kolesnikova Irina, my sister Anastasiya, and my grandparents for their continuous love and support. I also would like to thank my friends Ekaterina Pronina and Alexei Tcherniak for being like our family here. The last, but not least, I want to thank my biggest judge and my greatest supporter - my husband Artur Izmaylov for being my everything.

Table of Contents

Abstract	ii
Acknowledgments	iv
Table of Contents	vi
List of Figures	xii
CHAPTER 1: Introduction and Background	1
1.1 Triterpene Biosynthesis	1
1.2 Triterpene Discovery through Surveys of Natural Sources	2
1.3 Genome Mining to Uncover Triterpene Products	3
1.4 Oxidosqualene Cyclases	5
1.5 Protosteryl-Type Oxidosqualene Cyclases in Plants, Animals, and Fungi	6
1.5.1 Sterol Biosynthesis in Animals, Fungi, and Plants.	6
1.5.2 Catalytic Differences in Cycloartenol and Lanosterol Biosynthesis	8
1.6. Dammarenyl-type Oxidosqualene Cyclases in Plants	13
1.7. Oxidosqualene Cyclases from <i>Arabidopsis</i>	16
CHAPTER 2: Materials and Methods	19
2.1 Materials	19
2.2 Nuclear Magnetic Resonance (NMR)	19
2.3 Gas Chromatography – Mass Spectrometry (GC-MS)	20

2.4 Gas Chromatography-Flame Ionization Detection (GC-FID)	20
2.5 High-Performance Liquid Chromatography (HPLC)	20
2.6 UV Spectroscopy	21
2.7 Centrifugation	21
2.8 Incubations	21
2.9 Polymerase Chain Reaction (PCR)	22
2.10 DNA Plasmid Purification	23
2.11 DNA Restriction Digestion	24
2.12 DNA Gel Analysis	25
2.13 DNA Ligation	25
2.14 Bacterial Media	26
2.15 <i>E. coli</i> Transformation	26
2.16 Yeast Strains Used for Expression of Oxidosqualene Cyclases	27
2.17 Yeast Media	28
2.18 Yeast Transformation	29
2.19 (\pm)-2,3-Oxidosqualene Synthesis	30
2.20 Preparation of 20 \times (\pm)-2,3-Oxidosqualene Solutions	31
2.21 Small Scale Oxidosqualene Cyclase Assay	32
2.22 Large Scale Oxidosqualene Cyclase Assay	33
2.23 In vivo Assay	33
2.24 Isolation of Triterpene Alcohols from the Media with Resin	34
2.25 Purification of Triterpene Alcohols by Silica Gel Chromatography	34
2.26 Purification of Triterpene Alcohols by Preparative Thin Layer	

Chromatography	35
2.27 Trimethylsilyl (TMS) Derivatization of Triterpene Alcohols	35
CHAPTER 3: Cloning and Characterization of Lanosterol Synthase from <i>Arabidopsis thaliana</i> : Discovering Lanosterol Biosynthesis in Plants	
3.1 Previous Work	36
3.2 Experimental Procedures	37
3.2.1 Cloning of the Full-Length <i>At3g45130</i>	37
3.2.2 Heterologous Expression of <i>At3g45130</i> in Yeast	38
3.3 Results	39
3.3.1 Heterologous Expression of <i>Arabidopsis</i> Lanosterol Synthase in Yeast	39
3.4 Discussion	42
CHAPTER 4: Mutagenesis of Lanosterol Synthase to Biosynthesize Parkeol and Cycloartenol	
4.1 Experimental Procedures	49
4.1.1 Construction of <i>Arabidopsis</i> Lanosterol Synthase Asn477His/Val481Ile Double Mutant	49
4.1.2 Construction of <i>Arabidopsis</i> Lanosterol Synthase Asn477His and Val481Ile Single Mutants	50

4.1.3 Yeast Transformations	50
4.1.4 In Vitro Characterization of <i>Arabidopsis</i> Lanosterol Synthase Asn477His and Val481Ile Single Mutants and the Asn477His/Val481Ile Double Mutants	51
4.2 Results	52
4.2.1 Functional Characterization of <i>Arabidopsis</i> Lanosterol Synthase Val481Ile and Asn477His Single Mutants	52
4.2.2 Functional Characterization of <i>Arabidopsis</i> Lanosterol Synthase Asn477His/Val481Ile Double Mutant	53
4.2.3 Complementation Studies of <i>Arabidopsis</i> Lanosterol Synthase Mutants	53
4.3 Discussion	53
CHAPTER 5: Cloning and Characterization of Camelliol C Synthase From <i>Arabidopsis</i> <i>thaliana</i> : an Enzyme That Makes Monocycles Evolved From Those That Make Pentacycles	
5.1 Experimental Procedures	59
5.1.1 Cloning and Subcloning of <i>At1g78955</i> (<i>LUP3</i> , <i>CAMSI</i>)	59
5.1.2 Functional Characterization of the Camelliol C Synthase by Heterologous Expression in Yeast	60
5.2 Results	62
5.2.1 Functional Characterization of the Camelliol C Synthase	62

5.3 Discussion	65
CHAPTER 6: Cloning and Characterization of Arabidiol Synthase:	
Water Addition in Triterpene Biosynthesis	69
6.1 Experimental Procedures	69
6.1.1 Cloning and Subcloning of <i>At4g15340</i> (<i>PEN1</i>)	69
6.1.2 Functional Characterization of the Arabidiol Synthase by Heterologous Expression in Yeast: Analysis of the Dominant Compound Arabidiol	70
6.1.3 Degradation of the Major Product to the Corresponding Lactone	72
6.1.4 Purification of Minor Products	73
6.2 Results	74
6.2.1 Analysis and Structure Determination of the Major Product Arabidiol	74
6.2.2 Analysis of Media for Triterpene Products	76
6.2.3 Analysis and Structure Determination of Minor Products	77
6.2.4 Spectral Analysis of Arabidiol Synthase Minor Compounds	86
6.2.4 Quantitative Characterization of Arabidiol Synthase from In Vitro Analysis	89
6.3 Discussion	91
CHAPTER 7: Conclusions	
Appendix A: Additional Figures	101
Appendix B: List of Abbreviations	129

References

List of Figures

Scheme 1.1. Mechanism of enzymatic oxidosqualene cyclization.

Scheme 1.2. Biosynthesis of sterols in plants, animals, and fungi.

Figure 1.1. Sequence comparison of lanosterol synthases and cycloartenol synthases.

Scheme 1.3. Oxidosqualene cyclization by lanosterol and cycloartenol synthase.

Table 1.1. Percentage product composition of *A. thaliana* cycloartenol synthase (AtCAS1) mutants and *S. cerevisiae* lanosterol synthase (ScERG7) mutants.

Figure 1.2. Selected active site residues of human lanosterol synthase. Red dots correspond to ordered water molecules.

Table 1.2. Comparison of the corresponding active site residues between human lanosterol synthase (HsLSS), *Arabidopsis* cycloartenol synthase (AtCAS1), and yeast lanosterol synthase (ScERG7).

Scheme 1.4. Cation formation in protosteryl and dammarenyl type cyclases.

Figure 1.3. Computed phylogenetic relationships between *Arabidopsis* oxidosqualene cyclases.

Scheme 1.5. Biosynthesis of triterpenes in *Arabidopsis*.

Figure 3.1. GC-MS analysis of the total crude NSL from SMY8[pLH1.25] after induction with galactose.

Figure 3.2. Comparison of amino acid residues in lanosterol synthases and cycloartenol synthases.

Figure 3.3. Phylogenetic analysis of cycloartenol synthases and lanosterol synthases from plants.

Scheme 4.1. Cyclization of oxidosqualene by AtLSS1 Asn477His, Val481Ile, and Asn477His/Val481Ile mutants.

Table 4.1. Percentage product composition of *Arabidopsis* cycloartenol synthase (AtCAS1), *Arabidopsis* lanosterol synthase (AtLSS1) and yeast lanosterol synthase (ScERG7) mutants.

Figure 5.1. GC-MS analysis of the crude extract of RXY6[pMDK4.8] incubated with racemic oxidosqualene: total ion chromatogram, oven temperature 260 °C.

Figure 5.2. GC-MS analysis of the crude NSL of SMY8[pMDK4.8]: total ion chromatogram, oven temperature 270 °C.

Figure 5.3. ¹H NMR spectrum of the first PTLC fraction from the RXY6[pMDK4.8] in vitro reaction.

Figure 5.4. GC-MS analysis of fraction 5 from column separation of the crude SMY8[pMDK4.8] NSL: total ion chromatogram of underivatized triterpenes, oven temperature 260 °C.

Scheme 5.1. CAMS1 cyclizes oxidosqualene to monocyclic triterpenes and the pentacycle β-amyrin.

Figure 5.5. Phylogenetic analysis of CAMS1 and related plant oxidosqualene cyclases.

Figure 5.6. Partial alignment of oxidosqualene cyclase amino acid sequences. Active site residues alignment is based on a comparison with human lanosterol synthase.

Scheme 6.1. Synthesis of lactone.

Figure 6.1. Electron impact mass spectrum of the underivatized triterpene diol (arabidiol).

Figure 6.2. Structure of arabidiol with atoms numbered. Underlined are upfield methyl singlets (in ppm).

Figure 6.3. Total ion chromatogram of the medium extract.

Scheme 6.2. Scheme of chromatographic separation of NSL extract from a 30-L culture of SMY8[pMDK3.6].

Table 6.1. Triterpene alcohols identified by ^1H NMR analysis of column fraction 8 after HPLC separation. ξ indicates the stereochemistry was not determined.

Figure 6.4. Total ion chromatogram from GC-MS of fraction 9 from silica gel chromatography (no derivatization; 260 °C oven temperature), illustrating the multiplicity of triterpene alcohol products from arabidiol synthase. It should be noted that this chromatogram does not represent the relative abundance of the enzymatic products because fraction 9 represents only a portion of the monohydroxy triterpene products.

Figure 6.5. Electron impact mass spectrum of underivatized 14-epithalianol.

Table 6.2. Triterpene alcohols in HPLC-purified column fraction 9, detected by ^1H NMR analysis.

Figure 6.6. Triterpenes identified in a 30-L extract from in vivo expression of arabidiol synthase.

Scheme 6.3. Mechanistic pathways of product formation in arabidiol synthase.

Figure 6.7. Predicted steric conflict of water addition during formation of arabidiol. Black dots show positions of dummy oxygen atoms that would result in deprotonation from positions marked in red. The blue dot corresponds to a position of the dummy oxygen

atom that would successfully quench the carbocation and result in the formation of arabidiol.

Figure 6.8. Triterpene products formed by water addition.

CHAPTER 1

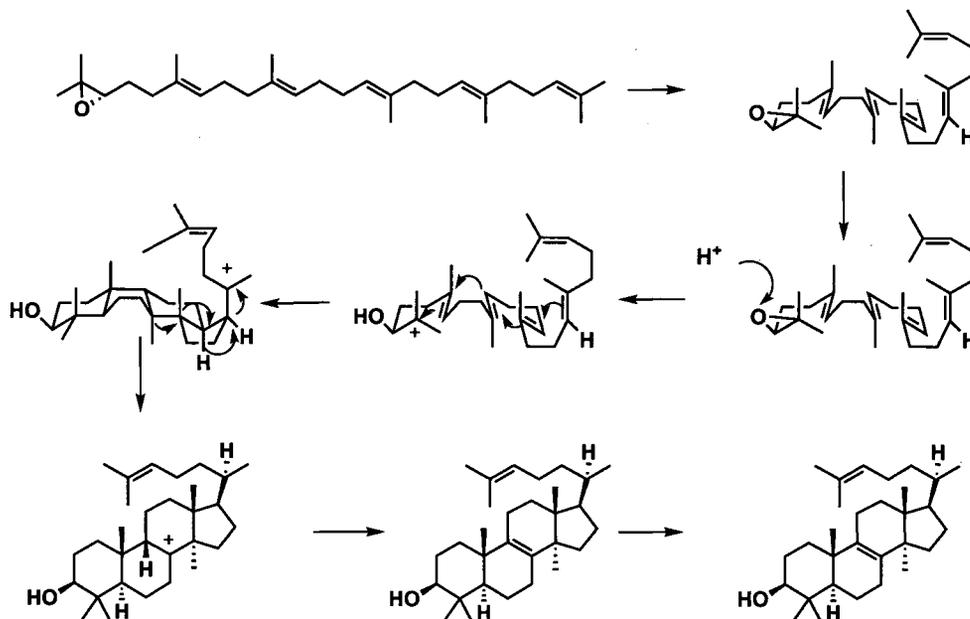
Introduction and Background

1.1 Triterpene Biosynthesis

Triterpenes and their derivatives, triterpenoids, comprise a large family of natural products derived from the C₃₀ acyclic molecule 2,3-oxidosqualene or, more rarely, squalene.¹ Triterpene synthases are a family of membrane proteins that mediate cyclization, rearrangement, and neutralization reactions to convert the acyclic substrate into triterpene products that can be polycyclic and contain many chiral centers. Triterpene synthases can be categorized as oxidosqualene cyclases (OSCs) and squalene-hopene cyclases (SHCs), depending on the nature of their native substrate. These enzymes perform cyclization reactions in a similar manner, differing in the initial protonation step. The substrate enters the enzyme active site, which guides initial pre-folding of the substrate to either chair-chair-chair or chair-boat-chair conformation. OSCs then protonate the epoxide, and SHCs protonate the 2,3-double bond to form a carbocation intermediate. The intermediate further undergoes a series of annulations, carbocation rearrangements (including possible ring expansions), 1,2-hydride and methyl shifts, deprotonation, or an alternative termination of the cyclization reaction with water addition. The product exits the active site (Scheme 1.1).^{2,3}

The enzyme-guided reactions promote the biosynthesis of over 100 diverse structures with different numbers and sizes of rings, functionalities, and stereochemistries.¹ The most abundant triterpenes found in nature are tetra- and pentacyclic protosteryl derivatives, including lanosterol^{4,5} and cycloartenol^{6,7} that are involved in the biosynthesis of sterols and steroids. Plants also produce numerous tetra-

and pentacyclic non-sterol triterpenes including β -amyrin^{8,9} and lupeol¹⁰ that are further utilized to give secondary metabolites: triterpenoids and triterpene saponins.^{11,12} The pentacyclic β -amyrin or its derivative β -amyrin saponins (modified triterpeneoids with an added sugar moiety) are often present in specific tissues¹³ or as components of plant waxes,^{14,15} which are important barriers that protect stems and other tissues from various environmental challenges such as infections. Some saponins have demonstrated potent antifungal activity.^{16,17} Lupeol was also frequently found in plant waxes¹⁷ and root nodules¹⁸ and both lupeol and its metabolites are actively studied for potential anti-cancer activity.¹⁹ Triterpenes with three rings or less have been found more rarely.¹



Scheme 1.1. Mechanism of enzymatic oxidosqualene cyclization.

1.2 Triterpene Discovery through Surveys of Natural Sources

The traditional approach to natural product discovery involves extraction and characterization of a product from a natural source, and was efficiently used to identify

hundreds of thousands of compounds. Many natural products, including valuable pharmaceuticals like artemisinin isolated from sweet wormwood (*Artemisia annua L*), paclitaxel (known as Taxol[®]) isolated from the pacific yew tree (*Taxus brevifolia*), and penicillin isolated from mold (*Penicillium notatum*) were found and extracted from natural sources. One possible limitation of this method is the compound must be present in sufficient amounts for detection and characterization. Classical approaches mandated that it have chemical and chromatographic properties that would allow facile detection in a crude mixture, but more modern approaches rely on activity screens. Advances in analytical methods and increasing sensitivity of existing techniques have improved the quality of the analysis of natural products and allowed for the detection of novel compounds present at much lower amounts than previously seen.

Application of this approach to *Arabidopsis thaliana*,^{15,20,21,22,23,24} allowed for the discovery of five triterpenes, including cycloartenol and three non-sterol lupeol, taraxasterol, β -amyirin, and α -amyirin.

1.3 Genome Mining to Uncover Triterpene Products

In recent years the genome mining approach has shown itself as a very valuable tool for discovery of new natural products from a number of sources, even those that were previously considered exhausted.²⁵ This strategy involves application of information obtained from genomes and available bioinformatics tools, such as genome databases and Basic Local Alignment Search Tool (BLAST) available through the National Center for Biotechnology Information (NCBI) website,²⁶ with application of molecular biological techniques and chemical methods. This approach becomes extremely important when a

targeted natural product is not easily accessible from its natural source due to limited production or production specific to certain conditions or tissue. For example, the maize sesquiterpene synthase TPS10 is expressed in response to herbivore attack.²⁷ Genome mining is also useful when a compound is immediately modified after production and cannot be easily detected as the unmodified intermediate.

Progress in sequencing technologies and bioinformatics tools provides increasing amounts of genomic information. The genome of *A. thaliana* was completely sequenced in 2000, and genomes of many other organisms, for example, rice (*Oryza sativa*) and black cottonwood (*Populus trichocarpa*) have been completed recently.²⁸ Thirteen oxidosqualene cyclases have been identified in the *Arabidopsis* genome, and over ten are predicted from each *Oryza sativa*. This data suggests the tremendous potential for the discovery of triterpenes and triterpenoids in plants by utilizing the genome mining approach.

Once obtained using bioinformatics tools, genes of interest can be then studied in known microbial expression systems such as *E. coli* or yeast. Further application of metabolic engineering approaches can significantly improve production of targeted compounds and avoid contamination with undesirable host products. Expression in heterologous hosts allows for scaling up production of the enzymes and therefore increasing the amounts of compounds produced without sacrificing large amounts of natural sources. This also provides better access to minor compounds present at significantly lower amounts than the major product.

1.4 Oxidosqualene Cyclases (OSCs)

Oxidosqualene cyclases have been found in mammals, plants, fungi, protists, and bacteria, and have been actively studied in past years.^{21,29} Analysis and characterization of both oxidosqualene cyclases and their products are valuable not only for studying triterpene biosynthesis, but they also help to understand how proteins guide catalytic reactions and how protein-substrate interactions affect the reaction outcome.

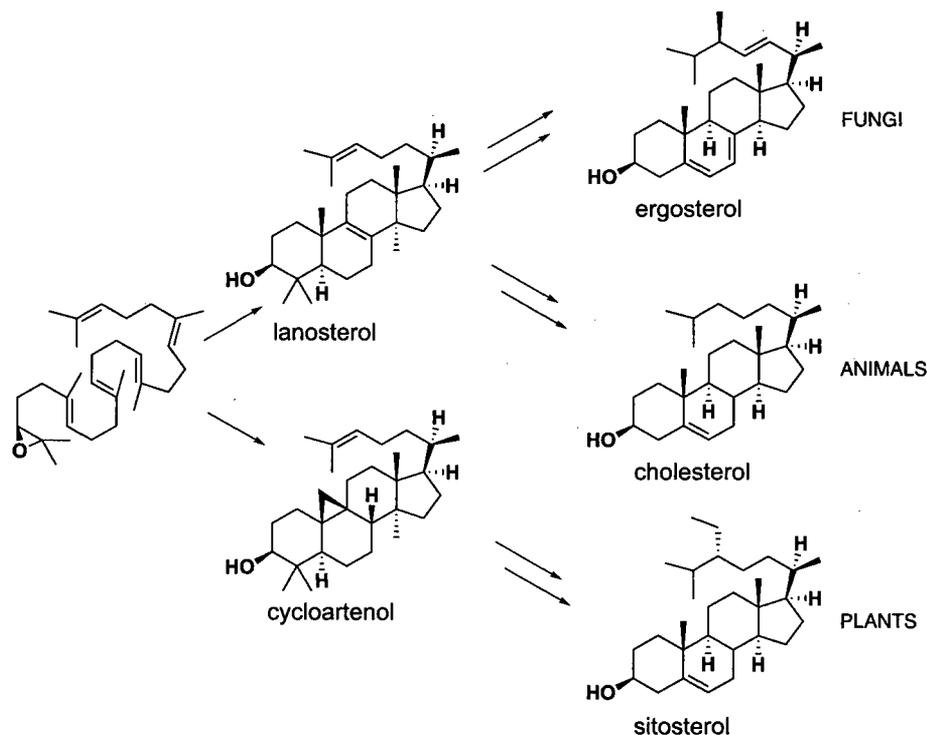
Characterization of oxidosqualene cyclases often involves analysis of triterpene products by expression in heterologous host, analysis of the OSC protein sequence with application of available bioinformatics tools and correlation of enzymatic features with the observed reaction outcome. More rigorous studies of OSCs, for example by application of crystallographic methods, are not easily accessible, because oxidosqualene cyclases are membrane-bound proteins and are a challenge for crystallographers. Crystal structure data is currently available for only a single OSC protein, human lanosterol synthase.³

However, because of the high similarity and conservation of essential structural domains among OSCs, we can make some assumptions about structures and active site constitutions of related cyclases based on the lanosterol synthase crystal structure, sequence identity between oxidosqualene cyclases and application of molecular modeling. As an alternative approach, we can correlate sequences with catalytic functions if enough cyclases with certain catalytic motifs and their active-site mutants have been well characterized. We further can apply this information to prediction of catalytic functions of uncharacterized enzymes.

1.5 Protosteryl-Type Oxidosqualene Cyclases in Plants, Animals, and Fungi

1.5.1 Sterol Biosynthesis in Animals, Fungi, and Plants.

Oxidosqualene cyclases that pre-fold oxidosqualene into a chair-boat-chair conformation are known as protosteryl-type cyclases. They are often involved in the biosynthesis of primary metabolites, triterpenes utilized as precursors in the biosynthesis of the essential membrane components, sterols. Animals and fungi biosynthesize sterols through the same precursor, lanosterol (Scheme 1.2).³⁰ In mammals, lanosterol is converted to cholesterol and steroidal hormones. Fungi utilize lanosterol to make the fungal sterol ergosterol. Plants utilize lanosterol to make the plant sterol sitosterol.



Scheme 1.2. Biosynthesis of sterols in plants, animals, and fungi.

In contrast, plants make sterols via cycloartenol (Scheme 1.2), which has an extra cyclopropyl ring that must be opened by an additional enzyme (cycloeucaleanol-obtusifoliol isomerase)³¹ on the way to the plant sterols sitosterol and stigmasterol.

Why plants maintain this less efficient pathway has been debated for decades. The use of cycloartenol in plants but lanosterol in the non-photosynthetic fungal and animal kingdoms initially suggested a connection between these structures and photosynthesis^{32,33} until diverse non-photosynthetic and dark-grown photosynthetic organisms were shown to retain the cycloartenol pathway.^{34,35,36,37,38} The prevailing hypothesis for the last several decades has been that the plant cycloartenol route is vestigial. Bloch proposed that cycloartenol would have been favored in ancient organisms that incorporated unmodified triterpene alcohols into membranes because cycloartenol condenses membrane lipids more effectively than lanosterol.^{39,40} Enzymes that catalyze demethylation, side chain modification, and B-ring rearrangement (including cyclopropyl ring opening) then evolved to generate tetracyclic sterols with optimized membrane properties. Some organisms (including opisthokonts and kinetoplastids)⁹⁰ later streamlined the process by evolving lanosterol synthase and losing the cyclopropyl isomerase, but others (including plants) did not find this more efficient route.

Although the sterol biosynthetic pathway through lanosterol has not been discovered in plants, some observations hint that lanosterol plays some role in plants. Lanosterol is metabolized by some plants^{41,42,43,44} and is biosynthesized by *Euphorbia lathryis* latex.⁴⁵ Additionally, the lanostane skeletons of several 4,4-dimethyl Δ^8 sterol saponins^{46,47,48} have been discovered in a number of plants. Because plants encode cycloeucaleanol isomerase to open the cyclopropyl ring to the Δ^8 olefin present in

lanosterol, most lanosterol metabolites could be biosynthesized readily from cycloartenol, with the notable exception of 4,4-dimethyl Δ^8 sterols. Because the 4 β -methyl group blocks access to the cyclopropyl ring, 4,4-dimethyl cyclopropyl sterols are poor substrates for plant cycloeucaenol isomerase.³¹ Therefore 4,4-dimethyl Δ^8 sterols such as peruvianosides,⁴⁶ scillasaponins,⁴⁷ lucilianosides⁴⁸ are more readily accessible from lanosterol.

1.5.2 Catalytic Differences in Cycloartenol and Lanosterol Biosynthesis

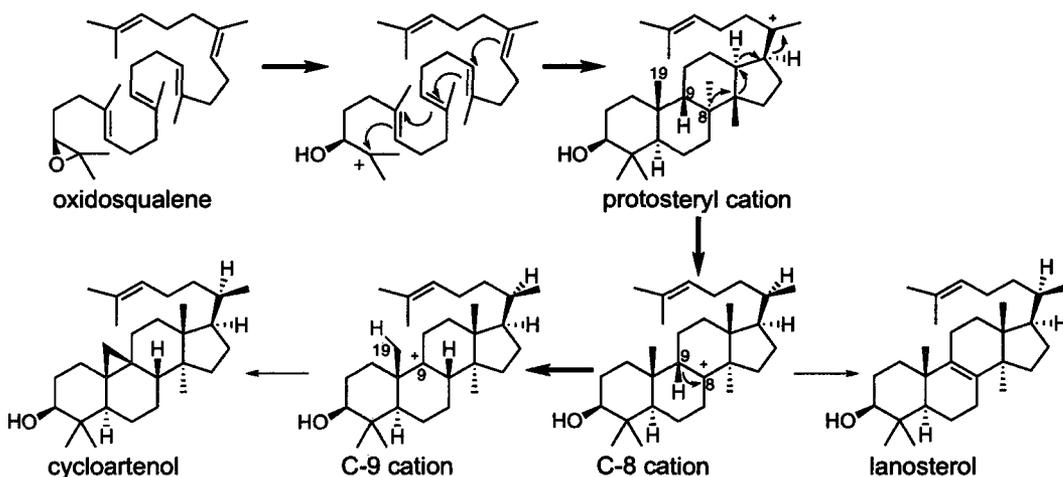
Enzymes involved in the biosynthesis of sterol precursors lanosterol (lanosterol synthases) and cycloartenol (cycloartenol synthases) have been studied more extensively than any other oxidosqualene cyclases. Comparison of amino acid sequences of the characterized enzymes has allowed for identification of conservative motifs and amino acid residues characteristic to cycloartenol synthases and lanosterol synthases (Figure 1.1).

	*		Δ	X	
<i>Trypanosoma brucei</i> ERG7	RMSGYNGSQLWDT	557	STRPQAWQVSDCTAEG	627	lanosterol synthases
<i>Trypanosoma cruzi</i> ERG7	RVCGYNGSQLWDT	548	STASQSWQVSDCTAEG	618	
<i>Saccharomyces cerevisiae</i> ERG7	TIMGTINGVQTWDC	392	STKTOGYTVADCTAEAT	462	
<i>Saccharomyces prombe</i> ERG7	LMRGTINGLQVWET	387	SNITQGYTVSDITSEAL	457	
<i>Mus musculus</i> LSS	KMQGTNGSQLWDT	390	STLDGGWIVADCTAEG	462	
<i>Rattus norvegicus</i> LSS	KMQGTNGSQLWDT	390	STLDGGWIVADCTAEAL	462	
<i>Homo sapiens</i> LSS	KMQGTNGSQLWDT	389	STLDGGWIVSDCTAEAL	461	
<i>Avena sativa</i> CAS1	KMQGYNGSQLWDT	419	STADHGWPISDCTAEG	490	
<i>Costus speciosus</i> CAS1	KMRGYNGSQLWDT	418	STADHGWPISDCTSEGL	489	
<i>Oryza sativa</i> CAS1	KMQGYNGSQLWDT	419	STADHGWPISDCTAEG	490	
<i>Zea mays</i> CAS1	KMQGYNGSQLWDT	416	STADHGWPISDCTAEG	487	cycloartenol synthases
<i>Hordeum vulgare</i> CAS1	KMQGYNGSQLWDT	419	STTDHGWPISDCTAEG	490	
<i>Panax ginseng</i> CAS1	KMQGYNGSQLWDT	418	STADHGWPISDCTAEGF	489	
<i>Pisum sativum</i> CAS1	KMQGYNGSQLWDT	418	STADHGWPISDCTAEG	489	
<i>Glycyrrhiza glabra</i> CAS1	KMQGYNGSQLWDT	418	STADHGWPISDCTAEG	489	
<i>Betula platyphylla</i> CAS1	KMQGYNGSQLWDT	418	STADHGWPISDCTAEG	489	
<i>Arabidopsis thaliana</i> CAS1	KMQGYNGSQLWDT	418	STADHGWPISDCTAEG	489	

Figure 1.1. Sequence comparison of lanosterol synthases and cycloartenol synthases.

For example, all cycloartenol synthases demonstrate conservation of several amino acid residues, including His477 and Ile481 (AtCAS1 numbering). Lanosterol synthases have a conserved valine instead of Ile481, but vary at position 477 with either Gln477, as in fungal and trypanosomal lanosterol synthases, or Cys477 in mammals.

Formation of both lanosterol and cycloartenol initially goes through the same steps, starting from chair-boat-chair pre-folding of oxidosqualene with subsequent formation of the protosteryl cation and formation of the C-8 cation (Scheme 1.3). Deprotonation from the C-9 position of the C-8 cation then gives lanosterol, while hydride migration from C-9 to C-8 and further deprotonation from C-19 yields cycloartenol.



Scheme 1.3. Oxidosqualene cyclization by lanosterol and cycloartenol synthase.

Because mechanisms of lanosterol and cycloartenol formation have similarities in the beginning and differ at the protonation step, a large number of mutagenesis studies and direct evolution experiments have been done to illuminate mechanisms, essential catalytic motifs, and particular amino acid residues affecting the reaction outcome. A

combination of knowledge from bioinformatics approaches with mutagenesis and site-directed mutagenesis allowed for identification of the essential amino acid residues dramatically affecting catalytic properties of lanosterol and cycloartenol synthases.

A summary of mutagenesis experiments performed on two extensively studied enzymes, *Arabidopsis* cycloartenol synthase (AtCAS1)⁴⁹ and lanosterol synthase (ScERG7)^{50,51} from *S. cerevisiae*, is provided in Table 1.1. This table presents data only for those experiments that have been done with careful analysis of the product profile. However there are a number of studies for which detailed information was not provided, but changes in product profile or enzyme activity were observed.^{52,53,54,55,56,57,58}

Current understanding of cycloartenol formation involves participation of the following active site residues: His257, Tyr410, His477, Ile481, and Tyr532 (*Arabidopsis* CAS1 numbering). Cycloartenol synthase protonates oxidosqualene, cyclizes the resultant carbocation to the protosteryl cation, and then facilitates rearrangement to the C-8 cation.¹ Cyclopropyl ring formation and deprotonation from C-19 are then promoted by His257, Tyr410, His477, Ile481, and Tyr532 (for illustration of the essential catalytic positions see Figure 1.2⁵⁹ and Table 1.2). The His257 and Tyr410 form a hydrogen-bonded pair, which His477 positions via a hydrogen bond to accept the proton from C-19.⁶⁰ The Ile481 residue promotes cycloartenol biosynthesis through steric bulk that positions the intermediate cation relative to His257 and Tyr410. Tyr532 is a part of the hydrogen-bonding network that is believed to facilitate deprotonation.

Table 1.1. Percentage product composition of *A. thaliana* cycloartenol synthase (AtCAS1) mutants and *S. cerevisiae* lanosterol synthase (ScERG7) mutants.

	1	2	3	4	5	6	7	8	Ref.
Mutations in AtCAS1									
none	99		1						49
I481L	83	1	16						61
I481V	54	25	21						62
I481A	12	54	15		13	6			61
I481G	17	23	4		44	12			61
Y410T		65	2	33					56
Y410T I481V		78	<1	22					56
Y410C		75		24	1				52
Y532H		45	31		24				52
H477N		88	12						63
H477G		22	73	5					63
Y410T H477N I481V		78		22					60
Y410T H477G I481V		78		22					60
H477N I481V		99	1						64
H477G I481V		94	6						64
Mutations in ScERG7									
None		100							50,51
V454I		100							65
V454L		100							65
V454A		95			5				65
V454G		83			17				65
T384Y		79	11				10		66
T384Y V454I		13	64				23		66
T384Y V454L		20	60				20		66
Y510P		95						5	67
Y510H		42	9		45			4	67

1 Cycloartenol

2 Lanosterol

3 Parkeol

4 9 β -lanosta-7,24-dien-3 β -ol

5 Achilleol A

6 Camelliol C

7 Lanost-24-ene-3 β ,9 α -diol

8 Isomalabaricatrienol

Mutations of the residues described above would change orientations of the hydrogen-bonding network within the active site and could easily shift the deprotonation base towards alternative deprotonation positions, including C-9 to give lanosterol, C-11 to give parkeol, or others. For example, the His477Asn mutation⁶³ promotes the biosynthesis of lanosterol (88%) and parkeol (12%) by relocating the polar His257 and Tyr410 residues close to C-9 and C-11. An Ile481Val mutation facilitates lanosterol production (25% lanosterol, 21% parkeol, and 54% cycloartenol) by enlarging the active site and reducing steric control.¹⁷ These mutations cooperate to promote lanosterol biosynthesis; an *AtCAS1* His477Asn Ile481Val double mutant biosynthesizes lanosterol accurately.⁶⁴ Disturbance of the considered residues easily causes changes in the cycloartenol synthase product profile, often completely abolishing biosynthesis of cycloartenol.

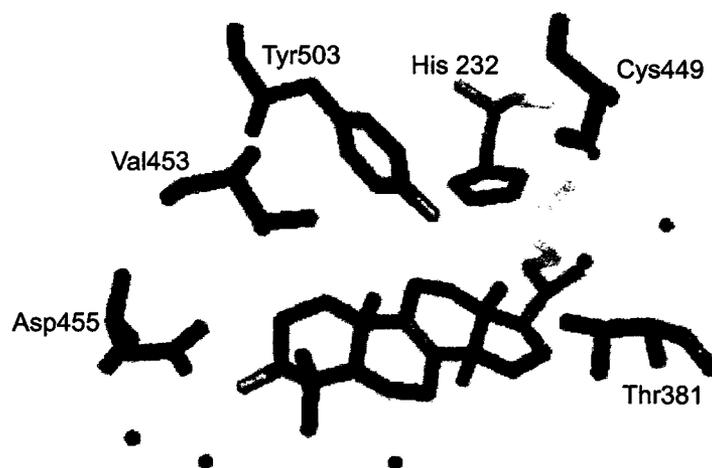


Figure 1.2. Selected active-site residues of human lanosterol synthase. Red dots correspond to ordered water molecules.

Table 1.2. Comparison of the corresponding active site residues between human lanosterol synthase (HsLSS), *Arabidopsis* cycloartenol synthase (AtCAS1), and yeast lanosterol synthase (ScERG7).

HsLSS	AtCAS1	ScERG7
His232	His257	His234
Thr381	Tyr410	Thr384
Cys449	His477	Gln450
Val453	Ile481	Val454
Asp455	Asp483	Asp456
Tyr503	Tyr532	Tyr510

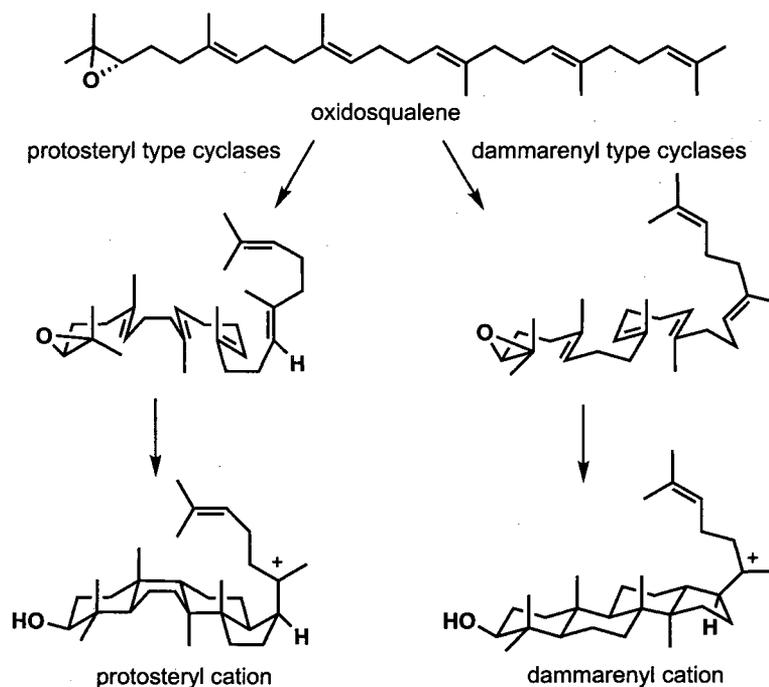
Lanosterol biosynthesis is affected by active site residues His234, Thr384, Val454, and Tyr510 (*S. cerevisiae* lanosterol synthase numbering), which correspond to *Arabidopsis* cycloartenol synthase residues His257, Tyr410, Ile481, and Tyr532. In lanosterol synthases it is believed that deprotonation from the C-9 proton is promoted by relocating the deprotonating base by a hydrogen bond between Tyr510 and His234, with initial deprotonation to Tyr510 with further transport of the proton to His234.³ Changes in steric bulk at Val454 influences the hydrogen bond orientation and is shown to affect cyclization but not deprotonation.⁶⁵ Thr384 decreases steric bulk compared to the strictly conserved Tyr410 in cycloartenol synthases. The Thr384Tyr mutation in lanosterol synthase broadens the product profile, suggesting Thr384 is essential for catalysis.⁶⁶ Notably, in all lanosterol synthase mutagenesis experiments that gave oxidosqualene cyclization products, lanosterol biosynthesis was never abolished completely.

1.6. Dammarenyl-type Oxidosqualene Cyclases in Plants

Another type of oxidosqualene cyclases is the dammarenyl-type cyclases. Dammarenyl-type oxidosqualene cyclases pre-fold oxidosqualene in the pre-chair-chair-

chair conformation (Scheme 1.4), compared to the pre-chair-boat-chair conformation in protosteryl-type cyclases. These cyclases are involved in the biosynthesis of secondary metabolites in plants. Plants are sessile organisms that require protection and communication methods different from those available to animals. Thus, plants make large diversity of secondary metabolites, including triterpenes and their derivatives, that can be successfully used for chemical protection. Cucurbitadienol synthase from *Cucurbita pepo* is the only reported example of a protosteryl-type oxidosqualene cyclase involved in secondary metabolism.⁶⁸ Only a few other cyclases, including marneral synthase⁶⁹ and thalianol synthase,⁷⁰ biosynthesize B-ring chair triterpenes involved in secondary metabolism but that do not access the tetracyclic dammarenyl cation and consequently cannot be classified as dammarenyl-type cyclases. Phylogenetic analysis and amino acid sequence comparison shows that thalianol synthase and marneral synthase arose from dammarenyl-type cyclases. Thalianol arises from a B-ring chair intermediate, and marneral is produced through initial cyclization to a bicyclic carbocation that undergoes Grob fragmentation to cleave the A-ring.

Protosteryl-type cyclases and dammarenyl-type cyclases promote different substrate conformations, and these changes are presumably due to significant differences mostly in hydrophobic or aromatic amino acids active site residues that preorganize the initial conformation of oxidosqualene. A phylogenetic tree constructed to account for all known (or most known) oxidosqualene cyclases demonstrates that dammarenyl-type cyclases and protosteryl-type cyclases diverged only once in the past, within the plant kingdom, and is a rare evolutionary event.^{71,72,73,74,75}



Scheme 1.4. Cation formation in protosteryl and dammarenyl type cyclases.

The most studied dammarenyl-type cyclases are lupeol synthases and β -amyirin synthases (reviewed in ^{21,29}). β -amyirin and lupeol are biosynthesized through the dammarenyl cation, with further D-ring expansion and annulation to give the lupyl cation, followed by C-29 deprotonation to give lupeol, or E-ring expansion and 1,2-hydride shifts to give β -amyirin.¹ Most lupeol and β -amyirin synthases have high product selectivity, however there are a number of enzymes with a broader spectrum of products, including mixed amyirin synthases from *Pisum sativum*,⁷⁶ *Lotus japonicus*,⁷⁷ *Costus speciosus*,⁷⁸ and *Kandelia candel (L.)*.⁷⁹ Cyclases that make different dammarenyl-type triterpenes, such as isomultiflorenol synthase from *Luffa cylindrica*,⁸⁰ or dammarenediol-II synthase from *Panax ginseng*,⁸¹ are known as well.

1.7. Oxidosqualene Cyclases from *Arabidopsis*

A total of thirteen oxidosqualene cyclases have been identified by a BLAST search from the *Arabidopsis* genome. Plants usually contain more oxidosqualene cyclases than animals and fungi, which usually have only one – lanosterol synthase. However, the number of oxidosqualene cyclases in different plants could vary, therefore sequencing of plant genomes would uncover the exact number of genes that encode oxidosqualene cyclases. The *Arabidopsis* cyclases share 44-84% amino acid sequence and can be divided into 3 groups (clades): CAS clade (CAS1 and LSS1), LUP clade (LUP1, LUP2, LUP3, LUP4, LUP5), and PEN clade (PEN1, PEN2, PEN3, PEN4, PEN5, PEN6), according to their sequence distances and phylogenetic relationship (Figure 1.3).²²

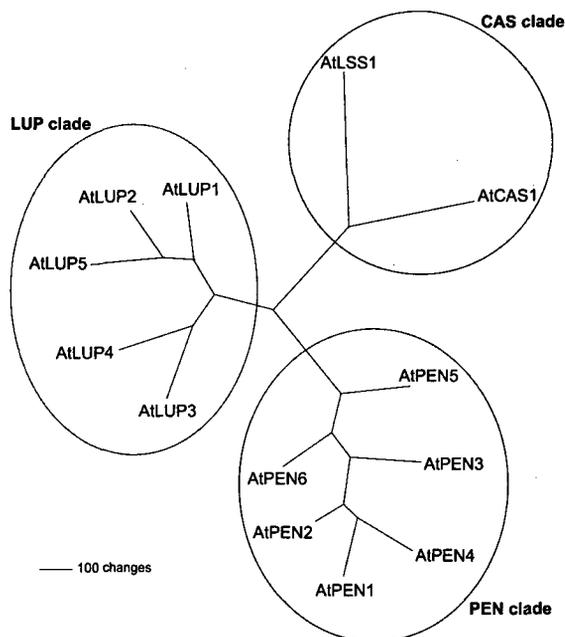
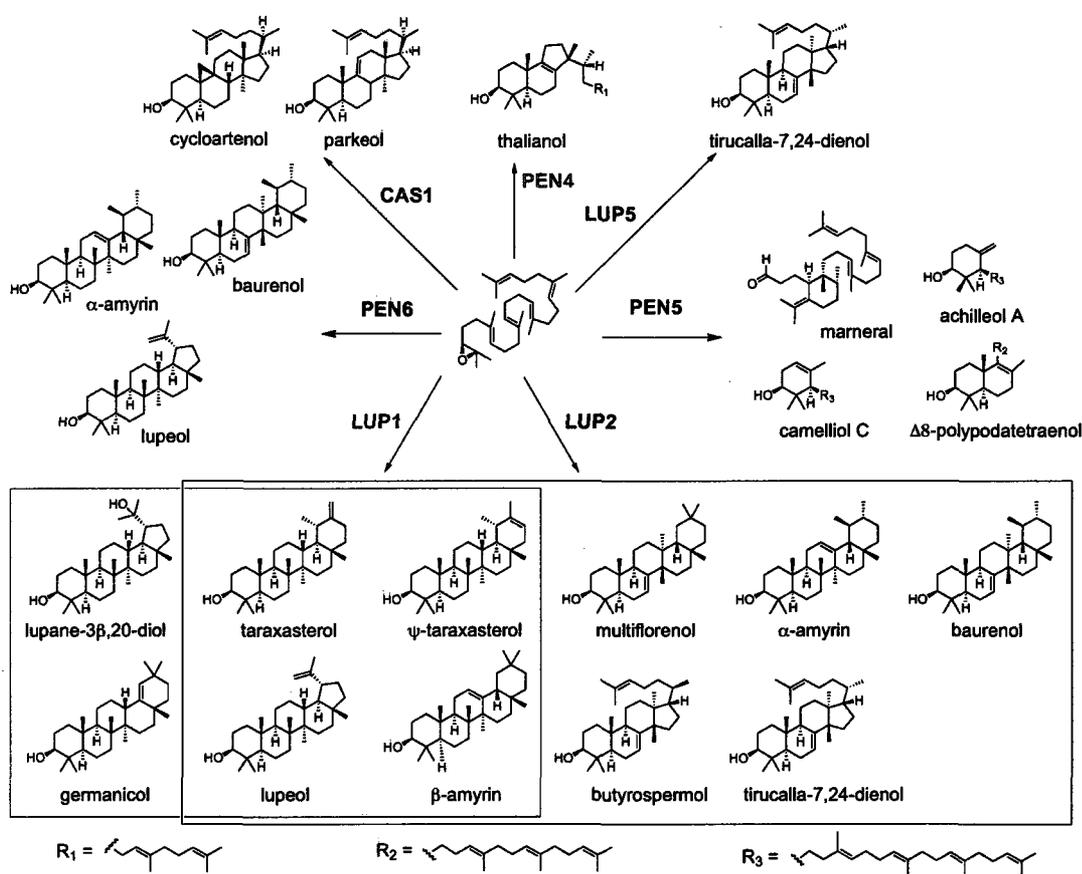


Figure 1.3. Computed phylogenetic relationships between *Arabidopsis* oxidosqualene cyclases. The tree was generated using the bootstrap method with 1000 replicates with equal weight given to all the characters and maximum parsimony as the optimality criterion.

Before the initiation of my thesis research, three oxidosqualene cyclases from *Arabidopsis*, including cycloartenol synthase (CAS1),⁴⁹ lupeol synthase (LUP1)^{74,75,82,83} and mixed lupeol synthase (LUP2)^{22,84} were characterized by heterologous expression in yeast (Scheme 1.5).



Scheme 1.5. Biosynthesis of triterpenes in *Arabidopsis*.

Later, two more enzymes, including marneral synthase (PEN5, MRN1)⁶⁹ an enzyme that makes a B-ring monocycle after rearrangement and Grob fragmentation, and thalianol synthase (PEN4, THAS1),⁷⁰ an enzyme that makes the novel tricyclic 6/6/5 triterpene thalianol, were characterized using this same approach. Another two enzymes

were partially characterized (PEN6 and LUP5) and reported to make α -amyrin, bauerenol, lupeol (PEN6), and tirucalla-7,24-dienol (LUP5).⁸⁵ Another cyclase, PEN1,²² was expressed in yeast, however oxidosqualene cyclization products were not detected.

Heterologous expression of just a few reported *Arabidopsis* oxidosqualene cyclases demonstrated that they are capable of production of more triterpenes than found by extraction from *Arabidopsis* plants. This indicates the tremendous potential for triterpene discovery, even within a simple model plant such as *Arabidopsis*.

This work presents characterization of three additional *Arabidopsis* oxidosqualene cyclases with novel functions not previously observed in plants, including the first plant lanosterol synthase (LSS1), the first enzyme (camelliol C synthase, CAMS1) that efficiently blocks B-ring formation to make a monocyclic triterpene, and an enzyme (PEN1) that makes a novel tricyclic triterpene diol and makes the most compounds (>25) ever observed for a triterpene synthase. It also provides additional insights into the mechanisms of oxidosqualene cyclization and catalytic differences in protosteryl-type cyclases.

CHAPTER 2

Materials and Methods

2.1 Materials

Restriction enzymes and Quick T4 DNA Ligation Kit were purchased from New England BioLabs (Beverly, MA). The Triple Master PCR System was from Eppendorf (Westbury, NY). A RETROscript kit for making *Arabidopsis* cDNA was purchased from Ambion (Austin, TX). Gene pure agarose LE (ISC BioExpress, Kaysville, UT) was used for gels. TOPO TA Cloning Kit for Sequencing was purchased from Invitrogen (Carlsbad, CA). Qiagen Gel Extraction Kit was used for DNA recovery from agarose gel (Qiagen, Inc., Valencia, CA). Media components were purchased from United States Biological (Swampscott, MA). Glass beads for yeast homogenization were purchased from BioSpec Products (Bartlesville, OK). Hemin chloride (heme), ergosterol, bis(trimethylsilyl)trifluoroacetamide (BSTFA) and Triton X-100 were from Sigma-Aldrich (St. Louis, MO). Pyridine, TLC plates, and solvents for extraction and saponification were purchased from EMD Chemicals, Inc. (Gibbstown, NJ).

2.2 Nuclear Magnetic Resonance (NMR)

^1H , ^{13}C , and 2D Nuclear Magnetic Resonance (NMR) spectra were acquired at 25 °C in deuterated chloroform (CDCl_3) solution that had been filtered through activated basic alumina. Samples for NMR measurements were prepared in 5 mm glass tubes (Wilmad Glass Co., Inc or Shigemi Inc). Spectra were obtained on Varian Inova 800 equipped with a cold probe, Varian Inova 600, and Bruker Avance 500 spectrometers. Spectra were referenced to tetramethylsilane at 0 ppm for ^1H and to CDCl_3 at 77.00 ppm for ^{13}C NMR.

2.3 Gas Chromatography – Mass Spectrometry (GC-MS)

Gas chromatography - mass spectrometry analysis was done with an Agilent 5973N GC-MSD instrument interfaced to an Agilent 6890N GC system containing an Rtx-35ms column (30 m × 0.25 mm × 0.1 μm, Restek). Samples (2 μL) were injected into the inlet using a 40:1 split; the injection temperature was 280 °C, and the isothermal oven temperature was either 260 °C (MDKTRIT2 method) or 270 °C (MDKTRIT1 method) for 30 min. MS data were acquired in the full-scan mode from 50 to 650 amu after a 3 min solvent delay.

2.4 Gas Chromatography-Flame Ionization Detection (GC-FID)

Gas chromatography-flame ionization detection was performed using a Hewlett-Packard 6890 instrument (equipped with a 30 m × 0.25 mm ID × 0.1 μm film thickness, Rtx-5 column from Restek). Two methods, MDKTRIT2 and MDKTRIT1, were used for analysis. The following conditions were used for the analysis: the inlet and FID were held at 290 °C, an isothermal oven program held at 270 °C (MDKTRIT1) or 260 °C (MDKTRIT2) for 30 min, the split ratio was 40:1, gas flow (helium) was 0.6 mL/min at a constant pressure.

2.5 High-Performance Liquid Chromatography (HPLC)

HPLC separations of triterpene products were done with an Agilent 1100 series instrument containing a Phenomenex column (ODS(3), 250 × 21 mm, 5-μm pore size, 100 Å pore size, Prodigy). Separations were performed isocratically at a flow rate of 8

mL/min or 7 mL/min with a mobile phase of water/methanol and UV detection at 210 nm.

2.6 UV Spectroscopy

UV visible spectrophotometric measurements were done on a Shimadzu UV-visible spectrophotometer UV-1601 at 600 nm, using 10 mm UV-grade polymethylmethacrylate (PMMA) cuvettes from VWR International Inc. (West Chester, PA).

2.7 Centrifugation

Centrifugations of 1.5-mL and 2-mL tubes were performed on Eppendorf Centrifuge models 5415D and 5810R. Centrifugations of 15- to 250-mL tubes and centrifuge cans were performed using Eppendorf Centrifuge 5810 (variable temperature and speed).

2.8 Incubations

Microbial plate cultures were grown in Fisher Scientific Isotemp incubators at 30 °C or 37 °C. All liquid cultures of *E. coli* and *S. cerevisiae* were grown in either a New Brunswick Series 25 Incubator Shaker or a New Brunswick C24 Incubator Shaker.

2.9 Polymerase Chain Reaction (PCR)

Polymerase chain reactions (PCR) and reverse transcriptase polymerase chain reactions (RT-PCR) were performed using an Eppendorf (Hamburg, Germany) Mastercycler Gradient thermocycler.

Oligonucleotides were purchased from Integrated DNA Technologies, Inc. (Houston, TX) or Sigma-Genosys (The Woodlands, TX). Oligonucleotides were dissolved in mqH_2O to a final concentration of 100 pmol/ μL . An aliquot of the 100 pmol/ μL solution was used to make a 20 pmol/ μL stock solution used for PCR reactions. Oligonucleotides for sequencing were diluted to 5 pmol/ μL . Stock solutions were stored at $-20\text{ }^\circ\text{C}$.

A reaction for PCR amplification typically contained template DNA (cDNA, plasmid 1-10 ng), a forward primer (20 pmol), a reverse primer (20 pmol), 5 μL 10 \times High Fidelity Buffer (5 prime or Eppendorf), deoxyribonucleotide triphosphates (dNTPs) (10 mmol each), Triple Master Polymerase (1-2 U) and mqH_2O to 50 μL total reaction. For PCR amplification a CAPS program was used that included the following program: 95 $^\circ\text{C}$, 1 min; 65 $^\circ\text{C}$, 30 s; 72 $^\circ\text{C}$, 3 min with final 5 min extension at 72 $^\circ\text{C}$, all repeated 36 times.

DNA mutagenesis was generally performed by introducing the desired mutation into the gene through PCR amplification with modified oligonucleotides. A mutation was usually designed by substituting a codon with minimal change to the nucleotide sequence. Unique restriction sites used for cloning were either found in the initial DNA sequence within 30 bp of the mutation, or by artificially introducing the site by nucleotide substitution without changing the amino acid sequence, also within 30 bp of the mutation.

That the desired mutation was introduced without changing other amino acids was established by sequencing the insert in the expression construct.

2.10 DNA Plasmid Purification

For mini-prep analysis of DNA sets of 12 or 24 bacterial cultures (2 mL each) were grown overnight in selective LB media. Bacterial cells were pelleted by centrifugation, followed by removal of supernatants. The cell pellets were then resuspended in 100 μ L cooled (4 $^{\circ}$ C) buffer P1 (50 mM Tris-HCl (pH 8.0), 10 mM ethylene diamine tetraacetic acid (EDTA)) containing 20 mg/mL RNase A. After 15 min incubation with P1, 100 μ L P2 lysis buffer (200 mM NaOH, 1% sodium dodecyl sulfate (SDS) (w/v)) were added and the samples were very gently mixed and incubated 3-5 min at room temperature. Addition of 100 μ L cooled P3 buffer (3.1 M potassium acetate, pH 5.5) and incubation at 4 $^{\circ}$ C for 30 min to 1 h terminated alkaline lyses. The cellular debris was pelleted by centrifugation and supernatants containing the DNA were transferred to new 1.5-mL tubes. To precipitate DNA 2.5 volumes of absolute EtOH were added to the supernatants, mixed by vortexing and incubated for at least 1 h at -20 $^{\circ}$ C. The ethanolic mixtures were centrifuged, the supernatants discarded and the DNA pellets were air-dried prior to dissolving in 100 μ L TE8 buffer (10 mM Tris-HCl, pH 8.0, 0.1 mM EDTA). Aliquots of 250 μ L absolute EtOH were added for DNA re-precipitation and the mixtures were chilled at -20 $^{\circ}$ C. The precipitation mixtures were centrifuged, the supernatants were discarded, and the DNA pellets were dried and resuspended in 50 μ L TE8. The DNA solutions obtained by this method were stored at -20 $^{\circ}$ C.

For preparative DNA purification, large-scale (30-100 mL) bacterial cultures were grown in LB media containing selective antibiotic for 12-16 h. Cell pellets obtained by centrifugation were resuspended in 1 or 2 mL cooled P1 buffer supplemented with RNase A. An equal amount of P2 buffer was added to the pellets, gently mixed and incubated for 5 min at room temperature, followed by addition of P3 buffer (1 or 2 mL). The mixture was then kept on ice for a minimum of 1 h before being transferred to 1.5-mL centrifuge tubes and centrifuged to precipitate cell debris. The supernatant was moved to clean centrifuge tubes, and 0.7 volumes of isopropyl alcohol was added prior to vortexing and chilling at $-20\text{ }^{\circ}\text{C}$ for at least 1 h to enhance yield. Following centrifugation, the supernatant was removed by decanting and drying residual EtOH. The DNA pellets were dissolved in TE8 and combined into a single tube. After combining all DNA into one tube, a second DNA precipitation was performed using 2.5 volumes of absolute EtOH. The DNA was pelleted by centrifugation, ethanolic supernatant was removed and dried DNA pellets were resuspended in 50-150 μL TE8 buffer. Obtained DNA solutions were stored at $-20\text{ }^{\circ}\text{C}$.

DNA was sequenced was performed by Seqwright DNA Technology Services (Houston, TX) and by Lone Star Labs, Inc. (Houston, TX).

2.11 DNA Restriction Digestion

The DNA digestion conditions were used according to the manufacturer recommendations provided in the New England Biolabs 2007-08 Catalog & Technical Reference. Reactions were incubated for a minimum of 1 h. For analytical purposes 10-

100 ng of DNA was used for a digest in 20 μ L reaction mixtures; for preparative purposes 100-1000 ng of DNA was used in 30-100 μ L reaction mixtures.

2.12 DNA Gel Analysis

Two types of agarose gels were used for DNA analysis. One type was used for analytical purposes and was prepared by microwaving 5 g agarose in 500 mL 1 \times Tris-Acetic acid-EDTA (TAE) buffer followed by addition of ethidium bromide (final concentration 5 μ g/100 mL gel). A stock 50 \times TAE buffer was prepared by dissolving 242 g Tris base, 57.1 g glacial acetic acid, and 18.6 g EDTA in 1 L mqH_2O . For preparative analysis of DNA GTAE buffer (TAE with addition of 0.283 g guanosine per 1 L of 1 \times TAE) was used.

For DNA analysis a 10 \times gel loading buffer (20% Ficoll 400, 0.1 mM EDTA, pH 8.0, 0.25% bromophenol blue, and 0.25% xylene cyanol) was added to each sample loaded. An appropriate molecular weight DNA marker (typically λ DNA digested with *BstE* II) was used as a control.

DNA fragments run on preparative gels were excised with razor blades and eluted with QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's manual.

2.13 DNA Ligation

A Quick T4 DNA Ligation Kit was used for DNA ligation. To a 10 μ L aliquot of 2 \times ligation buffer the vector DNA and insert DNA (or two insert DNA in case of 3-piece ligation) fragments with cohesive ends were added in amounts to achieve a 2:1 molar ratio of insert to vector (or 2:2:1 in case of a 3-piece ligation) and a total volume of 19

μL. A 1-μL aliquot of Quick Ligase was added and mixed by pipetting. After incubation at room temperature for 5-30 min the ligation was immediately used for *E. coli* transformation.

Several plasmid vectors were usually used for cloning and ligation, including pRS305Gal, pRS316Gal, and pRS426Gal.⁸⁶

2.14 Bacterial Media

E. coli cultures were grown in Luria-Bertani (LB) broth (LB, 10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) supplemented with a selection antibiotic. The LB media was sterilized by autoclaving at 121 °C for 35 min followed by cooling to room temperature, and addition of antibiotics ampicillin (to final concentration 100 μg/mL) or kanamycin (to final concentration 50 μg/mL). Solid LB media was prepared as described above except with the addition of agar (15 g/L) prior to sterilization. Selection antibiotics were stored at -20 °C as stock solutions of 25 mg/mL antibiotic in 1:1 ethanol/water solution. Incubations of bacterial cultures were performed at 37 °C.

During bacterial transformation procedure a rich SOC medium (20 g/L bacto-tryptone, 5 g/L yeast extract, 0.5 g/L sodium chloride, 0.186 g/L potassium chloride, 3.6 g/L glucose) was usually used for incubation of the bacteria.

2.15 *E. coli* Transformation

Three types of chemically competent cells, (stored at -80 °C) DH5α, NEB 5-alpha Competent *E. coli* (New England Biolabs, Inc.) and One Shot Mach1 – T1 Chemically Competent *E. coli*, (Invitrogen) were used for DNA transformations. For plasmid

transformations, cells were thawed on ice for 15 min. DNA was added to the cells and then incubated on ice for 20 min. The cells were heat-shocked at 42 °C for 30 s with further incubation of cells on ice for 2 min. An aliquot of 250 µL SOC medium was then added to cells and incubated with shaking at 37 °C for 1 h. After incubation aliquots of the transformation (25 µL and 250 µL) were plated on selective LB agar plates and spread with sterile glass beads.

2.16 Yeast Strains Used for Expression of Oxidosqualene Cyclases

Two types of yeast strains were used for the experiments. SMY8⁸⁷ (*MATa erg7::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2 GAL⁺*) is a lanosterol synthase deletion mutant and RXY6⁸⁸ (*MATa erg1::KanMX erg7::HIS3 hem1::TRP1 ura3-52 trp1-Δ63 leu2-3,112 his3-Δ200 ade2 GAL⁺*) is a lanosterol synthase / squalene epoxidase double deletion mutant.

Oxidosqualene cyclases that generate compounds similar to lanosterol may not be rigorously characterized by in vivo accumulation studies alone. Because yeast possesses enzymes that convert lanosterol to ergosterol, most of the lanosterol it biosynthesizes accumulates as ergosterol or other demethylated metabolites. These enzymes can accept a reasonably broad array of triterpene alcohols, including cycloartenol and parkeol.⁸⁹ Consequently, any byproducts generated by a lanosterol synthase might be metabolized more quickly or more slowly than lanosterol, and they would be correspondingly under- or overrepresented in the extract. The oxidative enzymes require NADPH, which depletes rapidly in a cell-free homogenate, and triterpene alcohols are consequently not significantly metabolized in vitro. However, if in vitro assays are conducted using

homogenates derived from strains that can produce substrate, triterpene alcohols that accumulated *in vivo* during the induction phase would compromise the analysis. The yeast strain RXY6 is an ideal host for expressing oxidosqualene cyclase protein for *in vitro* analysis. Its lanosterol synthase deletion abolishes *in vivo* and *in vitro* oxidosqualene cyclization by yeast enzymes and thereby ensures that any recovered $C_{30}H_{50}O$ compounds are products of the foreign oxidosqualene cyclase. The additional deletion of squalene epoxidase precludes the biosynthesis of the precursor oxidosqualene and thereby ensures that any cyclization occurs after cell lysis when oxidases are inactivated by low NADPH levels.⁹⁰

2.17 Yeast Media

Yeast cultures were grown in several types of media. For yeast strains RXY6 and SMY8 that did not carry plasmids a rich YP media (10 g/L yeast extract, 20 g/L peptone) was used. Yeast strains containing transformed plasmids with selection markers were grown in synthetic complete (SC; 1.7 g/L yeast nitrogen base, 5 g/L ammonium sulfate, 2 g/L amino acid mixture) media. Amino acid mixtures were prepared by combining 10 g of leucine and 2 g each of alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine, adenine, and uracil. Amino acid mixtures SC-Ura (SC media without uracil) or SC-Leu (SC media without leucine) were lacking either uracil or leucine, depending on the selective marker (selective media).

All cultures were supplemented with carbon source either dextrose or galactose with a final concentration of 20 g/L. Galactose was used as the inducing sugar in inducing media. Solid media were prepared by addition of 7.5 g agar to the sugar source. Additional components were heme (in the form of hemin chloride 13 $\mu\text{g}/\text{mL}$), ergosterol (20 $\mu\text{g}/\text{mL}$), and Tween 80 (5 g/L).

Stock solutions for media components (YP, SC-Ura, SC-Leu, dextrose, galactose) were prepared as 2 \times solutions, by dissolving components in mQH_2O and autoclaving at 121 $^\circ\text{C}$ for 35 min. Stock solutions for heme, ergosterol and Tween 80 were prepared as 100 \times Heme (65 mg hemin hydrochloride, 25 mL absolute EtOH, 25 mL dH_2O , and 0.75 mL 1M NaOH) and 100 \times ergosterol-Tween 80 solution (20 mg ergosterol, 5 mL absolute EtOH, and 5 mL Tween 80). All yeast cultures were cultivated at 30 $^\circ\text{C}$.

2.18 Yeast Transformation

A 10-mL yeast culture was grown to saturation at 30 $^\circ\text{C}$. Cells were collected by centrifugation for 5 min at 3000 rpm. After discarding the supernatant, the yeast pellet was washed twice with 20 mL sterile mQH_2O . The pellet was resuspended in the residual water. Plasmid DNA (5-15 μg) was added to the suspension of the yeast pellet followed by addition of pre-heated salmon sperm ssDNA solution (1 mg/mL). Two mL of yeast transformation buffer containing aqueous solution of 40% polyethylene glycol (PEG) 3350, 0.1 M lithium acetate, 10 mM tris buffer (pH 7.5), 1 mM EDTA, and 100 mM dithiothreitol (DTT) was added last and mixed by vortexing. The transformation mixture was incubated at room temperature for 8-16 h.

After incubation, a 20-mL aliquot of sterile mqH₂O was added to the yeast suspension and vortexed and the yeast pellet was collected by centrifugation. After discarding the supernatant the pellet was washed two times with 20 mL sterile water. The supernatant was discarded, and the pellet was resuspended in the residual water and plated on selective plates by spreading with sterile glass beads.

Yeast colonies grown on the selective transformation plates were then inoculated into 10 mL of selective liquid media with dextrose, grown to saturation and used for preparation of a glycerol stock and for further in vivo and in vitro assays. Glycerol stock solutions were prepared by mixing 0.8 mL of culture with 0.8 mL of 80% glycerol in water solution and were stored at -80 °C. All yeast cultures were cultivated at 30 °C.

2.19 (±)-2,3-Oxidosqualene Synthesis

Racemic oxidosqualene was synthesized from squalene as described previously.⁹¹ The starting material squalene (22.5 g) was dissolved in 500 mL tetrahydrofuran (THF) and cooled on ice for 10 min. Water (100 mL) was added to the solution while stirring until it became cloudy, and then THF (70 mL) was added until the solution became clear again. Freshly recrystallized *N*-bromosuccinimide (NBS, 11 g) was added in small portions over 10 min while constantly stirring for 20 min. The reaction mixture was concentrated 40-50% by evaporation in vacuo. Water (200 mL) was added to the concentrated reaction mixture and the crude bromohydrin was extracted with 4 × 200 mL hexanes. The combined hexane fractions were washed with 5 × 100 mL water and concentrated in vacuo to 80 mL. The concentrated crude bromohydrin was then purified by silica-gel column chromatography (280 g of silica gel 60), starting with 2% ether in

hexane and then increasing the ether concentration to 5% and ending with 7% ether. Fractions containing the bromohydrin were combined and concentrated in vacuo, yielding about 8 g of bromohydrin. Potassium carbonate (K_2CO_3 , 25 g) and methanol (200 mL) were added to the bromohydrin and were stirred for 3 h. Fifty milliliters of water was added to the reaction mixture, followed by extraction of the final oxidosqualene product with 4×100 mL hexanes. The combined hexane fractions were washed with 3×100 mL water and concentrated in vacuo. The racemic oxidosqualene was purified by silica-gel column chromatography using conditions described previously. A total of 27 fractions (~130 mL each) were collected. Fractions 15-18 containing oxidosqualene were combined and concentrated, giving 5.1 g of racemic oxidosqualene. A portion of the purified product was analyzed by 800 MHz NMR to determine the purity of the synthesized oxidosqualene. This racemic mixture was stored at -20 °C and used for preparation of the $20 \times$ racemic oxidosqualene solutions used for in vitro reactions.

2.20 Preparation of $20 \times$ (\pm)-2,3-Oxidosqualene Solutions

For preparation of a $20 \times$ stock oxidosqualene solution an aliquot of 100 mg of racemic synthetic oxidosqualene was mixed with 100 mg of either Tween 80 or Triton X-100 in the presence of methylene chloride (400 μ L). After 20 min of stirring on ice, methylene chloride was removed under vacuum. A 4.8-mL aliquot of mqH_2O was then added to the mixture and stirred for 40 min while keeping the mixture on ice. The final solution was a white emulsion and was stored at -20 °C.

2.21 Small Scale Oxidosqualene Cyclase Assay

Several yeast strains were tested for oxidosqualene conversion *in vitro* at small-scale. For these experiments 10-mL yeast cultures were grown in selective media in the presence of the inducing sugar galactose. The yeast cell pellets were collected by centrifugation and re-suspended in 100 mM sodium phosphate buffer (pH 6.2) at a ratio of 1 mL buffer per 1 g cell pellet. To this suspension, oxidosqualene was added to a final concentration of 1 mg/mL, followed by addition of 300-500 μ L of acid-washed glass beads (0.5 mm in diameter). Yeast pellets were lysed by vortexing for 3 min, followed by 1 min incubation on ice, and repeating this 2-step vortexing routine twice. Yeast cell lysates were then incubated at room temperature 12-24 h and later checked by TLC analysis. Negative control reactions were done in the same manner, except no substrate was added to the cell suspension.

2.22 Large Scale Oxidosqualene Cyclase Assay

Large-scale (2-L) cultures were grown in selective media with the inducing sugar galactose. Growth of yeast cultures was monitored by measuring the optical density (OD) using a 1:10 dilution of each culture. Once cells had reached a certain density and the OD was constant for two consequent measurements (between 0.5 to 0.8) taken within 3-4 h of each other, the cell pellets were collected by centrifugation and supernatants were discarded. The cells were resuspended in one volume 100 mM sodium phosphate buffer (pH 6.2) and homogenized using an EmulsiFlex-C5 cell homogenizer (Avestin Inc., Ottawa, Canada) at a minimum of 10,000 psi. A small aliquot of homogenate was typically collected for negative control, where no substrate was added. The remainder of

the homogenate was used for a reaction with synthetic oxidosqualene, added as a 20 × solution with either Triton X-100 or Tween 80 to a final concentration of 1 mg/mL OS. The homogenate was then mixed and incubated for 24-28 h at room temperature.

Upon completion, the homogenate was quenched twice with one volume of ethanol, mixed and separated by centrifugation on cell debris pellet and supernatant that was collected. Combined ethanolic fractions were then concentrated under vacuum and then run through a small (6-7 g) silica plug with 2% ether in hexane. Fractions containing squalene and unreacted oxidosqualene were then removed and the rest of the eluted components (cyclic products and ergosterol) were combined together and stored at -20 °C. Part of the combined material was used for GC-MS and GC-FID analysis and the rest of the extract was subjected to preparative thin layer chromatography (PTLC).

2.23 In vivo Assay

Yeast cultures (small scale 10- or 100-mL, or large scale 1- to 50-L) were cultivated in selective media supplemented with galactose. When cultures reached saturation, the cells were pelleted by centrifugation. Supernatants were often collected as well and used for further analysis of polar components. Collected cell pellets were then saponified with 10% KOH in 80% EtOH at 70 °C for 2 h, using 5 mL of saponification solution per 1 g of cell pellet. One volume of water was then added to the saponified mixture, followed by extraction of the non-saponifiable lipids (NSL) with hexane. For small-scale assays 4 × 15 mL of hexane was used for extraction (for both 10 and 100 mL cultures); for large scale assays 4 × 30 mL of hexane was typically used per liter of culture. Hexane fractions were combined, washed with water and concentrated in vacuo.

Sometimes, additional extraction with methyl tert-butyl ether (MTBE) was done and the hexane and MTBE extracts were combined and concentrated under vacuum. The obtained extracts of NSL were subjected to GC-MS, GC-FID, and NMR analyses, or further purified by silica-gel chromatography, HPLC, or PTLC.

2.24 Isolation of Triterpene Alcohols from the Media with Resin

If polar components were expected among triterpene products, the medium collected after separation from the cells was further incubated with activated Diaion HP-20 resin (20 g of resin per 1 L of media, activated with methanol for 15 min, Supelco). After 12-24 h of incubation the medium was separated from the resin by filtration and the collected resin was extracted with ethanol until the extracts were colorless. Combined ethanol fractions were further concentrated in vacuo and run through a silica-gel column using ether. Obtained fractions were then analyzed by GC-MS and subjected to further purification by silica-gel column chromatography or PTLC.

2.25 Purification of Triterpene Alcohols by Silica Gel Chromatography

Silica gel 60 (240-400 mesh) was used for silica-gel chromatography. Columns were wet-packed with solvent systems to be used for separation or dry-packed. The amount of silica used for separation was 500-1000 times more than the amount of crude material to be separated. Solvent systems were usually 2% ether in hexane with consequent increase of ether concentration, unless specifically noted. Fractions were collected in disposable borosilicate tubes. Column progress was monitored by TLC, developed in 1:1 hexane/ether or methylene chloride, and stained with anisaldehyde.

2.26 Purification of Triterpene Alcohols by Preparative Thin Layer Chromatography

Silica Gel 60 F₂₅₄ pre-coated plates for thin layer chromatography were used for all PTLC purifications. Prior to loading samples, all plates were pre-washed with ether (if expected developing system was hexane/ether), or methanol/methylene chloride 1:1 mixture (if the later developing system was methylene chloride). During preparation, the washing solvent was allowed to run to the end of the plate. The solvent was allowed to evaporate and the plate was heated in the oven at 120 °C overnight (no longer than 24 h). After the plate was cooled to room temperature, the sample was loaded as a thin line 20 mm from the bottom of the plate and the plate was developed in 200 mL of solvent. For good separation no more than 20 mg of sample was loaded per plate. The solvent was evaporated and the plate was checked under UV for the presence of UV-active components. A small vertical strip of the plate was cut, stained with anisaldehyde and analyzed for distribution of separated components. Areas of the plate that contained the desired compounds were scraped with a razor blade, ground with a mortar and pestle, eluted with ether, and analyzed by NMR or GC-MS.

2.27 Trimethylsilyl (TMS) Derivatization of Triterpene Alcohols

Samples for GC-MS and GC-FID analysis were injected either underivatized in toluene or as the TMS ether in the derivatization reagent. TMS ethers were made by dissolving each sample in 100 µL pyridine, followed by addition of 100 µL (bis-trimethylsilyl)-trifluoroacetamide (BSTFA). The mixture was kept at 37 °C for 1 h and then used directly for GC-MS or GC-FID analysis.

CHAPTER 3

Cloning and Characterization of Lanosterol Synthase from *Arabidopsis thaliana*:

Discovering Lanosterol Biosynthesis in Plants

This chapter describes functional characterization of the oxidosqualene cyclase AtLSS1, encoded by *At3g45130* in the *Arabidopsis* genome. Phylogenetic analysis of *Arabidopsis* oxidosqualene cyclases (Figure 1.3) showed that the AtLSS1 belongs to the same clade as *Arabidopsis* cycloartenol synthase. Comparison of amino acid sequences has predicted that AtLSS1 is a protosteryl-type oxidosqualene cyclase. Because before this experiment all characterized protosteryl-type cyclases from plants were cycloartenol synthases, it was interesting to uncover the function of the AtLSS1, the second putative protosteryl-type cyclase from *Arabidopsis*. Therefore the functional characterization of AtLSS1 by heterologous expression in yeast was undertaken.

3.1 Previous Work

Initial analysis, cloning, and expression of the *At3g45130* gene were done by Dr. Ling Hua before I joined the project.⁹² Sequences similar to the *Arabidopsis* cycloartenol synthase were obtained by a tblastn⁹³ search against the *A. thaliana* genome.⁹⁴ A cDNA corresponding to the *A. thaliana* coding sequence *At3g45130* was cloned by hybridizing a radiolabeled probe obtained by PCR-amplification of cDNA. After screening 3×10^5 colonies from an *A. thaliana* young seedling cDNA library,⁹⁵ a clone with a 2.4 kbp insert was obtained and named pLH1.1. The *Not* I fragment of pLH1.1 was subcloned into pBluescript II KS (+) to give pLH1.2. The plasmid pLH1.2 was sequenced, revealing that the second exon of LH1.2 (62 amino acids) was absent. A *Not* I – *Sal* I fragment was

cloned from the plasmid pLH1.2 into the galactose-inducible yeast expression vector pRS316GAL to give pLH1.19. The initial clones were transformed into SMY8 yeast strain, however no expression was detected and no oxidosqualene cyclization products were observed.

3. 2 Experimental Procedures

3.2.1 Cloning of the Full-Length *At3g45130*

To obtain a full-length *AtLSS1* cDNA, the primer pair LH1-SalI-F 5'-TAATgtcgcacTAATATGTGGAGGTTAAAGTTA-3' and LH1-BsrGI-R 5'-TATGAGAGCACgtgacaAAACATGGTGCTATT-3' was designed to amplify the 538 bp at the 5' end of the expected coding sequence and thereby obtain a properly spliced fragment.⁹⁶ These primers were used to PCR-amplify an *A. thaliana* cDNA. The PCR reaction mixtures (50 μ L) contained 0.2 μ g of the 7 day-old seedlings cDNA pool, 20 μ L 2.5 \times Triple Master Mix, 20 pmol of each primer, and 3 units of Triple Master polymerase. The amplicon was gel-purified and cloned into pLH1.19 using Sal I and BsrG I restriction enzymes. The resultant expression plasmid, named pLH1.25, was confirmed to contain the full-length ORF (2271 bp), including the exon that was missing previously. The plasmid pLH1.25 was transformed into two yeast strains, SMY8 (lanosterol synthase deletion mutant) and RXY6 (lanosterol synthase and squalene epoxydase double deletion mutant) and the obtained transformants were plated on selective media.

3.2.2 Heterologous Expression of *At3g45130* in Yeast

To study *in vitro* expression of the AtLSS1 in yeast, the RXY6[pLH1.25] was grown to saturation in 100 mL induction medium. The 1.1-g pellet was lysed by vortexing for 10 min with 1 mL acid-washed glass beads, 1.1 mg synthetic racemic 2,3-oxidosqualene, and 1.1 mg Tween 80 and 1.1 mL of sodium phosphate buffer pH 6.2. The lysates were incubated for 30 h at 25 °C. A reaction with a 0.5 g of yeast pellet incubated without oxidosqualene was used as a control. The suspensions were extracted three times with 10 mL hexane, and the combined organic extracts were concentrated to dryness and further studied by GC-FID, GC-MS, and NMR.

To obtain enough triterpene material for detailed analysis, characterization and ratio determination of AtLSS1 minor products a large scale (2L) RXY6[pLH1.25] culture was grown in inducing medium. A total of 15 g cell pellet was collected by centrifugation. The cell pellet was lysed by vortexing 3 × 5 min with glass beads (5 mL) in presence of 15 mL of 100 mM sodium phosphate buffer (pH 6.2). After homogenization, 300 µL of mixture containing 20 mg/mL racemic oxidosqualene and 20 mg/mL Triton X-100 was added to the cell lysate and incubated overnight at room temperature. After incubation the cell homogenate was extracted 4 × 20 mL of ethanol. Ethanol fractions were combined together and concentrated under vacuum. The resultant residue was dissolved in 15 mL of water and extracted with 4 × 30 mL of hexane. The combined hexane fractions were concentrated to dryness. The extract was further subjected to a short silica gel column to remove oxidosqualene, using 5.5 g of silica gel and eluting with 2% ether in hexane until oxidosqualene was completely removed. The remaining organics were eluted from the column with ether, and this fraction lacking

oxidosqualene was further subjected to preparative TLC (1:1 hexane/ether). The silica from the preparative TLC plate above ergosterol (where all triterpene alcohols should run) was scraped from the plate and eluted with ether. Eluted fraction was further analyzed by ^1H NMR.

For in vivo studies, a 100-mL culture of SMY8[pLH1.25] was grown to saturation in induction medium. After 2 days at 30 °C, a 1.2 g cell pellet was collected and saponified with 6 mL 10% KOH in 80% ethanol and 2 mg of the antioxidant butylated hydroxytoluene. The suspension was deoxygenated under a nitrogen stream for 5 min and was incubated at 70 °C for 2 h. The reaction mixture was diluted with 10 mL of water and partitioned with 3 × 10-mL aliquots of hexane. The combined hexane layers were washed with distilled water and brine followed by evaporation in vacuo to give the nonsaponifiable lipid (NSL) fraction, which was analyzed by GC-MS and NMR.

To study if the *Arabidopsis* lanosterol synthase could complement the lanosterol synthase deletion, the SMY8[pLH1.25] strain was grown on sterol-free inducing medium, solidified with agar. Growth on this media would indicate the strain was able to rescue lanosterol, and thus ergosterol, biosynthesis.

3.3 Results

3.3.1 Heterologous Expression of *Arabidopsis* Lanosterol Synthase in Yeast

The plasmid pLH1.25 was expressed in the yeast lanosterol synthase deletion mutant SMY8 to generate oxidosqualene cyclization products in vivo. Nonsaponifiable lipids were obtained from an induced culture of SMY8[pLH1.25]. GC-MS analysis of the trimethylsilyl ethers revealed a component that was absent in an uninduced control, with

the mass spectral fragmentation pattern and retention time (9.88 min) of the product experimentally indistinguishable from a trimethylsilyl lanosterol standard. No other signals with m/z equal to 498 (corresponding to the trimethylsilyl ether of a $C_{30}H_{50}O$ triterpene alcohol) were observed within 1% peak intensity of lanosterol (Figure 3.1). The 1H NMR spectrum exhibited signals characteristic of lanosterol at δ 0.690 (H-18, s), 0.811 (H-31, s), 0.876 (H-32, s), 0.983 (H-19, s), and 1.001 (H-30, s)⁹⁷ and of ergosterol at δ 0.631 (H-18, s), 0.948 (H-19, s), 0.918 (H-28, d, 6.8 Hz), 1.037 (H-21, d, 6.7 Hz), 5.385 (H-7, dt, 5.6, 2.7 Hz), and 5.573 (H-6, dd, 5.6 Hz).⁹⁸ 1H NMR analysis revealed no signals corresponding to the alternative sterol precursors cycloartenol⁹⁹ and parkeol¹⁰⁰ or the monocyclic compounds achilleol A¹⁰¹ and camelliol C¹⁰² (within a 1% detection limit). That SMY8 completely lacks lanosterol synthase activity was established through control reactions previously.⁹⁰ These data unambiguously established that recombinant *At3g45130* is a lanosterol synthase.

The lanosterol yield was calculated to be 4 μ g/100 mL by comparing the H-18 signal intensity of lanosterol to that of the internal standard epicoprostanol (δ 0.641). 1H NMR and GC-MS of a control NSL similarly obtained from a culture grown on glucose did not reveal any signals corresponding to lanosterol or any other triterpene monoalcohol products.

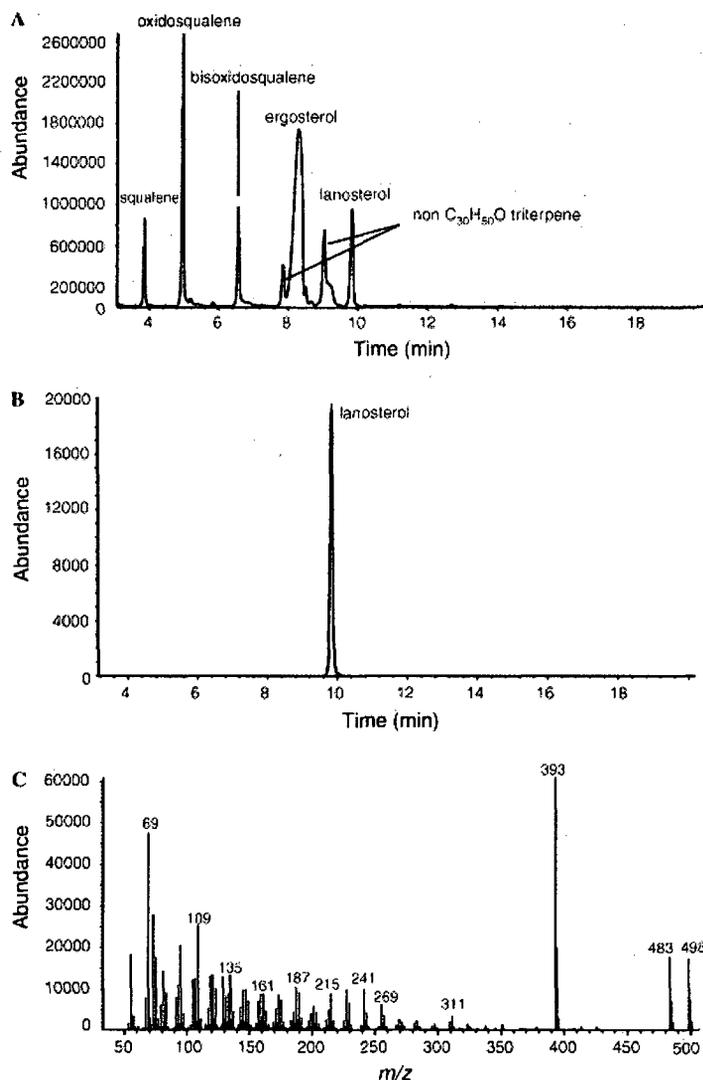


Figure 3.1. GC-MS analysis of the total crude NSL from SMY8[pLH1.25] after induction with galactose. (A) Total ion chromatogram (signals labeled as “non C₃₀H₅₀O triterpene” are oxidosqualene cyclase products further metabolized by yeast enzymes). (B) Selective ion chromatogram of the components with m/z 498. (C) Electron-impact mass spectrum of the peak from total ion chromatogram (B) that demonstrates a mass spectrum identical to a TMS-lanosterol standard.

The cell homogenate of RXY6[pLH1.25] cultured in induction medium was incubated with synthetic oxidosqualene. GC-MS and ¹H NMR analysis of the crude

hexane extract of the reaction mixture demonstrated lanosterol production (2.6 $\mu\text{g}/100$ mL), which was absent in a control reaction with no oxidosqualene added. No signals were observed for any other products from oxidosqualene cyclization within a 1% detection limit. These results established that the *At3g45130* gene encodes a lanosterol synthase with catalytic fidelity >99%.

Analysis of the triterpenes obtained from a large scale in vitro reaction showed presence of several minor compounds, including epilanosterol, epiparkeol, isomalabaricatrienol, and parkeol at levels 2.5%, 0.4%, 0.3% and 0.1% correspondingly (Figure 1, Appendix A). Epilanosterol and epiparkeol are not mechanistically accessible from 3(*S*)-oxidosqualene, and their presence is consequently attributed to the ability of AtLSS1 to utilize the 3*R* epimer of 2,3-oxidosqualene from the racemic stock.

Complementation studies with SMY8[pLH1.25] on sterol-free induction medium showed that pLH1.25 allowed sufficient sterol biosynthesis to compensate for the yeast lanosterol synthase mutation. GC-MS and ^1H NMR of the NSL confirmed the presence of lanosterol and ergosterol, which were 5 $\mu\text{g}/100$ mL and 42 $\mu\text{g}/100$ mL, respectively. The low lanosterol to ergosterol ratio indicates that most of the lanosterol was metabolized further by the yeast sterol metabolic pathway.

3.4 Discussion

Heterologous expression of *At3g45130* in yeast showed conversion of oxidosqualene to lanosterol (99%). The gene product of *At3g45130* was named LSS1 (Lanosterol Synthase) and was the first example of a lanosterol synthase cloned from a plant.

Plant lanosterol synthases use catalytic motifs distinct from those in animal, fungal, and protozoal lanosterol synthases, consistent with the independent evolution of these groups from cycloartenol synthase. These residues are most readily discussed in the context of known structure-function relationships in cycloartenol synthase. The active site residues His257, Tyr410, His477 and Ile481 (*Arabidopsis* cycloartenol synthase numbering) are essential in cycloartenol biosynthesis (see Chapter 1). The *Arabidopsis* lanosterol synthase has Asn477 and Val481 residues in the positions where cycloartenol synthases has His477 and Ile481 (Figure 3.2). That these are the catalytically important differences between plant lanosterol synthases and cycloartenol synthases is apparent in light of previous mutagenesis experiments with cycloartenol synthase (for details see Chapter 1). The *AtCAS1* His477Asn Ile481Val double mutations cooperate to promote accurate lanosterol biosynthesis.⁶⁵ The catalytically important His257, Tyr410, Asn477, and Val481 are conserved in several uncharacterized eudicot oxidosqualene cyclases, suggesting that these are also lanosterol synthases (Figure 3.2).

The Asn477 Val481 pair that this *Arabidopsis* enzyme uses to promote lanosterol biosynthesis is distinct from catalytic motifs in known lanosterol synthases. Previously described lanosterol synthases also enlarge the active site with the Val substitution, but they use other motifs to reposition polarity. In kinetoplastid lanosterol synthases, a His477Gln mutation relocates Tyr410 to approximately the right location, but a third unknown mutation is needed to make lanosterol accurately. Opisthokonts move polarity through a Tyr410Thr mutation, but again another compensatory mutation is necessary. The Tyr410Thr Ile481Val double mutant is a lanosterol synthase that generates 25% 9 β -lanosta-7,24-dien-3 β -ol (9 β - Δ 7 lanosterol) and parkeol as byproducts.⁶⁶

	Δ		Δ	Δ										
<i>Trypanosoma cruzi</i> ERG7	C	G	Y	N	G	498	S	Q	S	W	Q	V	S	567
<i>Trypanosoma brucei</i> ERG7	S	G	Y	N	G	386	P	Q	A	W	Q	V	S	619
<i>Saccharomyces cerevisiae</i> ERG7	M	G	T	N	G	381	T	Q	G	Y	T	V	A	455
<i>Saccharomyces pombe</i> ERG7	R	G	T	N	G	380	T	Q	G	Y	T	V	S	450
<i>Candida albicans</i> ERG7	M	G	T	N	G	383	E	Q	G	Y	T	V	S	449
<i>Homo sapiens</i> ERG7	Q	G	T	N	G	384	D	C	G	W	I	V	S	454
<i>Rattus norvegicus</i> ERG7	Q	G	T	N	G	384	D	C	G	W	I	V	A	455
<i>Mus musculus</i> ERG7	Q	G	T	N	G	383	D	C	G	W	I	V	A	455
<i>Stigmatella aurantiaca</i> CAS1	N	G	Y	N	S	321	D	H	G	W	P	I	S	390
<i>Arabidopsis thaliana</i> CAS1	Q	G	Y	N	G	412	D	H	G	W	P	I	S	482
<i>Arabidopsis thaliana</i> CAS1 mut*	Q	G	Y	N	G	412	D	N	G	W	P	V	S	482
<i>Arabidopsis thaliana</i> LSS1	Q	G	Y	N	G	412	D	N	P	W	P	V	S	482
<i>Arabidopsis thaliana</i> LSS1 mut**	Q	G	Y	N	G	412	D	H	P	W	P	I	S	482
<i>Luffa cylindrica</i> OSC2	Q	G	Y	N	G	412	D	N	G	W	P	V	S	482
<i>Taraxacum officinale</i> TRV	Q	G	Y	N	G	410	D	N	G	W	P	V	S	480
<i>Cucurbita pepo</i> CPR	Q	G	Y	H	G	411	D	N	A	W	P	V	S	481

Figure 3.2. Comparison of amino acid residues in lanosterol synthases and cycloartenol synthases. Deltas (Δ) denote essential catalytic positions corresponding to Tyr410, Asn477, and Val481 in *Arabidopsis* LSS1. Blue highlights lanosterol synthases, pink shows cycloartenol synthases. One asterisk (*) shows *Arabidopsis* cycloartenol synthase His477Asn and Ile481Val double mutants. Two asterisks (**) indicate the *Arabidopsis* lanosterol synthase Asn477His Val481Ile double mutant.

The *Arabidopsis* lanosterol synthase predicted protein sequence is closely related (62-66% identical) to cycloartenol synthases from the gymnosperm *Abies magnifica* (Genbank # AF216755), several monocots (*Avena clauda*, *Avena longiglumis*, *Avena prostrata*, *Avena strigosa*, *Avena ventricosa*, and *Costus speciosus*), and a range of eudicots (*A. thaliana*, *Cucurbita pepo*, *Luffa cylindrica*, *Betula platyphylla*, *Pisum sativum*, *Glycyrrhiza glabra*, and *Panax ginseng*).^{17,49,68,78,,103,104,105,106,107} A maximum parsimony tree was generated to reconstruct the relationship between these cycloartenol synthases, the *A. thaliana* lanosterol synthase, putative lanosterol synthases from other plants, and uncharacterized cyclases from the green alga *Ostreococcus tauri* (Genbank # CR954203) and the lycophyte *Selaginella moellendorffii* (Genbank # AC158179) (Figure

3.3). The *A. thaliana* lanosterol synthase forms a clade with a family of uncharacterized enzymes both from other eurosids (*LcOSC2*⁸⁰ from *L. cylindrica* and *CPR*⁶⁸ from *C. pepo*) and from asterids (*OSCPNZ1*¹⁰⁸ from *P. ginseng* and *TRV*⁷⁵ from *Taraxacum officinale*). Their phylogenetic grouping with *At3g45130* suggests that these enzymes may be lanosterol synthases, and the conservation of key catalytic residues supports this position.

The presence of probable lanosterol synthase orthologs in both eurosids and asterids establishes that this clade arose before the divergence of eurosids and asterids. The phylogenetic tree (Figure 3.3) suggests an even earlier origin. In this tree, the plant lanosterol synthases arose after the evolution of angiosperms (marked by the divergence of cycloartenol synthases from the gymnosperm *Abies magnifica* and the angiosperms), but before the monocot cycloartenol synthases diverge from the eudicot cycloartenol synthases. However, it is worth noting that the sequenced rice genome and existing sequence on other monocots lack obvious lanosterol synthase orthologs, and we cannot preclude the possibility that the placement prior to the divergence of monocots and eudicots is an artifact of rapidly mutating lanosterol synthase sequence.

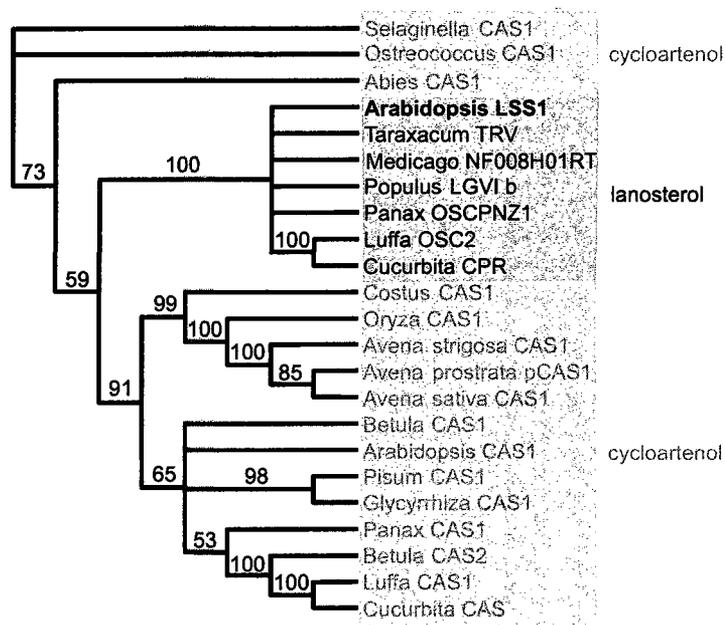


Figure 3.3. Phylogenetic analysis of cycloartenol synthases and lanosterol synthases from plants.

Lanosterol synthases from opisthokonts (animals and fungi) and kinetoplastids were shown previously to have evolved independently in the unikonts and in the euglenids, respectively.¹⁰⁹ These evolutionary events are much more ancient than the origin of higher plants, and plant lanosterol synthases consequently comprise a third evolutionarily distinct class of lanosterol synthases.

The role of lanosterol synthases in plants is still unclear. Early experiments in oxidosqualene cyclization established that lanosterol is the first carbocyclic intermediate in mammals and fungi.^{30,110} Similar experiments in diverse plants from red algae to angiosperms^{111,32,33} uncovered the biosynthesis of cycloartenol rather than lanosterol. Cycloartenol is detected routinely in plants,^{17,49,68,78,103,104,105,106,107} and cloning and genomic information suggest that cycloartenol synthases are ubiquitous plant enzymes.¹¹²

Experiments in tobacco⁴² and blackberry^{31,32} established that cycloartenol is a plant sterol precursor; exogenous cycloartenol is metabolized to stigmasterol, sitosterol, and other phytosterols by a multi-step sequence including the opening of the cyclopropyl ring. The plant sterol biosynthetic pathway can also accept lanosterol,^{42,43,44} which already has the appropriate skeleton to be elaborated into conventional plant sterols. Lanosterol is, in a structural sense, a shortcut to sterols because it lacks the cyclopropyl ring, although some plant enzymes seem to be optimized for substrates with the cyclopropyl group and metabolize lanosterol less efficiently.^{42,44}

Lanosterol has been precluded as a plant sterol biosynthetic intermediate because its biosynthesis cannot be readily detected in plants.^{34,35,36,45,111} The new discoveries do not challenge the long-standing acceptance of cycloartenol as the primary precursor of plant sterols. Although the cycloartenol route to plant sterols clearly predominates, a low-level route through lanosterol may exist. Latex from the eurosid *Euphorbia lathyris* has been shown to biosynthesize lanosterol from oxidosqualene,^{35,45} however this is apparently a secondary metabolic process; *E. lathyris* continues to utilize the cycloartenol pathway to sterols.

The conservation of lanosterol synthases across eudicots in the presence of a cycloartenol route to sitosterol and stigmasterol suggests a role for lanosterol separate from being a membrane sterol precursor. Because plants encode cycloeucaleenol isomerase to open the cyclopropyl ring to the Δ^8 olefin present in lanosterol, most lanosterol metabolites could be biosynthesized readily from cycloartenol, with the notable exception of 4,4-dimethyl Δ^8 sterols, as described in Chapter 1. One possibility is that 4,4-dimethyl Δ^8 sterols derive from lanosterol than cycloartenol, because the 4 β -

dimethyl group blocks access to cyclopropyl ring in cycloeucaleenol isomerase.³¹ Kinetic preferences of other enzymes could divert lanosterol to other compounds distinct from those derived from cycloartenol.

The sporadic detection of lanosterol and lanostane saponins in plants was previously attributed to inaccurate cycloartenol synthases. The discovery of plant lanosterol synthases reveals a more direct pathway to the lanosterol metabolites, and this correlates well with protein expression data. Expression profiles of *Arabidopsis* LSS1 indicate increased expression of *AtLSS1* as a response to methyl jasmonate treatment that serves as a signal for biotic and abiotic stress in plants.¹¹³ Infection with *Pseudomonas syringae* increases lanosterol synthase expression in *Arabidopsis* as well. These data suggest possible defense role of lanosterol or metabolites thereof in plants.

After my discovery of the *Arabidopsis* lanosterol synthase it was also reported by another group.¹¹⁴ Another plant lanosterol synthase was characterized from *L. japonicus* and had similar amino acid residues, corresponding to Asn477 and Val481 in *AtLSS1* confirming that lanosterol synthases are distributed in plants.¹¹⁵

CHAPTER 4

Mutagenesis of Lanosterol Synthase to Biosynthesize Parkeol and Cycloartenol

Discovery of a lanosterol synthase from a plant and the fact that plant lanosterol synthases have evolved from plant cycloartenol synthase has raised a question how the transition between cycloartenol and lanosterol biosynthesis was promoted. Another question was what changes within the active site should be made to change catalytic properties of the enzyme and if it is possible to design an enzyme that would produce more complicated structure than its ancestor does.

4.1 Experimental Procedures

4.1.1 Construction of *Arabidopsis* Lanosterol Synthase Asn477His/Val481Ile

Double Mutant

Two pairs of primers, 5'-CTCCaccggtGACcATCCATGGCCTaTCTCTG-3' with 5'-TAATgtcgacTAATATGTGGAGGTTAAAGTTA-3', and 5'-CAGAGAtAGGCCATGGATgGTCaccggtGGAG-3' with 5'-TAATgcccgcTAATAGTCTATACTCACAAAGA-3', were designed to introduce both Asn477His and Val481Ile mutations into *A. thaliana* lanosterol synthase (AtLSS1). Additional silent mutations were introduced into the primers in order to obtain the restriction site Age I. These primers were used for PCR amplification of two parts of AtLSS1, the front piece (containing the two introduced mutations) and the back piece, using the pLH1.25 plasmid as template. Obtained PCR fragments were gel-purified, digested with Sal I – Age I and Age I – Not I restriction enzymes, respectively, and cloned together into pRS426Gal vector, digested with Sal I and Not I, using a 3-piece

ligation method. The resulting expression plasmid was named pMDK14.3, and sequencing confirmed the presence of the two intended amino acid mutations Asn477His and Val481Ile, and the absence of undesired point mutations.

4.1.2 Construction of *Arabidopsis* Lanosterol Synthase Asn477His and Val481Ile Single Mutants

To construct the Asn477His and Val481Ile single mutants a similar strategy was used, with primers containing the desired single mutation and a silent mutation to introduce the *Age* I restriction site. To prepare the Asn477His mutant the front part of AthLSS1 was PCR-amplified using a pair of primers 5'-CTCCaccggtGACcATCCATGGCCTGTCTCTG-3' and 5'-TAATgtcgacCTAATATGTGGAGGTTAAAGTTA-3'. The 3- end of the Val481Ile mutant was PCR-amplified with the primers 5'-CTCCaccggtGACAATCCATGGCCTaTCTCTG-3' and 5'-TAATgtcgacCTAATATGTGGAGGTTAAAGTTA-3'. The 5' end was identical to that in used in the double mutant, described above. The resultant plasmids, containing Asn477His and Val481Ile single mutants of *Arabidopsis* lanosterol synthase, were named pDAL6.0 and pDAL7.0, respectively.

4.1.3 Yeast Transformations

The yeast strains SMY8 and RXY6 were transformed with the plasmids pMDK14.3, pDAL6.0, and pDAL7.0 using the lithium acetate method to generate yeast strains SMY8[pMDK14.3], RXY6[pMDK14.3], SMY8[pDAL6.0], RXY6[pDAL6.0],

SMY8[pDAL7.0], and RXY6[pDAL7.0]. Transformants were selected on selective medium lacking uracil and solidified with 1.5% agar. To determine whether the obtained mutants biosynthesize enough lanosterol to support yeast growth, SMY8[pMDK14.3], SMY8[pDAL6.0], and SMY8[pDAL7.0] were plated on sterol-free inducing medium.

4.1.4 In Vitro Characterization of *Arabidopsis* Lanosterol Synthase Asn477His and Val481Ile Single Mutants and the Asn477His/Val481Ile Double Mutant

RXY6[pMDK14.3], RXY6[pDAL6.0], and RXY6[pDAL7.0] were grown to saturation in 100 mL of inducing medium. The cells were collected by centrifugation giving 0.94 g, 0.76 g, and 0.92 g cell pellets for RXY6[pMDK14.3], RXY6[pDAL6.0], and RXY6[pDAL7.0], respectively. The cell pellets were lysed with 200 μ L of acid-washed glass beads in 100 mM sodium phosphate buffer pH 6.2. Aliquots (94 μ L, 76 μ L, and 92 μ L, respectively) of a mixture containing synthetic racemic 2,3-oxidosqualene (2 mg/mL) and Triton X-100 (2 mg/mL) were added to each yeast homogenate and left incubating for 30 h at room temperature. After incubation with oxidosqualene, each cell homogenate was quenched with 3 \times 20 mL ethanol and each set of the ethanolic fractions were combined. Water (60 mL) and brine (10 mL) were added to each ethanolic extract, and the triterpene alcohols were extracted with 3 \times 30 mL hexanes. Each extract was subjected to preparative TLC, developed in a 50% hexane – 50% ethyl ether mixture, to separate the triterpene alcohols from ergosterol and oxidosqualene. The fractions were then analyzed by 1 H NMR and GC-MS.

To show that changes in the product profile of AtLSS1 mutants were indeed introduced by the installed desired mutations, each yeast strain RXY6[pMDK14.3], RXY6[pDAL6.0], and RXY6[pDAL7.0] was used for colony PCR amplification of the encoded *AtLSS1* mutant with LH1-SalI-F 5'-TAATgtcgcacTAATATGTGGAGGTTAAAGTTA-3' and LH1-NotI-R 5'-TAATgcggccgcTAATAGTCTATACTCACAAAGA-3' pair of primers. Obtained amplicons were further sequenced.

4.2 Results

4.2.1 Functional Characterization of *Arabidopsis* Lanosterol Synthase Val481Ile and Asn477His Single Mutants

The triterpene fraction of RXY6[pDAL7.0], corresponding to the Val481Ile single mutant of AtLSS1, contained 80% lanosterol and 20% parkeol (established from ¹H NMR spectra, Figure 2, Appendix A). Cycloartenol was not observed among the cyclization products. No other triterpene alcohols were detected within a 1% detection limit.

¹H NMR 800 MHz data of the RXY6[pDAL6.0] triterpene fraction, corresponding to the Asn477His single mutant, allowed for identification of lanosterol (81%), parkeol (16%), and cycloartenol (3%) (Figure 3, Appendix A). Each compound was identified by comparison of chemical shifts of the upfield methyl signals with those of authentic standards. No other triterpene alcohol was detected within a 1% detection limit.

4.2.2 Functional Characterization of *Arabidopsis* Lanosterol Synthase Asn477His/Val481Ile Double Mutant

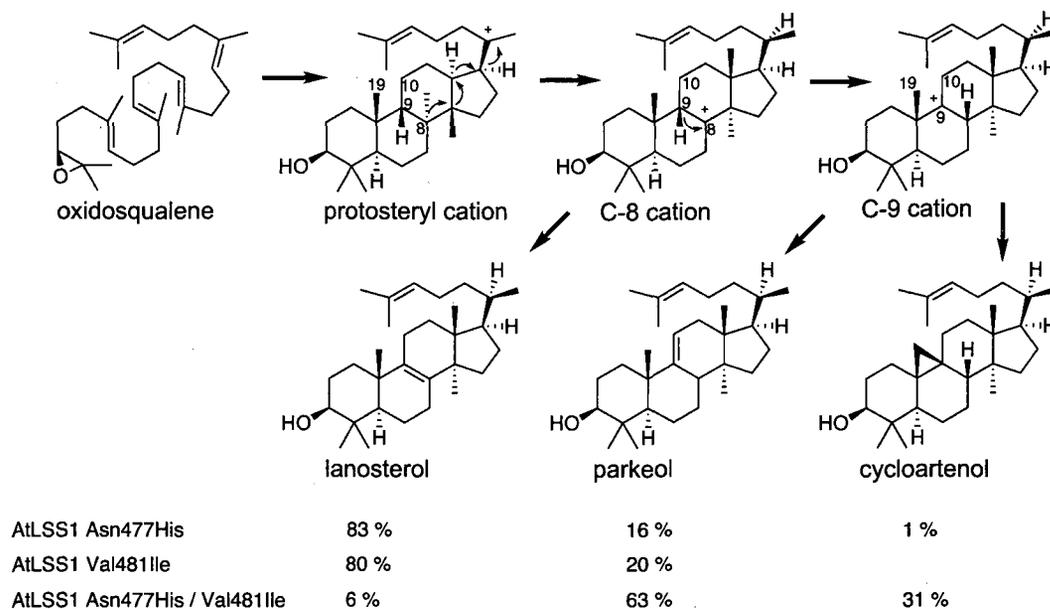
The AtLSS1 N477H V481I double mutant produces 31% cycloartenol, 63% parkeol, and 6% lanosterol (Figure 4, Appendix A). The absence of other compounds, including monocyclic and bicyclic compounds (which move faster on silica), more polar compounds, and alternative deprotonation structures such as ketones or ethers (which move faster on silica) was established through analysis of ¹H NMR data and the absence of methyl shifts characteristic of those compounds (within a detection limit <1%).

4.2.3 Complementation Studies of *Arabidopsis* Lanosterol Synthase Mutants

All three AtLSS1 mutants, including the Asn477His and Val481Ile single mutants and the Asn477His/Val481Ile double mutant, expressed in SMY8 grew on sterol-free inducing medium. This indicated that all three mutants were able to complement lanosterol production in vivo.

4.3 Discussion

In efforts to obtain information on the catalytic differences between *Arabidopsis* lanosterol synthase and cycloartenol synthase, mutagenesis experiments were undertaken in which residues from AtLSS1 were replaced with corresponding residues from AtCAS1. First, the AtLSS1 Val481Ile mutant was expressed in yeast and produced 80% lanosterol and 20% parkeol (Scheme 4.1), as determined by ¹H NMR. In contrast, the corresponding mutation in yeast lanosterol synthase (Val454Ile) had no effect on the product profile.⁶⁵



Scheme 4.1. Cyclization of oxidosqualene by AtLSS1 Asn477His, Val481Ile, and Asn477His/Val481Ile mutants.

This Val481Ile mutation introduced into AtLSS1 a cycloartenol synthase residue that added one methylene group, slightly decreasing the available volume in the active site. The mutant has the same triad of catalytically important residues (Tyr410, Asn477, and Ile481) as the parallel CAS1 His477Asn mutant, which produces 88% lanosterol and 12% parkeol.⁶³ The Ile481 residue in AtCAS1 promotes cycloartenol biosynthesis through steric bulk that positions the intermediate cation relative to His257 and Tyr410, residues that participate in deprotonation at C-19.⁶³

The next mutant studied was the AtLSS1 Asn477His single mutant, which replaced the lanosterol synthase asparagine residue with the corresponding cycloartenol synthase histidine residue. When expressed in yeast, the enzyme still produced lanosterol as the major product (83%) together with parkeol (16%), and cycloartenol (1%). This

product profile suggests that introduced changes in polarity moved the deprotonating base closer to the C19-C9-C10 plane and away from the C9-C8, resulting in biosynthesis of 1% cycloartenol and 16% parkeol. This is the first example in which a single mutation in a lanosterol synthase allowed for biosynthesis of cycloartenol. The analogous mutation was not studied in ScERG7 and therefore cannot be compared.

Modeling and mutation studies of *Arabidopsis* cycloartenol synthase suggest that the His477 residue affects the location of polar His257 and Tyr410 residues situated close to C-9 and C-11.^{60,63,65} The Asn477His mutation in AtLSS1 alters polarity at the 477 position, likely causing the increased production of cycloartenol. The AtLSS1 Asn477His mutant also has the same triad of catalytically important residues (Tyr410, His477, and Val481) as the corresponding AtCAS1 Ile481Val mutant, which makes 54% cycloartenol, 25% lanosterol and 21% parkeol,⁶² indicating importance of the polarity in this position in both AtCAS1 and AtLSS1.

Finally, the AtLSS1 double mutant Asn477His/Val481Ile was expressed in yeast and showed a dramatically different product profile compared to the single mutants. This mutant produced parkeol as the major product (61%), but it also produced a significant amount of cycloartenol (31%). Lanosterol was a minor product, composing only 6% of the total triterpene profile.

These two mutations together have an additive effect in the AtLSS1 Asn477His/Val481Ile double mutant. The parental ability to deprotonate from C-9 and form lanosterol is nearly eliminated by the combination of steric and polar mutations. The increased ratio of cycloartenol to parkeol in the double mutant suggests that the deprotonating base was moved closer to C-19 and C-10. However, while the AtCAS1

double mutant His47Asn/Ile481Val transforms the cycloartenol synthase into an accurate lanosterol synthase,⁶⁴ the analogous transformation in lanosterol synthase does not achieve such accuracy. Additional modifications in the enzyme will be necessary to force the deprotonating base closer to C-19. Nevertheless, the AtLSS1 Asn477His/Val481Ile double mutant produces the largest amount of cycloartenol ever detected in a lanosterol synthase background.

Comparing the effects of mutations in positions 477 and 481 in native and mutant *Arabidopsis* cycloartenol synthase with those in lanosterol synthases from *Arabidopsis* and *S. cerevisiae* helps to illuminate the functions of these key residues (Table 4.1).

Table 4.1. Percentage product composition of *Arabidopsis* cycloartenol synthase (AtCAS1), *Arabidopsis* lanosterol synthase (AtLSS1) and yeast lanosterol synthase (ScERG7) mutants.

Protein, mutation	Cycloartenol	Lanosterol	Parkeol	Ref
AtCAS1, none	99		1	⁴⁹
AtCAS1, I481V	54	25	21	⁶²
AtCAS1, H477N		88	12	⁶³
AtCAS1, H477N/I481V		99	1	⁶⁴
ScERG7, none		100		^{50,51}
ScERG7, V454I		100		⁶⁵
AtLSS1, none		99.5	<0.5	⁹⁶
AtLSS1, V481I		80	20	This work
AtLSS1, N477H	1	83	16	This work
AtLSS1, N477H/V481I	31	6	63	This work

These mutagenesis experiments show that the active site composition of AtLSS1 resembles that of AtCAS1 more closely than that of ScERG7. That the AtLSS1 and

ScERG7 mutants demonstrate such different results also suggests that AtLSS1 likely utilizes a different set of catalytic residues that observed of ScERG7 with similar general mechanism of bond rearrangement. Comparable results between the product profiles of AtCAS1 and the AtLSS1 mutants that have residues Tyr410, Asn477, and Ile481 support the suggestion that the two enzymatic active sites are similar in their overall organization. Application of molecular modeling experiments that were not performed at this point could be a way to confirm the suggestions. It also demonstrates that AtLSS1 has acquired additional catalytically relevant mutations than these two, and that reverse conversion of lanosterol synthase into cycloartenol synthase requires additional changes.

Evolutionary relationships between cycloartenol synthases and plant lanosterol synthases have shown that *Arabidopsis* lanosterol synthase diverged from cycloartenol synthase by acquiring specific mutations that allowed for changes in the deprotonation strategy of the enzymes (see Chapter 3). Previous mutagenesis studies were carried out using AtCAS1 and ScERG7, but these enzymes have a low sequence identity (~40%), making it more difficult to compare catalytic activity and determine evolutionary relationships between the enzyme families. The higher sequence similarity between *Arabidopsis* cycloartenol synthase and lanosterol synthase (65% identity) helps to illuminate potential amino acid residues that allow catalytic changes between the two enzymes.

Previous mutagenesis experiments on *Arabidopsis* cycloartenol synthase have shown that only two mutations, His477Asn and Ile481Val, could convert an accurate cycloartenol synthase (99% cycloartenol) into an accurate lanosterol synthase (99% lanosterol).⁶⁴ All previous efforts to introduce cycloartenol synthase activity into a

lanosterol synthase background have been unsuccessful until this work. Therefore the use of a native *Arabidopsis* lanosterol synthase has been demonstrated to be an effective tool to elucidate catalytically relevant residues for the production of cycloartenol.

During the course of our analyses, a paper appeared in which the Asn477His Val481Ile double mutant was generated in the lanosterol synthase isolated from *Lotus japonicus*.¹¹⁵ The authors reported results similar to present work. Introduction of the Asn477His and Val481Ile mutations into *Lotus* lanosterol synthase (LjLAS) produced 4% lanosterol, 13% cycloartenol, and 83% parkeol, compared to the results observed in our AtLSS1 Asn477His/Val481Ile double mutant that makes 6% lanosterol, 31% cycloartenol, and 61% parkeol. The AtLSS1 and LjLAS1 enzymes are 66% identical, and the LjLAS1 and LjCAS1 shows 66% homology, similar to that between AtCAS1 and AtLSS1. Expression of the mutant *Lotus* lanosterol synthase was done in a yeast strain that only allows in vivo production of triterpenes, and therefore reported ratio of the triterpene products is not reliable, because lanosterol, cycloartenol, and parkeol are metabolized by yeast at different rates. However, general conclusions about the evolutionary relationships between plant cycloartenol synthases and plant lanosterol synthases and the importance of specific active-site residues in the bifurcation of lanosterol and cycloartenol biosynthesis support our results and conclusions.

CHAPTER 5

Cloning and Characterization of Camelliol C Synthase From *Arabidopsis thaliana*: an Enzyme That Makes Monocycles Evolved From Those That Make Pentacycles

This chapter describes functional characterization of the *Arabidopsis* oxidosqualene cyclase (LUP3), encoded by *At1g78955*. Initially this gene was annotated as a fusion with another gene with unknown function (*At1g78950*). LUP3 belongs to the LUP clade of *Arabidopsis* oxidosqualene cyclases (Figure 1.3) and demonstrates high amino acid identity with enzymes that make pentacycles (lupeol and β -amyrin synthases).

5.1 Experimental Procedures

5.1.1 Cloning and Subcloning of *At1g78955* (LUP3, CAMS1)¹¹⁶

To obtain expression construct for heterologous expression of LUP3 in yeast an *A. thaliana* sequence *At1g78955* similar to *AtCAS1* was found by BLASTN search against the *A. thaliana* genome. *A. thaliana* cDNA was made by RT-PCR from mRNA isolated from siliques using a RETROscript kit. Two pairs of primers, ORF2 Sall F 5'-(TAAAgctgcacTAAAATGTGGAAGTTGAAGATAGC)-3' with ORF2 EcoRI R 5'-(TGCAATCCACAAGTAATCAGAAAT)-3' and ORF2 EcoRI F 5'-(TAGCCGTTATATTACCATTGGATGTGTT)-3' with ORF2 NotI R 5'-(TTTAgcggccgcTTTACTAACTTCTTCTCTCTGTT)-3', were used to PCR-amplify the two parts of *At1g78955* from *A. thaliana* cDNA. The two PCR amplicons obtained were gel-purified and cloned into TOPO vector using a TOPO-TA Cloning kit according to manufacturer's instructions, yielding two plasmids named pMDK4.6 and pMDK4.7. The plasmids pMDK4.6 and pMDK4.7 contained the front and back fragments of

At1g78955, respectively. Fragments were digested with Sal I - EcoR I and EcoR I - Not I restriction enzymes, respectively, and the gel-purified fragments were subcloned together into the yeast expression vector pRS426Gal digested with Sal I and Not I. The resultant plasmid containing the 2310-bp ORF was named pMDK4.8. The insert was sequenced and confirmed to be identical to the predicted nucleotide sequence of *At1g78955*.

5.1.2 Functional Characterization of the Camelliol C Synthase by Heterologous Expression in Yeast

Two *S. cerevisiae* strains RXY6 and SMY8 were transformed with the plasmid pMDK4.8 using the lithium acetate method. Transformants were selected on synthetic complete medium lacking uracil.

For in vitro studies, a 100-mL culture of RXY6[pMDK4.8] was grown to saturation in selective medium. The resulting 1.1-g cell pellet was collected by centrifugation and lysed by vortexing with 1.1 mL 100 mM sodium phosphate buffer (pH 6.2) and 10 mL acid-washed glass beads. A 110- μ L aliquot of a solution containing 2.2 mg synthetic racemic 2,3-oxidosqualene and 2.2 mg Triton X-100 dissolved in water was added to the lysate. A control reaction with 0.6 g of the cell lysates was incubated with 600 μ L 100 mM sodium phosphate buffer (pH 6.2) containing 1.2 mg Triton X-100 without addition of oxidosqualene. After 48-h incubation at 25 °C, ethanol (20 mL) was added to the yeast suspension, followed by vortex mixing and centrifugation. The pellet was subjected two more times to this procedure. The combined ethanol fractions (3 \times 20 mL) were diluted with water (40 mL) and the mixture was extracted with hexanes (3 \times 35 mL). The combined extracts were concentrated in vacuo, yielding 56.5 mg of crude

extract. The residue was used for GC-MS analysis and further separation (Figure 5.1). To avoid non-enzymatic cyclization, we removed unreacted oxidosqualene and squalene from the triterpene products and ergosterol with a short silica gel column (6 g, 230-400 mesh, elution with gradients of ethyl ether in hexanes). The triterpene fraction was further subjected to preparative TLC in dichloromethane, giving three bands. Each of those was scraped onto small columns and eluted with ethyl ether.

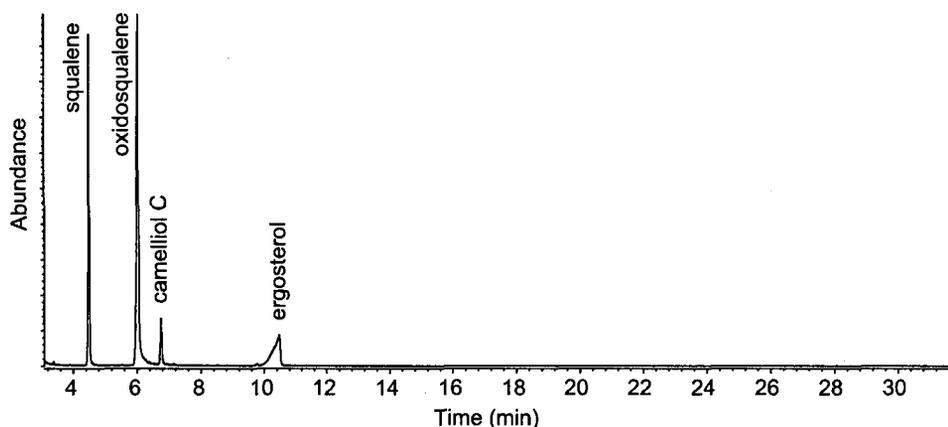


Figure 5.1. GC-MS analysis of the crude extract of RXY6[pMDK4.8] incubated with racemic oxidosqualene: total ion chromatogram, oven temperature 260 °C.

To study *in vivo* expression of LUP3, a 4-L yeast culture of SMY8[pMDK4.8] was grown to saturation in inducing medium at 30 °C. A 41.3-g yeast pellet was collected and saponified with 160 mL ethanolic potassium hydroxide (10% KOH in 80% EtOH) and 80 mg butylated hydroxytoluene (BHT, as an antioxidant agent) at 70 °C for 2 h. The reaction mixture was diluted with an equal amount of water (160 mL) and extracted with 5 × 100 mL hexanes. The organic extracts were combined, washed with water, and concentrated *in vacuo* to a residue (0.256 g) of nonsaponifiable lipids (NSL). An aliquot (1%) of the NSL was analyzed by GC-MS (Figure 5.2) and ¹H NMR, and the remaining

NSL was subjected to column chromatography (40 g silica gel, elution with 2% ether in hexane for 6 fractions followed by elution with ether). A total of 10 fractions were collected and analyzed by GC-MS and ^1H NMR.

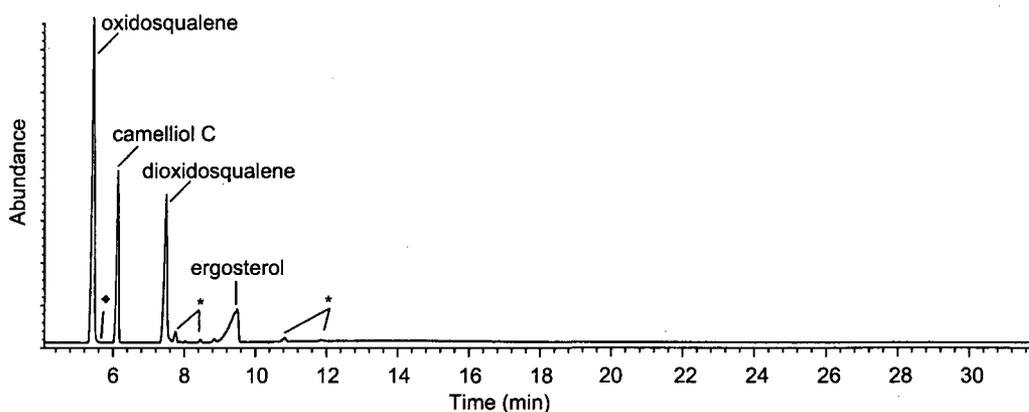


Figure 5.2. GC-MS analysis of the crude NSL of SMY8[pMDK4.8]: total ion chromatogram, oven temperature 270 °C. Asterisks (*) denote non-triterpene components, as judged by their mass spectra and the black diamond (◆) corresponds to achilleol A.

5.2 Results

5.2.1 Functional Characterization of the Camelliol C Synthase

All three fractions from the RXY6[pMDK4.8] preparative TLC were analyzed by 800 MHz ^1H NMR. The first band contained the major product camelliol C¹⁰² together with achilleol A¹⁰¹ (Figure 5.3). In the second fraction, β -amyrin⁹ was identified on the basis of several resolved upfield methyl signals that matched those of an authentic standard within 0.0004 ppm. These methyl signals showed very similar splitting behavior upon resolution enhancement, notably splitting of the singlet at 1.136 ppm into a doublet ($J = 0.9$ Hz) in spectra of both the standard and the second fraction. The third fraction

contained mainly of ergosterol with no detectable triterpene alcohols. Whereas the *in vitro* reaction with oxidosqualene generated camelliol C, achilleol A, and β -amyryn in a 98:2:0.2 ratio, no triterpenes were detected in the parallel control reaction lacking oxidosqualene.

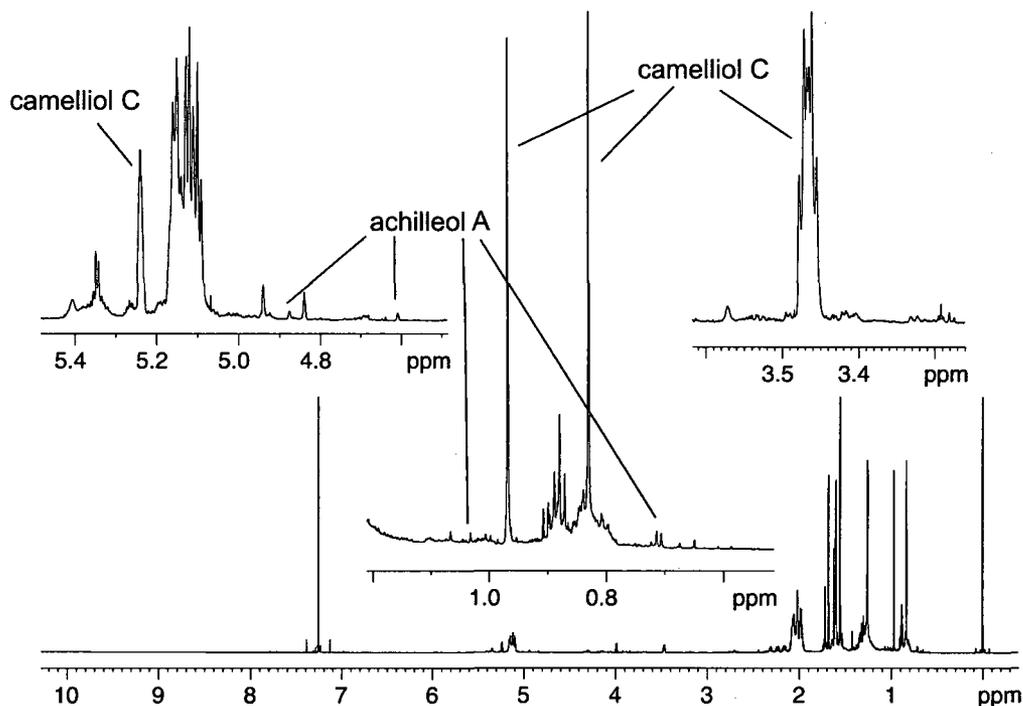


Figure 5.3. ¹H NMR spectrum of the first PTLC fraction from the RXY6[pMDK4.8] *in vitro* reaction.

Analysis of the fractions obtained by column chromatography separation of the NSL obtained from SMY8[pMDK4.8] showed that fractions 1-4 contained squalene, oxidosqualene, and dioxidosqualene. Fractions 6-10 contained ergosterol, ergosterol derivatives, acetylated glycerols, and epoxide ring opening products of oxidosqualene and dioxidosqualene that occurred during saponification (resulting in diol and hydroxy-

ethoxy derivatives). Only fraction 5 contained cyclized triterpene alcohols, namely the major product camelliol C and minor products achilleol A and β -amyryn (Figure 5.4).

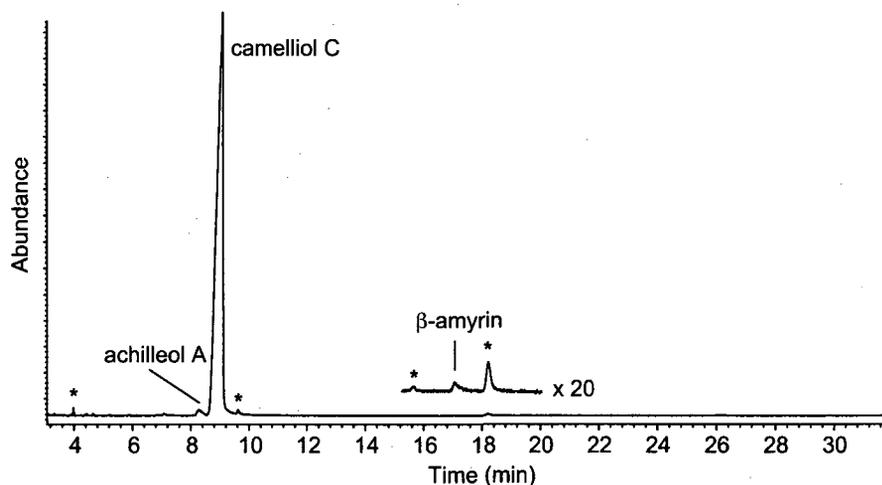
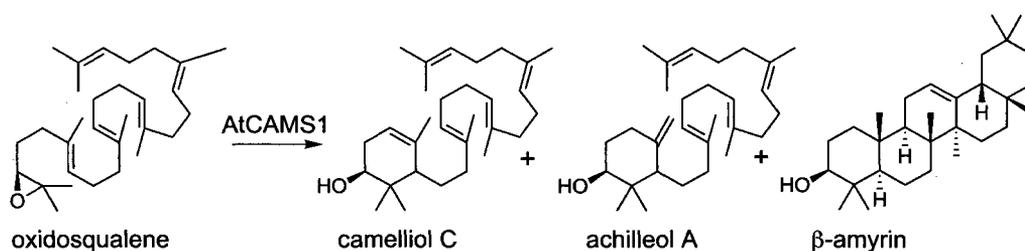


Figure 5.4. GC-MS analysis of fraction 5 from column separation of the crude SMY8[pMDK4.8] NSL: total ion chromatogram of underivatized triterpenes, oven temperature 260 °C. Asterisks (*) denote non-triterpene components, as identified through their mass spectral patterns.

The three products camelliol C, achilleol A, and β -amyryn were separated by HPLC and characterized by GC-MS and 800 MHz ^1H NMR (Figures 5 – 7, Appendix A). HPLC fractions 44-45, 47-52, and 76-79 contained achilleol A, camelliol C, and β -amyryn, respectively. No other HPLC fractions showed the presence of triterpene alcohols by 800 MHz ^1H NMR. Additionally, HSQC, HMBC, DEPT, and ^{13}C NMR spectra of camelliol C were collected using triterpene material from another CAMS1 in vivo experiment, which gave similar results.

5.3 Discussion

Most known triterpene synthases make major products containing four or five rings.²¹ The *Arabidopsis* oxidosqualene cyclase encoded by *At1g78955* converts oxidosqualene to the monocyclic triterpene camelliol C (98%) and two minor by-products: the monocycle achilleol A (2%) and the pentacycle β -amyrin (0.02%). This enzyme, CAMS1, is the first example of a native triterpene synthase that makes an A-ring monocyclic triterpene as the dominant product (Scheme 5.1).



Scheme 5.1. CAMS1 cyclizes oxidosqualene to monocyclic triterpenes and the pentacycle β -amyrin.

CAMS1 is 70-78% identical to the *Arabidopsis* enzymes LUP1, LUP2, and LUP5 that predominantly make penta- and tetracycles. More distantly related are families of enzymes that make the pentacycles β -amyrin (72-75% identical) and lupeol (59-61% identical).²¹ *Arabidopsis* enzymes that predominantly make tricycles (THAS1 and PEN1)⁷⁰ and B-ring monocycle marneral (MRN1)⁶⁹ are more distantly related than some β -amyrin synthases. Phylogenetic analysis suggests that camelliol C synthase evolved from β -amyrin synthases (Figure 5.5). One might imagine that polycyclic triterpene synthases evolved from enzymes that form smaller ring systems by iterative addition of

motifs favoring an extra ring. However, CAMS1 shows the reverse evolutionary order, being a descendant of enzymes that form polycyclic triterpenes.

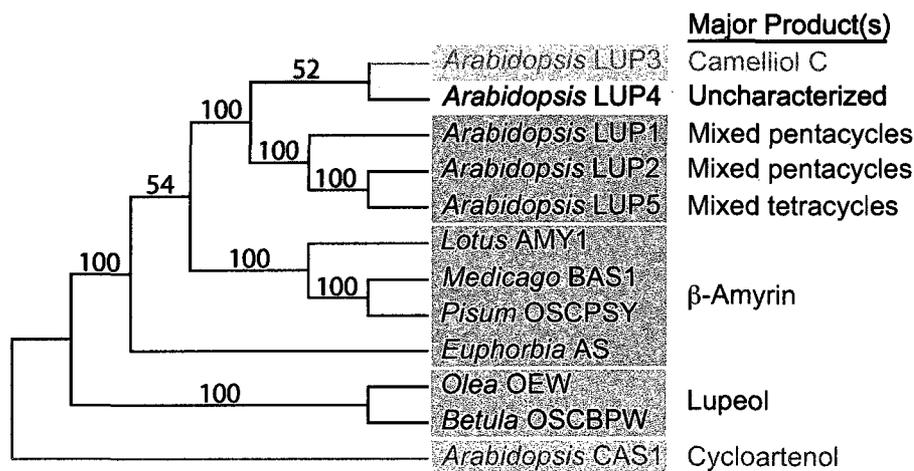


Figure 5.5. Phylogenetic analysis of CAMS1 and related plant oxidosqualene cyclases.

Camelliol C synthase demonstrates high product specificity compared to many enzymes involved in secondary metabolism (e.g., *Arabidopsis* arabidiol synthase, baruol synthase and lupeol synthase). Such product accuracy can be attributed to a short biosynthetic route that stops cyclization immediately after A-ring formation. Accuracy would require simply excluding rearrangement, further cyclization, and alternate deprotonation. If a small portion of the monocyclic carbocation escapes deprotonation, it undergoes further annulation to the bicyclic carbocation and continues cyclization to β-amyrin.

To understand how camelliol C synthase produces monocycles, we compared CAMS1 amino acid residues with available sequences of other oxidosqualene cyclases. The active-site residues of CAMS1 and β-amyrin synthases are identical except for Ala484, which is Val or Ile in nearly all other plant cyclases (Figure 5.6). Previous

mutation studies^{61,65} suggested that decreased steric bulk at this position (notably mutations to alanine or glycine) promotes the formation of monocycles. For example, Ile481Ala mutation of AtCAS1 results in the production of 13% achilleol A and 6% camelliol C, and Ile481Gly shows 44% achilleol A and 12% camelliol C.⁶¹ The analogous Val454Ala mutation in yeast lanosterol synthase allows for 5% biosynthesis of achilleol A.⁶⁵ Notably, in all experiments, achilleol A production was dominant over camelliol C. Mutations in the corresponding positions of dammarenyl type cyclases have not been studied. However, we predict that cyclases utilizing smaller residues (alanine or glycine) in this position may likewise be compromised in B-ring formation. An uncharacterized cyclase from *Betula platyphylla* (*OSCBPD*),¹¹⁷ for example, encodes alanine at this position and may produce monocycles as well.

	*	
<i>Olea europa</i> OEW	H G W Q V S D C T	485
<i>Betula platyphylla</i> OSCBPY	H G W Q V S D C T	487
<i>Euphorbia tirucalli</i> AS	H G W Q V S D C T	487
<i>Medicago truncatula</i> BAS1	H G W Q V S D C T	487
<i>Glycyrrhiza glabra</i> AS1	H G W Q V S D C T	487
<i>Lotus japonicus</i> AMY1	H G W Q V S D C T	475
<i>Panax ginseng</i> OSCPNY1	H G W Q V S D C T	488
<i>Pisum sativum</i> OSCPSY	H G W Q V S D C T	487
<i>Arabidopsis thaliana</i> PEN5 (MRN1)	Q G L P I S D G T	489
<i>Arabidopsis thaliana</i> LUP3 (CAMS1)	H G W Q A S D C T	488
<i>Betula platyphylla</i> OSCBPD	H G W Q A S D C T	460
<i>Arabidopsis thaliana</i> LUP4	H G W Q V S D C T	488
<i>Arabidopsis thaliana</i> LUP2	H G W Q V S D C T	488
<i>Arabidopsis thaliana</i> LUP1	H G W Q V S D C T	485
<i>Arabidopsis thaliana</i> LUP5	H G W Q V S D C T	488
<i>Arabidopsis thaliana</i> CAS1	H G W P I S D C T	485
<i>Saccharomyces cerevisiae</i> ERG7	Q G Y T V A D C T	458
<i>Homo sapiens</i> LSS	C G W I V S D C T	457

Figure 5.6. Partial alignment of oxidosqualene cyclase amino acid sequences.^{21,117,118} An asterisk (*) denotes the position corresponding to Ala484 in CAMS1, corresponding to Val481 in *Arabidopsis* LSS1. Active site residues alignment is based on a comparison with human lanosterol synthase.³

If the transition between enzymes that make polycycles and monocycles requires relatively few mutations, then evolution of this kind could be more frequent, which is consistent with the punctate distribution of monocycles across the vast diversity of higher plants. Camelliol C and achilleol A have been found in asterids (*Camellia sasanqua*, *Camellia japonica*, *Achillea odorata*, *Bupleurum spinosum*, and *Santolina elegans*), two eurosids (*Euphorbia antiquorum* and *Garcinia speciosa*),^{101,102,119,120,121} the monocots wheat and rice, and the fern *Polypodiodes formosana*.^{102,122,123,124}

The dominance of camelliol C instead of achilleol A in the product profile of CAMS1 suggests that camelliol C or a metabolite thereof provides a competitive advantage that achilleol A does not replicate. While the biological role of enzymes and their products can often be illuminated by microarray data, the initial annotation of the *Arabidopsis* genome included *At1g78955* (*CAMS1*) as a fusion with *At1g78950* (*LUP4*), and these genes were not distinguished in early microarrays. The limited available nonarray expression data (*Arabidopsis* MPSS Plus: Gene Analysis) show elevated amounts of *At1g78955* mRNA in *Arabidopsis* inflorescence tissue.¹²⁵ However because the data is limited, the actual distribution of camelliol C may be not limited to the flowering parts of the plant. Further investigation is necessary to establish the actual biological role of camelliol C.

Because monocyclic and polycyclic triterpenes have rather different spectral and chromatographic signatures, most monocycles were described as components of plant oils (usually lacking polycyclic triterpenes) rather than in surveys of triterpene distribution. This suggests that monocyclic triterpenes are more widespread than literature suggests.

CHAPTER 6

Cloning and Characterization of Arabidiol Synthase: Water Addition in Triterpene Biosynthesis

This chapter describes functional characterization of another oxidosqualene cyclase from *Arabidopsis*, PEN1. The oxidosqualene cyclase PEN1 encoded by *At4g15340* belongs to PEN clade of *Arabidopsis* oxidosqualene cyclases (Figure 1.3). Functional characterization of several other cyclases from the same clade has resulted in identification of cyclases with novel catalytic activities that make either previously unknown triterpene structures (e.g., thalianol synthase)⁷⁰ or perform mechanistically unusual oxidosqualene cyclization (e.g., marneral synthase).⁶⁹ Therefore, the closely related to thalianol synthase PEN1 was a good candidate for analysis with potential for discovery of novel triterpene products or enzymatic functionalities.

The PEN1 characterization was once attempted in the past, but no cyclization products were found.²² I undertook functional characterization of PEN1 by heterologous expression in yeast resulting in successful identification of PEN1 cyclization products.

6.1 Experimental Procedures

6.1.1 Cloning and Subcloning of *At4g15340* (*PEN1*)

To design an expression construct containing full-length *PEN1*, the *A. thaliana* sequence *At4g15340* similar to *AtCAS1* was found by a BLASTN search against the *A. thaliana* genome. An *Arabidopsis* cDNA library was made from mRNA¹²⁶ isolated from 7-day old seedlings with use of a RETROscript kit (Ambion) and was used for PCR amplification with the pair of primers 5'-

(TAAAgtcgacTAAAATGTGGAGACTAAGAATTGGAGCT)-3' and 5'-
(TTTAgcggccgcTTTATCAAGGCTGAAGCC)-3'. The PCR-amplified fragment was gel-purified, digested with *Sal* I and *Not* I restriction enzymes, and cloned into the yeast expression vector pRS426Gal. The resultant plasmid containing the 2301-bp ORF was named pMDK3.6. The insert was sequenced and confirmed to be identical to *At4g15340*.¹²⁷

6.1.2 Functional Characterization of the Arabidiol Synthase by Heterologous Expression in Yeast: Analysis of the Dominant Compound Arabidiol

The plasmid pMDK3.6 was used to transform *S. cerevisiae* strains RXY6 and SMY8. Transformants were selected on synthetic complete medium lacking uracil, solidified with 1.5% agar, and supplemented with 2% glucose, 13 mg/L hemin chloride, 20 mg/L ergosterol, and 5 g/L Tween 80.

A 2-L RXY6[pMDK3.6] culture was grown to saturation in selective inducing medium, yielding an 18-g cell pellet, which was collected by centrifugation and lysed by homogenization using a cell disruptor with 30 mL of 100 mM sodium phosphate buffer (pH 6.2). An ethanol solution (400 μ L) containing 8 mg synthetic racemic 2,3-oxidosqualene and 8 mg Triton X-100 was added to the cell homogenate and this mixture was incubated at room temperature for 24 h. After incubation, ethanol (30 mL) was added to the cell homogenate and the mixture was vortexed and centrifuged to separate ethanolic fraction from the cellular debris. After removal of the ethanolic fraction, the cell debris was again extracted with 30 mL ethanol, followed by 2 \times 30 mL of hexane. The combined ethanolic and hexane supernatants were concentrated in vacuo. To the

obtained extract, 30 mL water was added and the mixture was extracted with hexanes (3 × 30 mL) and 2 × 30 mL methyl *tert*-butyl ether (MTBE). The combined hexanes and MTBE extracts were concentrated in vacuo and 5% of the crude extract was used for GC-MS and ¹H NMR analysis. The remainder of the residue was subjected to a short silica gel column run with 2% ether in hexane to remove unreacted oxidosqualene and to prevent the possibility of non-enzymatic cyclization during the remaining workup. A portion (5%) of the purified triterpene fraction was used for ¹H NMR analysis to establish the ratio of the major component to the minor compounds. The remainder of the extract was further subjected to preparative TLC developed in 1:1 ether/hexane, allowing separation of the minor products from and ergosterol and the major triterpene product for further detailed analysis of the minor compounds ¹H NMR. A 1-L culture of SMY8[pMDK3.6] was grown to saturation at 30 °C in inducing medium containing 2% galactose, 13 mg/L hemin chloride, 20 mg/L ergosterol, 5 g/L Tween 80, and synthetic complete medium lacking uracil. The resulting 8.2 g yeast pellet was collected and saponified with 45 mL of ethanolic potassium hydroxide (10% KOH, in 80% EtOH and 20 mg BHT) at 70 °C for 2 h. The reaction mixture was diluted with an equal amount of water (45 mL) and extracted with 3 × 50 mL of hexanes. The organic extracts were combined and concentrated in vacuo. This nonsaponifiable lipid (NSL) fraction was analyzed by GC-MS, GC-FID and ¹H NMR.

The medium that remained after recovering yeast SMY8[pMDK3.6] from a 1-L culture was used to identify possible oxidosqualene cyclization products that could have been emitted from the yeast. Diaion HP-20 resin (15 g) was activated with 50 mL methanol for 15 min, and the resin was added to the medium after removing the methanol

by decanting. The medium was incubated with activated resin overnight at 30 °C while shaking (200 rpm) and the resin was recovered from the medium by filtering and extracted with 4 × 50 mL ethanol. The ethanol fractions were combined together and concentrated in vacuo. The obtained extract was further subjected to silica-gel column chromatography (10 g of silica), eluting with ether. The collected ether fractions (12 mL each) were combined, concentrated in vacuo, and used for GC-MS analysis.

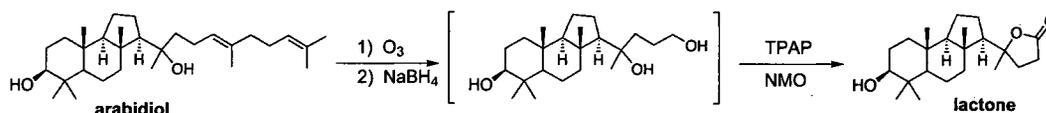
To obtain sufficient material to determine structures of minor compounds, a large-scale (30 L) culture was grown in inducing medium. A 180-g cell pellet was collected, saponified for 2 h at 70 °C with 450 mL of 10% KOH in 80% EtOH and 360 mg BHT. The cooled reaction mixture was extracted with 5 × 150 mL of hexanes. The combined hexane fractions were washed with water and brine and concentrated in vacuo. A small aliquot of the resulting residue of NSL (0.76 g) was analyzed by GC-MS. The remaining NSL residue was subjected to chromatography on a silica gel column using 100 g silica gel as the stationary phase and gradients of 2% - 50% ether in hexanes as the mobile phase. Collected fractions were combined, giving a total of 16 fractions. Each fraction was analyzed by GC-MS and ¹H NMR and was further purified by HPLC or PTLC, as necessary.

6.1.3 Degradation of the Major Product to the Corresponding Lactone

In principle, the C14 stereochemistry of arabidiol could be resolved through a series of molecular modeling experiments to predict the NMR chemical shifts of both epimers, and comparing these values to experimental results. However, the flexible side chain of arabidiol introduced a large number of conformers that complicated calculations.

To reduce the number of conformers, I obtained 5 mg of arabidiol from in vivo culture and an undergraduate student Allie Obermeyer modified arabidiol¹²⁸ to a lactone through a sequence of reactions that preserved the initial C-14 configuration of arabidiol (Scheme 6.1). Material for chemical degradation of the lactone was obtained from the extract of a large-scale culture purified by a silica gel column. Fraction 11, containing arabidiol and ergosterol, was subjected to PTLC (developed with 50% ether in hexanes). Ozonolysis of arabidiol was conducted in a CH₂Cl₂/CH₃OH solution (1:1) at -78 °C, followed by treatment with NaBH₄. The triol intermediate was oxidized to the lactone with *N*-methylmorpholine-*N*-oxide (NMO), tetrapropylammonium perruthenate (TPAP), and 4 Å molecular sieves in CH₂Cl₂ at room temperature.

Scheme 6.1. Synthesis of lactone.



6.1.4 Purification of Minor Products

Most minor compounds were present in column fractions 8 and 9, with just traces of triterpenes in fractions 7 and 10. During column chromatography of the large scale in vivo extract, the precautions against non-enzymatic cyclization were not undertaken and therefore some of the mono-, bi-, and tricyclic compounds may be artifacts.¹²⁹

From the 30-L NSL column fractions, fractions 1 – 7 were analyzed by GC-MS and ¹H NMR and did not require further purification or analysis. Fractions 10 and 12 – 16 were analyzed and did not undergo further purification or analysis. Fraction 11 was further purified by PTLC (see section 6.1.4) to separate ergosterol from arabidiol.

Fraction 8 (12.2 mg, dissolved in 80 μ L methanol) was further subjected to preparative HPLC separation using a reverse-phase column. A mobile phase of 95% methanol with 5% was used to separate the mixture of triterpene alcohols, and the eluted material was collected and divided into approximately 200 fractions. HPLC fractions were combined based on UV detection results and were further analyzed by ^1H NMR (800 MHz) and GC-MS.

Fraction 9 (14.6 mg) was dissolved in methanol with 5% MTBE (total 150 μ L) and split in half. Both parts were subjected to reverse-phase HPLC with a 95:5 methanol-water mobile phase at a flow rate of 7.5 mL/min. Collected fractions (over 200 in each HPLC run) were combined together based on UV activity detected by HPLC and were rigorously analyzed by 800 MHz ^1H NMR. All compounds found in the HPLC fractions were either identified by comparison to authentic standards or literature data or were analyzed by ^1H , ^{13}C and 2D NMR (HSQC, HMBC, NOESY) for full spectral data.

6.2 Results

6.2.1 Analysis and Structure Determination of the Major Product Arabidiol

GC-MS analysis of extracts obtained from both in vitro (RXY6) and in vivo (SMY8) experiments demonstrated the presence of a dominant oxidosqualene cyclization product with m/z 444 underivatized and m/z 514 derivatized, corresponding to a putative triterpene diol (Figure 6.1). Further ^1H NMR analysis showed five upfield methyl singlets at 0.788 ppm, 0.859 ppm, 0.972 ppm, 0.979 ppm, and 1.268 ppm, suggesting a tricyclic triterpene skeleton. The more downfield singlet at 1.268 ppm was likely to be adjacent to a second alcohol group, further indication of a triterpene diol.

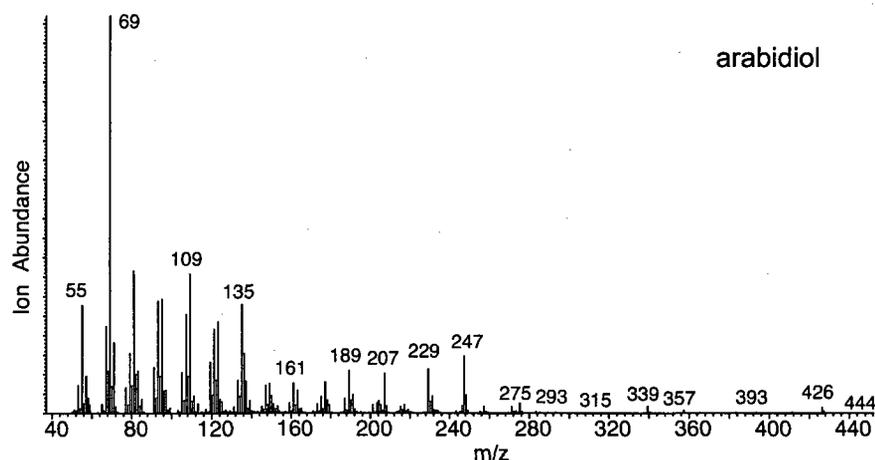


Figure 6.1. Electron impact mass spectrum of the underivatized triterpene diol (arabidiol).

Initial NMR analysis (^1H , ^{13}C , DEPT, HMBC, and HSQC spectra are shown in Appendix A, Figures 5-9, respectively) allowed for determination of the basic structure and connectivity of the major product arabidiol.¹²⁷ Arabidiol was determined to be a tricyclic triterpene with hydroxyl groups at C-3 and C-14. The stereochemistry of most positions was determined by NOE experiments (Figure 10, Appendix A). However, establishing the C-14 configuration of arabidiol was nontrivial. The C-14 stereocenter is located in the side chain, where free rotation about C–C bonds generates conformational heterogeneity that complicates interpretation of NOE results. To reduce the number of conformers, arabidiol was modified to a lactone through a sequence of reactions that preserved the initial C-14 configuration of arabidiol (Scheme 6.1).

To resolve the C-14 stereochemistry of the lactone and arabidiol, a series of quantum mechanical calculations was performed by William K. Wilson.¹²⁷ The quantum mechanical calculations¹²⁷ predicted that the results of NOESY and 1D NOE difference experiments for the lactone favored the 14*R* epimer. Analogous calculations and NOESY

data for arabidiol also strongly favored the 14*R* configuration. Based on these two independent structure analyses the structure of arabidiol was assigned as (13*R*,14*R*,17*E*)-malabarica-17,21-diene-3 β ,14-diol (Figure 6.2). Detailed NMR assignment is provided in Table 1, Appendix A.

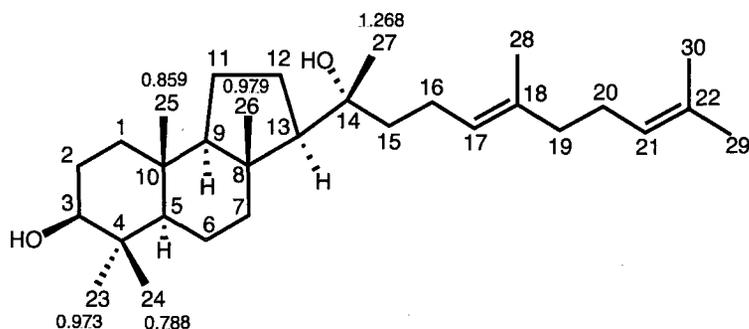


Figure 6.2. Structure of arabidiol with atoms numbered. Underlined are upfield methyl singlets (in ppm).

6.2.2 Analysis of Media for Triterpene Products

Analysis of the resin extraction of the medium revealed large amounts of arabidiol (Figure 6.3), suggesting that increase in polarity could promote triterpene emission to the medium.

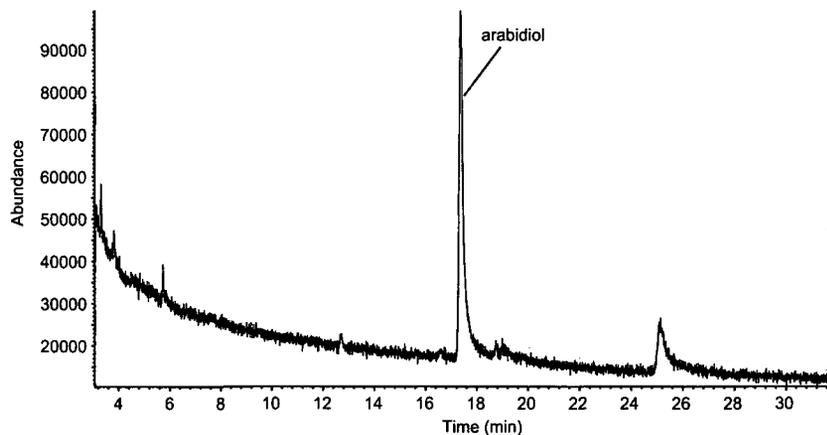


Figure 6.3. Total ion chromatogram of the medium extract.

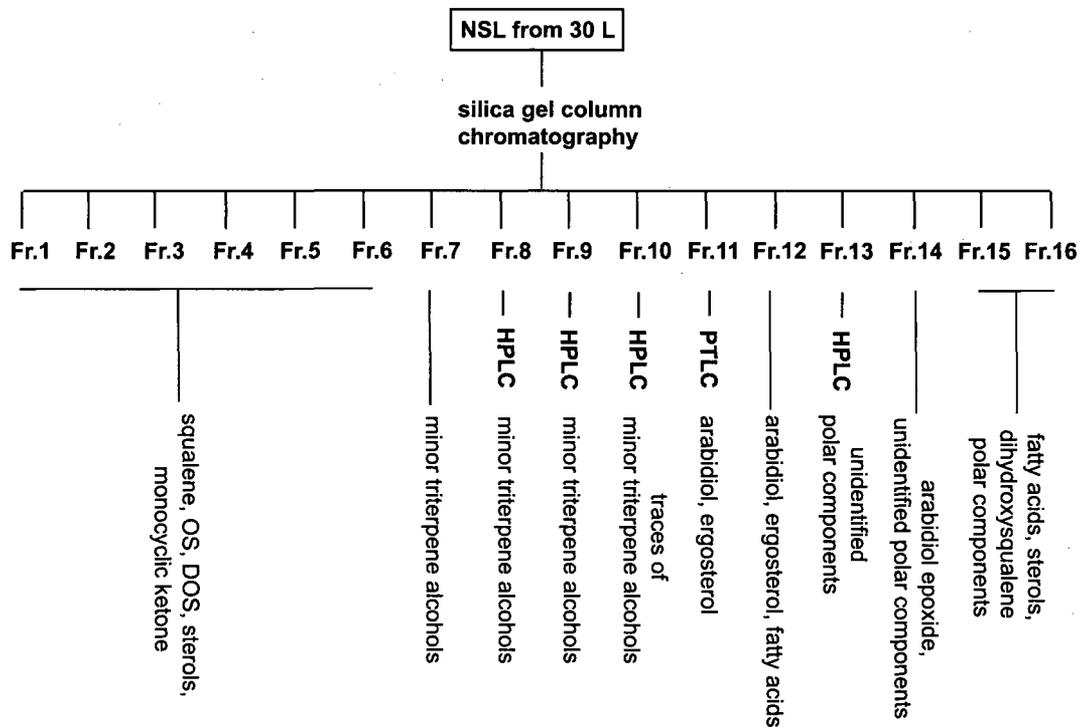
6.2.3 Analysis and Structure Determination of Minor Products

Careful GC-MS and ^1H NMR analyses of crude arabidiol synthase products indicated the presence of many minor triterpene alcohols. For detailed analysis of the minor products, a large-scale (30 L) culture was grown to provide enough material for analysis and structure elucidation. After subjecting the obtained NSL to column chromatography and separating the initial material into 16 fractions (Scheme 6.2), all fractions were analyzed by GC-MS and ^1H NMR (500 MHz) (Figures 11-13, Appendix A), fractions 8, 9 and 10 were further separated by HPLC and rigorously analyzed by ^1H 800 MHz NMR (Figures 14-25, Appendix A).

Fractions 1 to 6 contained squalene, oxidosqualene, dioxidosqualene, monocyclic ketone,¹³⁰ isocamelliol,¹²⁹ and other non-triterpene compounds, as judged by their mass-spectral characteristics and absence of ^1H NMR chemical shifts characteristic for other triterpene alcohols. The presence of the monocyclic ketone and isocamelliol was attributed to non-enzymatic cyclization of oxidosqualene. Sterols were most likely introduced as impurities of ergosterol added to the medium during incubation of SMY8[pMDK3.6]. Fractions 1 – 6 were not analyzed further.

GC-MS and ^1H NMR analysis of fraction 7 showed small amounts of several monocyclic, bicyclic and tricyclic triterpene alcohols, including achilleol A, camelliol C, Δ^7 -polypodatetraenol, and (13*R*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol. All compounds were identified by comparison with authentic standards or spectral data available from the literature. The novel tricycle (13*R*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol was identified from the material obtained from fractions 8 and 9 and is described below.

Scheme 6.2. Scheme of chromatographic separation of NSL extract from a 30-L culture of SMY8[pMDK3.6].



Initial GC-MS and ^1H NMR (Figure 11, Appendix A) analysis of column fraction 8 indicated that several triterpene alcohols were present. Fraction 8 was further subjected to preparative HPLC separation, and the purified components were further analyzed by ^1H 800 MHz NMR and GC-MS (Table 6.1).

Fraction 9 contained the majority of the minor triterpene products (Figure 6.4). This fraction was further purified by HPLC and the combined HPLC fractions were rigorously analyzed by 800 MHz ^1H NMR.

Table 6.1. Triterpene alcohols identified by ^1H NMR analysis of column fraction 8 after HPLC separation. ξ indicates the stereochemistry was not determined.

Fraction number	Compound	Ratio relative to major compound in the fraction
1-74	nothing	
75-80	achilleol A	1
81-87	(13 <i>R</i> ,14 <i>E</i> ,17 <i>E</i>)- malabarica-14,17,21-trien-3-ol	1
	Δ 8(26)-polypodatetraenol	0.1
	camelliol C	0.1
	isocamelliol	0.1
	unidentified peaks	<0.05
88-90	none	
91-92	thalianol	1
	Δ 8-polypodatetraenol	0.12
	14-epithalianol	0.09
	(13 <i>R</i> ,14 <i>E</i> ,17 <i>E</i>)- malabarica-14,17,21-trien-3 β -ol	0.06
	isocamelliol	0.01
	(13 <i>R</i> ,14 ξ ,17 <i>E</i>)-podioda-7,17,21-trien-3 β -ol	0.01
93-98	Δ 7-polypodatetraenol	1
	14-epithalianol	0.09
99-200	sterols or non triterpenes	

The most abundant minor product in both in vivo and in vitro experiments was another tricyclic that had a mass spectrum (Figure 6.5) essentially identical to that of thalianol, with a GC retention time ca. 1 min longer than that of thalianol. 1D (^1H , ^{13}C) and 2D (NOESY, HSQC, and HMBC) NMR experiments revealed a podiodatrienol skeleton for 14-epithalianol but were insufficient for determining the stereochemistry at C-14. Several groups have isolated podiodatrienols or mixtures thereof,^{134,135,130,131,132,133} but assignment of the C-14 configuration of 14-epithalianol was not possible from comparisons with any of literature values. However, previous determination⁷⁰ of the C-14 configuration of thalianol as 14*R* would suggest that by default 14-epithalianol should be the 14*S* epimer of thalianol. For a rigorous proof of structure, we confirmed the C-14 configuration of thalianol using the methodology described above for arabidiol.¹²⁷

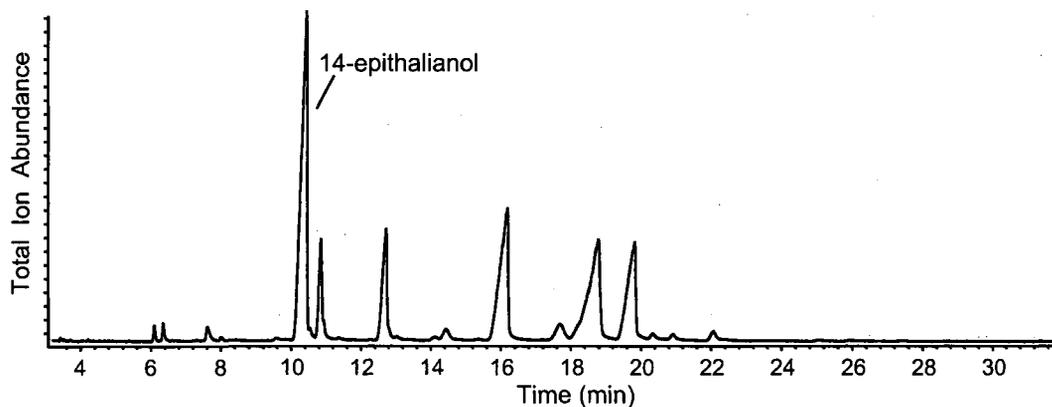


Figure 6.4. Total ion chromatogram from GC-MS of fraction 9 from silica gel chromatography (no derivatization; 260 °C oven temperature), illustrating the multiplicity of triterpene alcohol products from arabidiol synthase. It should be noted that this chromatogram does not represent the relative abundance of the enzymatic products because fraction 9 represents only a portion of the monohydroxy triterpene products.

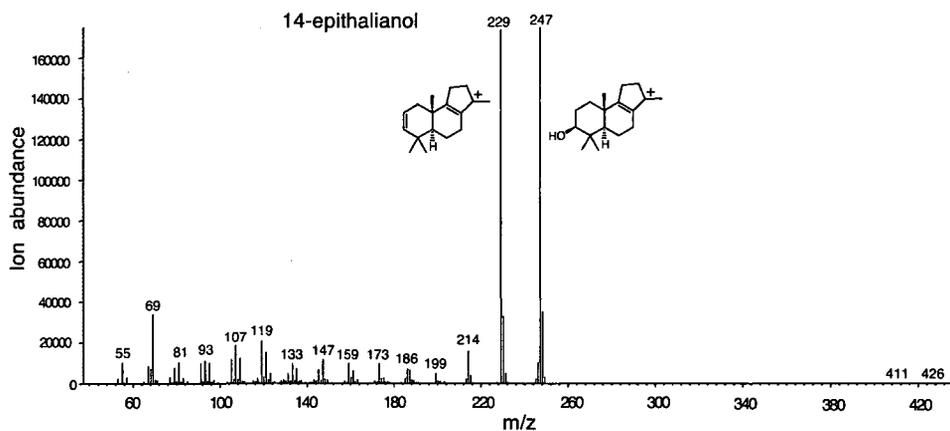


Figure 6.5. Electron impact mass spectrum of underivatized 14-epithalianol.

From HPLC fractions 126-132, the ^1H NMR spectra of the next most abundant compound (1/3 intensity of the major compound 14-epithalianol) had 4 distinct methyl singlets, suggesting a tricyclic triterpene structure. Carbon connectivity was determined from additional ^{13}C , DEPT, HSQC, HMBC, and NOESY data as a 6/6/5 tricycle with unknown C-13 and C14-C15 double bond stereochemistry.

A series of quantum mechanical predictions of stereochemistry for different C-13 structures and 14-15 double bond isomers and comparison to recent literature values for a malabaricatrienol C14-C15 double bond skeleton¹²⁹ allowed for assignment of the stereochemistry for the unknown tricycle to be $13R$ and the C14-C15 double bond to possess a Z configuration, yielding the structure (13*R*,14*Z*,17*E*)-malabarica-14,17,21-trien-3 β -ol.¹²⁹

Another previously unreported tricyclic isomer was characterized from the products of arabidiol synthase. ^1H NMR spectra showed four upfield methyl singlets, corresponding to another 6/6/5 tricyclic triterpene alcohol. Using 2D NMR data for initial structure determination and molecular modeling for establishing stereochemistry at C-14 and C-13, the structure of the tricycle was determined as (13*R*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol

Six tetracycles were identified by comparison of literature values, including euphol, butyrospermol, tirucalla-7,24-dienol, tirucalla-8,24-dienol, dammara-13(17),24-dienol, and bacchara-12,21-dienol. The structure of another tetracycle, dammara-12,24-dienol, was predicted based on available data from the corresponding 3-desoxy analog characterized by NMR. Finally, the structure of a novel tetracyclic triterpene alcohol was

deduced through ^{13}C , DEPT, HMBC, HSQC, and NOE data to be bacchara-13(18),21-dienol.

Pentacycles lupeol, β -amyrin and germanicol were also determined from ^1H NMR. While lupeol and germanicol were present in large amounts, β -amyrin was present only in a trace amount.

Relative ratios of the compounds found in fraction 9 were obtained from ^1H NMR of the fraction prior to HPLC separation (Figures 12-13, Appendix A). In vivo production of tetracycles and pentacycles is significantly lower than that of tricycles, with 20 times less tetra- and pentacycles seen in comparison with tricycles. The ratio of arabidiol to the dominant tetra- and pentacycles was at most 100:0.2. The final composition of fraction 9 is summarized in Table 6.2.

Fraction 10 contained small amounts of (13*S*,17*E*)-malabarica-14(27),17,21-trien-3 β -ol, sterols, and non-triterpene components. Fraction 11 contained the major product arabidiol and ergosterol which were separated by preparative TLC. Part of the purified arabidiol was used for synthesis of the lactone and determination of the arabidiol stereochemistry as described above. Fraction 12 contained small amounts of arabidiol, ergosterol and fatty acids.

GC-MS analysis of fractions 13 and 14 revealed several minor polar compounds with m/z 426 and 444, corresponding to triterpene alcohols and triterpene diols. However, the amount of material available was not sufficient for structure determination and these products were not analyzed in detail. Fraction 14 contained large amounts of arabidiol epoxide, the product of enzymatic cyclization of dioxidosqualene. The structure of arabidiol epoxide with undetermined C-14 stereochemistry was described previously.¹²⁸

Fractions 15 and 16 contained fatty acids, dihydroxysqualene, sterols, and other minor polar components.

Table 6.2. Triterpene alcohols in HPLC-purified column fraction 9, detected by ^1H NMR analysis.

Fraction number	Compound	Ratio relative to major compound in the fraction
88-94	(13 <i>R</i> ,14 <i>E</i> ,17 <i>E</i>)- malabarica-14,17,21-trien-3 β -ol	1
	camelliol C	0.07
	isocamelliol	0.06
	Δ 8(26)-polypodatetraenol	0.04
	unidentified peaks	<0.01
97-98	small unidentified peaks	
101-102	small unidentified peaks	
110-111	Δ 8(26)-polypodatetraenol	1
	isocamelliol	0.02
112-114	(13 <i>R</i> ,14 <i>E</i> ,17 <i>E</i>)- malabarica-14,17,21-trien-3 β -ol	1
	Δ 8(26)-polypodatetraenol	0.4
	isocamelliol	0.2
	unidentified peaks	<0.2
115-117	isocamelliol	1
	unidentified peaks	0.1
118-120	(13 <i>R</i> ,17 <i>E</i>)-malabarica-14(27),17,21-trien-3 β -ol	1
126-132	14-epithalianol	1
	(13 <i>R</i> ,14 <i>Z</i> ,17 <i>E</i>)-malabarica-14,17,21-trien-3 β -ol	0.33
	(13 <i>R</i> ,14 ξ ,17 <i>E</i>)-podioda-7,17,21-trien-3 β -ol	0.16
	dammara-20,24-dienol	<0.03
	unidentified peaks	<0.01
136-137	(13 <i>S</i> ,14 <i>E</i> ,17 <i>E</i>)- malabarica-14,17,21-trien-3 β -ol	1
	(13 <i>R</i> ,14 ξ ,17 <i>E</i>)-podioda-7,17,21-trien-3 -ol	0.01
	14-epithalianol	0.01
	(13 <i>R</i> ,14 <i>Z</i> ,17 <i>E</i>)-malabarica-14,17,21-trien-3 β -ol	<0.01
	(13 <i>S</i> ,17 <i>E</i>)-malabarica-14(27),17,21-trien-3 β -ol	<0.01
138-139	(13 <i>S</i> ,14 <i>E</i> ,17 <i>E</i>)- malabarica-14,17,21-trien-3 β -ol	1
	(13 <i>S</i> ,17 <i>E</i>)-malabarica-14(27),17,21-trien-3 β -ol	0.08
	14-epithalianol	<0.01
	unidentified peaks	<0.01
141-143	(13 <i>S</i> ,17 <i>E</i>)-malabarica-14(27),17,21-trien-3 β -ol	1
	(13 <i>S</i> ,14 <i>E</i> ,17 <i>E</i>)- malabarica-14,17,21-trien-3 β -ol	0.08
	unidentified peaks	<0.01

Fraction Number	Compound	Ratio relative to major compound in the fraction
145-147	(13 <i>S</i> ,17 <i>E</i>)-malabarica-14(27),17,21-trien-3 β -ol	1
	lupeol	0.6
	dammara13(17),24-dienol	0.1
	(13 <i>S</i> ,14 <i>E</i> ,17 <i>E</i>)- malabarica-14,17,21-trien-3 β -ol	0.05
	14-epithalianol	<0.01
148-152	unidentified peaks	<0.01
	bacchara-13(18),21-dienol	1
	(13 <i>S</i> ,17 <i>E</i>)-malabarica-14(27),17,21-trien-3 β -ol	0.07
	lupeol	0.07
	(13 <i>S</i> ,14 <i>E</i> ,17 <i>E</i>)- malabarica-14,17,21-trien-3 -ol	<0.01
162-166	unidentified peaks	<0.07
	bacchara-12,21-dienol	1
	tirucalla-7,24-dienol	0.7
	dammara-12,24-dienol	0.55
	tirucalla-8,24-dienol	0.1
167-171	unidentified peaks	<0.01
	euphol	1
	tirucalla-7,24-dienol	0.4
	germanicol	0.04
	butyrospermol	0.04
172-175	unidentified peaks	<0.04
	butyrospermol	1
	β -amyryn	0.0125
	germanicol	<0.01
	euphol	<0.01
All other fractions combined and all fractions after 180	unidentified peaks	<0.01
	(13 <i>R</i> ,17 <i>E</i>)-malabarica-14(27),17,21-trien-3 β -ol	1
	(13 <i>R</i> ,14 ξ ,17 <i>E</i>)-podioda-7,17,21-trien-3 β -ol	0.5
	14-epithalianol,	<0.5 other
	(13 <i>S</i> ,17 <i>E</i>)-malabarica-14(27),17,21-trien-3 β -ol, Δ 7-polypodatrienol, thalianol, achilleol A, camelliol C, isocamelliol, unidentified peaks	triterpenes

Overall, the large-scale culture of arabidiol synthase showed the presence of two monocyclic triterpenes, achilleol A and camelliol C, and three bicyclic triterpenes, Δ 7-polypodatetraenol, Δ 8-polypodatetraenol, and Δ 8(26)-polypodatetraenol. Nine tricyclic triterpenes were identified as arabidiol, 14-epithalianol, thalianol, (13*R*,14*Z*,17*E*)-malabarica-14,17,21-trien-3 β -ol, (13*R*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol,

(13*S*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol, (13*R*,17*E*)-malabarica-14(27),17,21-trien-3 β -ol, (13*S*,17*E*)-malabarica-14(27),17,21-trien-3 β -ol, and (13*R*,14 ξ ,17*E*)-podioda-7,17,21-trien-3 β -ol. Further analysis uncovered eight tetracycles: euphol, butyrospermol, tirucalla-7,24-dienol, tirucalla-8,24-dienol, dammara-12,24-dienol, dammara-13(17),24-dienol, bacchara-13(18),21-dienol, and bacchara-12,21-dienol. Finally, three pentacycles were identified as lupeol, β -amyrin, and germanicol (Figure 6.6).

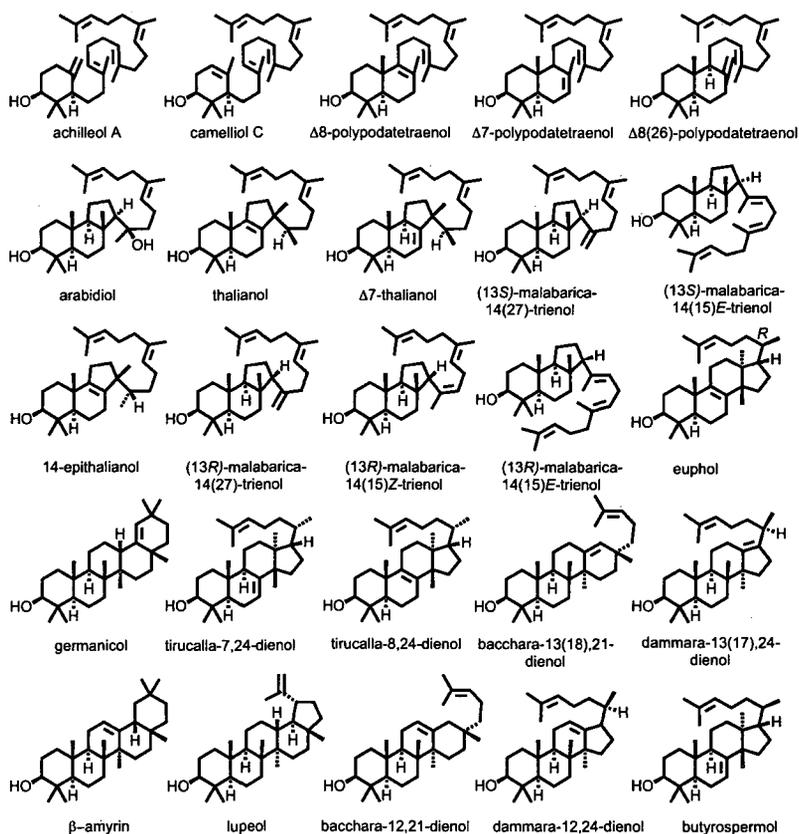


Figure 6.6. Triterpenes identified in a 30-L extract from in vivo expression of arabidiol synthase.

We do not preclude the possibility that other triterpenes could be formed as by-products of arabidiol synthase, but these would be found at much lower levels (<0.01% of the

total). That some of the mono-, bi-, and tricycles are products of non-enzymatic cyclization of oxidosqualene is also a possibility. Monocyclic ketone and isocamelliol are the most abundant products of non enzymatic cyclization and therefore we suggest that those are artifacts. Non-enzymatic formation of tetra- and pentacycles was not previously observed.¹²⁹

6.2.4 Spectral Analysis of Arabidiol Synthase Minor Compounds

Identification of minor compounds was done by comparison of obtained ¹H NMR spectra with spectra available from the literature or comparison with available ¹H NMR spectra of authentic standards. Usually, upfield methyl peaks and olefinic signals were used for identification of the compound.

Two monocycles including monocyclic ketone¹³⁴ and isocamelliol,¹²⁹ were identified by upfield signals at 0.578 ppm (s), 0.900 ppm (d) and 0.931 ppm (d) for monocyclic ketone, and by two methyl singlets at 1.012 ppm and 1.074 ppm for isocamelliol.

The monocyclic triterpene achilleol A¹⁰¹ was identified by two characteristic methyl singlets at 0.715 ppm (H-23, s) and 1.032 ppm (H-24, s), and two olefinic signals at 4.607 ppm and 4.874 ppm. Distinctive NMR signals, including singlets at 0.831 ppm (H-24) and 0.969 ppm (H-23) and an olefinic peak at 5.237 ppm, allowed for camelliol C¹⁰² identification.

Bicyclic triterpenes were identified by comparison of NMR values with those reported previously. Each of the polypodatetraenols was identified by three distinctive methyl singlets for H-23, H-24, and H-25 and olefinic peaks. Δ 7-polypodatetraenol¹³⁵

(0.971 ppm, 0.851 ppm, 0.744 ppm, and 5.390 ppm), Δ 8-polypodatetraenol⁶⁹ (1.007 ppm, 0.804 ppm, and 0.946 ppm) and Δ 8(26)-polypodatetraenol¹³⁶ (0.992 ppm, 0.770 ppm, 0.672 ppm, 4.557 ppm, and 4.846 ppm) were thus identified.

Previously reported tricyclic triterpenes were identified from ¹H NMR spectra, including distinctive upfield signals for thalianol⁷⁰ (0.820 ppm (s), 0.836 ppm (d), 0.934 ppm (s), 0.947 ppm (s), 1.017 ppm (s), and two olefinic signals at 5.097 ppm and 5.112 ppm). The ¹H NMR spectrum of (13*S*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol¹²⁹ indicated the presence of characteristic singlets at 0.621 ppm, 0.792 ppm, 0.839 ppm, 0.979 ppm, and a broad triplet at 5.118 ppm. The spectrum of (13*R*,17*E*)-malabarica-14(27),17,21-trien-3 β -ol¹³⁷ showed singlets at 0.781 ppm, 0.855 ppm, 0.959 ppm, 0.979 ppm and two olefinic signals at 4.589 ppm and 4.877 ppm. The 13*S* isomer (13*S*,17*E*)-malabarica-14(27),17,21-trien-3 β -ol¹³³ was also identified by upfield singlets at 0.661 ppm, 0.795 ppm, 0.852 ppm, 0.980 ppm, and two olefinic signals at 4.742 ppm and 4.904 ppm. The ¹H NMR spectra of (13*R*,14 ξ ,17*E*)-podioda-7,17,21-trien-3 β -ol^{138,139} with undetermined C14 stereochemistry showed presence of previously reported signals at 0.707 ppm (s), 0.808 ppm (d), 0.876 ppm, 0.927 ppm, 0.997 ppm and 5.220 ppm (dt).

The tricycle 14-epithalianol was characterized through 1D (¹H, ¹³C) and 2D (NOESY, HSQC, and HMBC) NMR experiments. The C-14 stereocenter was confirmed using the methodology described above for arabidiol (see section 6.1.3) and explained in detail in Kolesnikova et al.¹²⁷

¹H NMR spectra of a novel tricycle (13*R*,14*Z*,17*E*)-malabarica-14,17,21-trien-3 β -ol (from column fraction 9, HPLC fractions 126-132) showed 4 methyl singlets (0.776

ppm, 0.849 ppm, 0.966 ppm, and 0.978 ppm) in the upfield region. Carbon connectivity was determined from additional ^{13}C , DEPT, HSQC, HMBC, and NOESY data.

The ^1H NMR spectrum of another novel tricycle showed four upfield methyl singlets (0.781 ppm, 0.843 ppm, 0.955 ppm, and 0.957 ppm). 2D NMR data established the structure as (13*R*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol, and molecular modeling established the stereochemistry at C-14 and C-13.

Euphol⁹ (0.754 ppm (s), 0.801 ppm (s), 0.855 ppm (d), 0.875 ppm (s), 0.951 ppm (s), 1.003 ppm (s), and 5.093 ppm), butyrospermol⁹ (0.744 ppm (s), 0.805 ppm (s), 0.849 ppm (d), 0.860 ppm (s), 0.970 ppm (s), 0.974 ppm (s), 5.101 ppm, and 5.255 ppm), tirucalla-7,24-dienol^{9,140,141} (0.747 ppm (s), 0.809 ppm (s), 0.861 ppm (s), 0.882 ppm (d), 0.968 ppm (s), 0.970 ppm (s), 5.099 ppm, and 5.256 ppm), tirucalla-8,24-dienol⁹ (0.756 ppm (s), 0.801 ppm (s), 0.867 ppm (s), 0.917 ppm (d), 0.953 ppm, 1.003 ppm, and 5.101 ppm), and bacchara-12,21-dienol¹⁴² (0.748 ppm (s), 0.792 ppm (s), 0.968 ppm (s), 0.994 ppm (s), and 1.001 ppm (s)) were also identified.

The structure of the novel tetracyclic triterpene bacchara-13(18),21-dienol was determined from 1D and 2D NMR spectra of the HPLC fractions 148-152. The six upfield methyl singlets at 0.770 ppm, 0.881 ppm, 0.918 ppm, 0.939 ppm, 0.991 ppm, and 1.098 ppm, along with further assignments from ^{13}C , DEPT, HMBC, HSQC and NOE data allowed for determination of its structure to be bacchara-13(18),21-dienol.

The tetracycles dammara-12,24-dienol and dammara-13(17),24-dienol were identified by comparison of the ^1H NMR with data available for their 3-desoxy analogs.^{143,144} Comparison of our spectral data for dammara-12,24-dienol (0.784 ppm (d, $J=7.0$), 0.798 ppm (s), 0.931 (d, $J=7.0$ Hz), 0.937 ppm (s), 0.953 ppm (s), 0.998 ppm (s),

0.999 ppm (s), and olefinic peak at 5.255 ppm), and dammara-13(17),24-dienol (0.770 ppm (s), 0.815 ppm (s), 0.846 ppm (s), 0.960 ppm (d), 0.987 ppm (s), 1.082 ppm (s), and 5.077 ppm) with the signals for the 3-deoxy analogs predicted the structure of the tetracycles as dammara-12,24-dienol and dammara-13(17),24-dienol correspondingly.

Lupeol⁹ was identified in ¹H NMR spectra by distinctive signals (methyl singlets at 0.761 ppm, 0.788 ppm, 0.831 ppm, 0.944 ppm, 0.968 ppm, and 1.030 ppm, and olefinic signals at 4.566 ppm and 4.687), germanicol⁹ was identified by (methyl singlets at 0.736 ppm, 0.769 ppm, 0.879 ppm, 0.937 ppm, 0.940 ppm, 0.972 ppm, 1.019 ppm, and 1.078 ppm, and olefinic signal at 4.855 ppm), and β-amyrin⁹ by (methyl singlets at 0.792 ppm, 0.833 ppm, 0.870 ppm, 0.872 ppm, 0.939 ppm, 0.969 ppm, 0.998 ppm, and 1.136 ppm).

6.2.5 Quantitative Characterization of Arabidiol Synthase from In Vitro Analysis

Analysis of a 2-L in vitro extract allowed for clarification of the relative amounts of triterpene products in the arabidiol synthase product profile. Analysis of the crude extract established the ratio of arabidiol to the dominant minor product 14-epithalianol as roughly 100:4. Estimated ratios of the other minor products were calculated from the preparative TLC fraction containing only minor products (Figures 26-27, Appendix A). Ratios of 14-epithalianol, (13*S*,17*E*)-malabarica-14(27),17,21-trien-3β-ol, (13*S*,14*Z*,17*E*)-malabarica-14,17,21-trien-3β-ol, (13*S*,14*E*,17*E*)-malabarica-14,17,21-trien-3β-ol, (13*R*,14*ξ*,17*E*)-podioda-7,17,21-trien-3β-ol, and Δ⁷-polypodatetraenol were determined from ¹H NMR as 4:4:3.6:1.6:0.4:0.4. Two tricyclic triterpenes, (13*R*,17*E*)-

malabarica-14(27),17,21-trien-3 β -ol and (13*R*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol, did not have well-separated signals for quantitation, however together they had a ratio to 14-epithalianol of 2.8:4. Bicyclic triterpenes Δ 8-polypodatetraenol and Δ 8(26)-polypodatetraenol were present at levels lower than 0.2% compared to arabidiol.

Monocyclic triterpenes were not distinctively identified in the mixture of the triterpenes because they have only 2 upfield methyl peaks each and were covered by other peaks, however a peak at 1.032 ppm suggests the presence of achilleol A in the mixture at low levels (less than 0.1%). Among the tetra- and pentacycles identified from the large-scale in vivo experiment, only signals for lupeol and bacchara-13(18),21-dienol that were present in dominant amounts compared to other penta- and most tetracycles in vivo, were identified from in vitro extracts, and only at very low levels (less than 0.1% of arabidiol). However, because other tetracyclic triterpenes were present in vivo at amounts significantly lower than that for tricycles, and because formation of tetra- and pentacycles was not previously observed as a result of non-enzymatic oxidosqualene cyclization, we conclude that penta- and tetracyclic triterpenes are produced at levels significantly lower than 0.1% of the total. Therefore estimating total product ratios from available ^1H NMR data, arabidiol synthase makes 85.4% arabidiol, 3.4% each 14-epithalianol and (13*S*,17*E*)-malabarica-14(27),17,21-trien-3 β -ol, 3% (13*S*,14*Z*,17*E*)-malabarica-14,17,21-trien-3 β -ol, 1.4% (13*S*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol, 2.4% between both (13*R*,17*E*)-malabarica-14(27),17,21-trien-3 β -ol and (13*R*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol, 0.3% each (13*R*,14 ξ ,17*E*)-podioda-7,17,21-trien-3 β -ol and Δ 7-polypodatetraenol, and less than 0.2% each Δ 8-polypodatetraenol and Δ 8(26)-

polypodatetraenol. All other triterpenes were present at levels less than 0.1% of the total product yield.

6.3 Discussion

Expression of *At4g15340* (PEN1) in yeast resulted in isolation of the major product arabidiol and numerous minor triterpene products, and therefore PEN1 was named arabidiol synthase.

Arabidiol synthase belongs to the PEN clade of *Arabidopsis* oxidosqualene cyclases. This enzyme is 72-82% identical to other cyclases from the PEN clade, including 82% to baruol synthase (BARS1, PEN2), 81% to thalianol synthase (THAS1, PEN4), marneral synthase (MRN1, PEN5), seco-amyrin synthase (PEN6), and an uncharacterized enzyme (PEN3). It also shows 78% homology with AtCAS1. More distantly (59-61%), PEN1 is related to several β -amyrin synthases, including those from *Panax ginseng*,¹⁰⁸ *Euphorbia tirucalli*,¹⁴⁵ and *Betula platyphylla*.¹⁰⁴

Although PEN1 has more amino acid residues in common with baruol synthase, its phylogenetic origin shows that it is closer to thalianol synthase (Figure 1.3). In fact, analysis of their major products arabidiol and thalianol shows that these most likely arrived from the same cation A (Scheme 6.3), and therefore active site organization of PEN1 and thalianol synthase are more similar than PEN1 and baruol synthase.¹²⁹

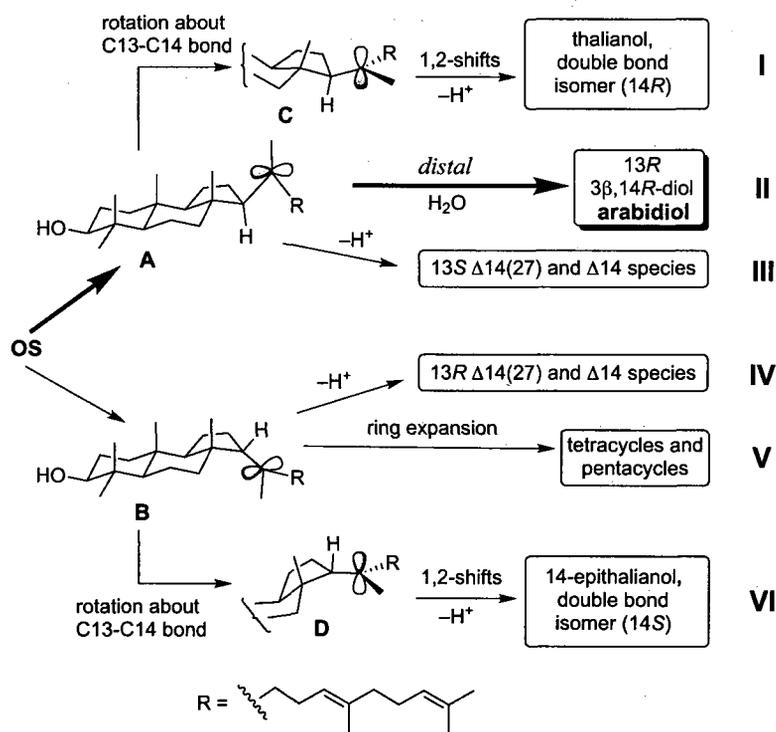
Arabidiol synthase produces 85.4% arabidiol and 14.6% minor compounds. These minor products include three bicycles and eight more tricycles. Additionally, small amounts of monocycles (achilleol A), eight tetracycles and six pentacyclic triterpenes were identified. From these, five novel structures were characterized through extensive

GC-MS and NMR studies, bringing the total number of triterpenes isolated from *Arabidopsis* triterpene synthases up to over 40.

Through molecular modeling experiments and degradation of the diol to the corresponding lactone, we were able to deduce the stereochemistry of arabidiol and 14-epithalianol. Furthermore, analysis of the media revealed that arabidiol was being released into the medium, underlining the significance of in vitro experiments for establishing product ratios. Differences in product polarities could cause significant underestimation of polar products compared to less polar triterpenes.

How arabidiol synthase biosynthesizes arabidiol and these numerous minor products is illustrated in Scheme 6.3. The typical tricyclic precursor of tetracyclic and pentacyclic triterpenes is the 13*R* malabaricadienyl cation **B**. However, its 13*S* epimer **A** is clearly the precursor of arabidiol and thalianol. Rotations of **A** about the C13-C14 bond to form cation **B** or **D** are unlikely because of restricted volume in the cyclase active site cavity. To form arabidiol (pathway **II**), a replenishable ordered water in the active site might attack the C-14 cation of **A** on its proximal side (facing the cyclized core) or its distal side (facing the entrance channel of the active site) or it could attack cation **C** from above or below. Molecular modeling calculations with addition of dummy oxygen atoms to a B3LYP/6-31G* model¹⁴⁶ of cation **A** were prepared to investigate these possibilities. The obtained results suggested that the only place where water would preferentially attack **A** is on the distal face of the carbocation. All other positions of water prior the attack would result in deprotonation (such as pathway **III**) or would be impossible because of steric conflict (Figure 6.7).

Scheme 6.3. Mechanistic pathways of product formation in arabidiol synthase.



These thoughts lead us to the conclusion that the C-14 configuration could have been deduced a priori. Similar thoughts and molecular modeling of horizontal conformers of the isomalabaricadienyl, epidammarenyl, lupanyl, and hopanyl cations shows a stronger bias for water addition on the distal face of the cation. This reasoning allowed to assign stereochemistry for several other triterpenes previously reported with ambiguous stereochemistry (Figure 6.8).^{81,83,147,148} and suggested a unifying rule that water attacks only the distal face of triterpene cationic intermediates.

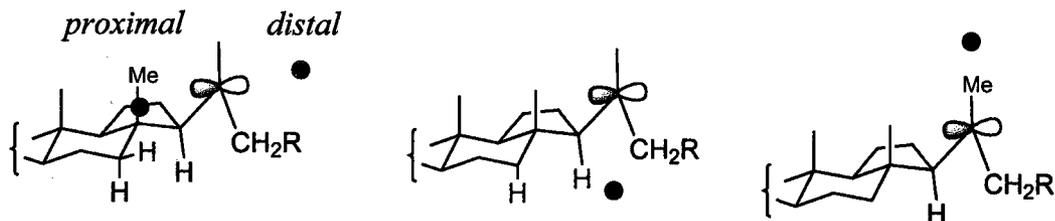


Figure 6.7. Predicted steric conflict of water addition during formation of arabidiol. Black dots show positions of dummy oxygen atoms that would result in deprotonation from positions marked in red. The blue dot corresponds to a position of the dummy oxygen atom that would successfully quench the carbocation and result in the formation of arabidiol.

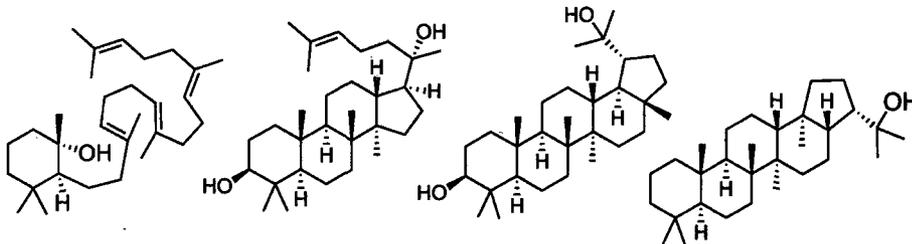


Figure 6.8. Triterpene products formed by water addition.

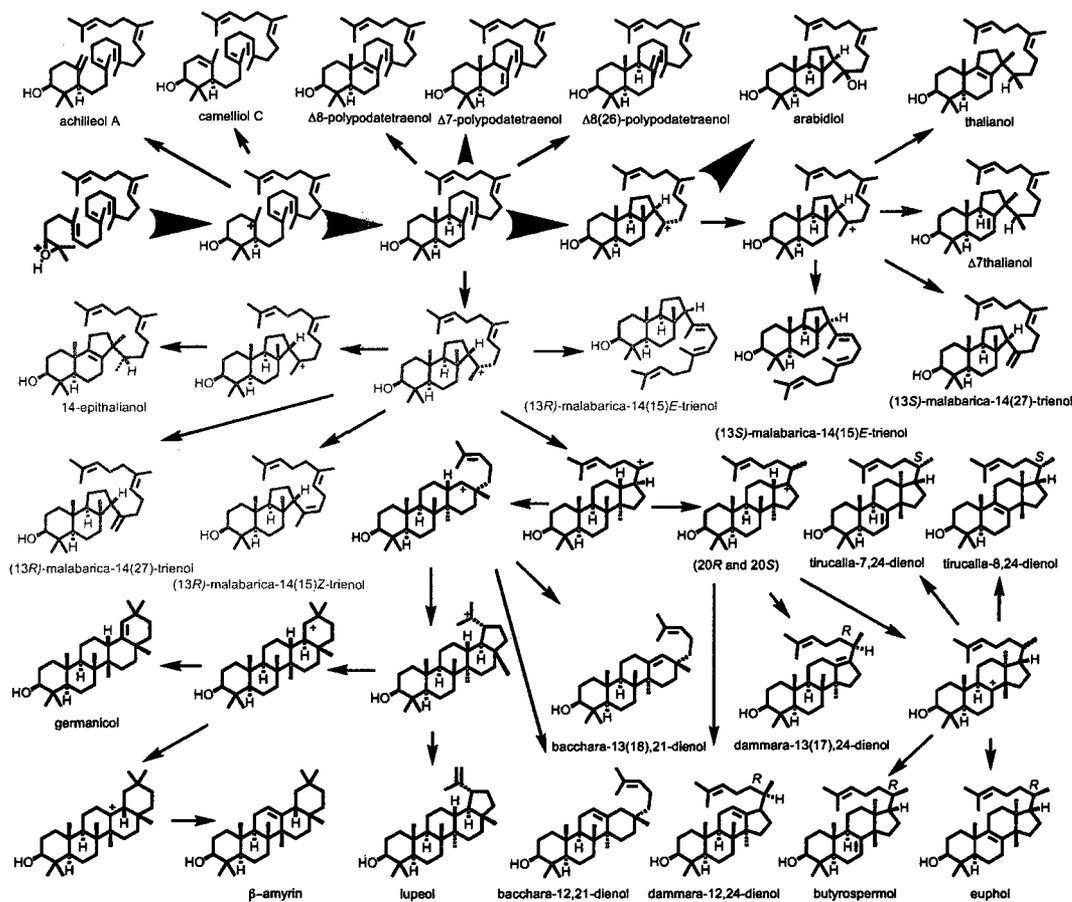
Water addition in triterpene biosynthesis is uncommon when compared to biosynthesis of triterpenes obtained by deprotonation. There are just two other reported examples of native oxidosqualene cyclases terminating triterpene biosynthesis by water addition, *Arabidopsis* lupeol synthase (LUP1)^{74,75,80,81} and dammarendiol-II synthase from *Panax ginseng*.⁸¹ Hydroxylation is more difficult for cyclases to carry out and is favored over deprotonation only if the ordered water is efficiently replenished for each new substrate, and if the trajectory of attack closely follows the axis of the empty 2p orbital. This insight aligns with observations that cyclases usually terminate the cationic

cascade by deprotonation and that cyclases that do perform hydroxylation also generate olefinic byproducts as shown in this work.

Occasional misplacement or absence of replenishable water in the enzyme causes formation of cation **B** and could lead to the formation of the minor products via pathways **IV**, **V**, or **VI**. A detailed scheme of the arabidiol synthase product formation is illustrated in Scheme 6.4. In arabidiol synthase, there are at least 24 minor products generated from this deprotonation side reaction. Tetracycles and pentacycles are likely derived from deprotonation of the 13*R* malabaricadienyl cation **B** (pathway **V**), while the tricycles are produced by further ring expansion, rearrangements and deprotonation.

The ability of oxidosqualene cyclases to make numerous products and the reasoning for such biosynthetic diversity in the product profiles of cyclases has been discussed previously.¹²⁹ Assuming that arabidiol synthase produces all tetra- and pentacyclic compounds at lower levels (less than 0.1% of the total), the total number of positions from which deprotonation can occur is close to that seen in baruol synthase. Therefore the arabidiol synthase is another example of a cyclase that relatively efficiently biosynthesizes a dominant product (85.4% of the total triterpenes). However, because of low selective pressure it does not improve its accuracy and allows for minor products to be formed. These results support previous suggestions that biosynthetic diversity is a default for oxidosqualene cyclases.¹²⁹

Scheme 6.4. Proposed mechanism of triterpene formation by arabidiol synthase.



Arabidiol synthase was previously reported as an accurate enzyme that makes a single product arabidiol.¹²⁸ However the present work demonstrated that the arabidiol synthase in fact makes numerous minor products and is a first example of the oxidosqualene cyclase that can make a spectrum of triterpenes, from mono- to pentacycles. This enzyme generates more triterpene products than any other triterpene synthase to date, dramatically increasing the metabolic profile of this small plant.

CHAPTER 7

Conclusions

Genome mining is a powerful approach to uncover new enzymes with unknown functions. Comparison of closely related enzymes with similar active site organization while tracking small changes within the active site, including changes in polarity or steric bulk of the residues, and correlating it with product outcome allows elucidation of catalytic functions of enzymatic residues and prediction of mechanisms of enzymatic catalysis. Sequence comparison and phylogenetic analysis further supports elucidating how enzymes evolved and what mechanisms were developed to introduce novel functionality and to achieve product specificity.

The genome of the model plant *Arabidopsis thaliana* encodes thirteen oxidosqualene cyclases. Initial expectations of triterpene production in the plant were more modest than the obtained results, in terms of diversity of characterized triterpene structures and variety of enzymatic functions. To date eleven *Arabidopsis* oxidosqualene cyclases have been characterized, and eight of these (MRN1, THAS1, CAMS1, PEN1, LSS1, BARS1, LUP5, PEN6) have catalytic activity that previously were not typically found in *Arabidopsis*.¹¹⁶(and references therein) This work has reported characterization of three of these oxidosqualene cyclases.

Discovery of *Arabidopsis* lanosterol synthase, the first example of a lanosterol synthase found in plants, has demonstrated the biosynthetic ability of plants to make the alternative sterol precursor lanosterol that is presumably used in plants as a secondary metabolite, in contrast with animals and fungi which utilize lanosterol for primary

metabolism. It was also shown that lanosterol synthases in plants evolved independently from lanosterol synthases in animals and fungi and phylogenetic analysis suggests evolution from plant cycloartenol synthases. This enzyme also utilizes a different set of active site residues to biosynthesize lanosterol than previously seen in animals and fungi. Previous mutagenesis experiments converted an accurate cycloartenol synthase into an accurate lanosterol synthase, showing an evolutionary path to altering enzymatic activity by just changing two AtCAS1 active site residues, His477 and Ile481. These changes in polarity and steric bulk allowed for the deprotonating base to shift within the active site. The reverse mutations in *Arabidopsis* lanosterol synthase to change polarity and steric bulk in the active site to those in cycloartenol synthase resulted in a partial success, allowing the deprotonating base to shift back to C-19. However, this work also demonstrated that during evolution plant lanosterol synthases acquired additional mutations that would need to be altered to improve accuracy in cycloartenol synthase formation. Nevertheless, the observed production of cycloartenol (31%) in a lanosterol synthase background is the highest known level for introducing the ability to form the cyclopropyl ring through protein engineering.

The characterization of *Arabidopsis* camelliol C synthase has demonstrated that the physiological role and distribution of monocycles might have been previously underestimated. It also showed that evolution of oxidosqualene cyclases does not necessarily go through continuous iterative addition of motifs that change the product outcome from smaller ring systems to larger ones, but can be done spontaneously. Identification of a potential active site residue that might be responsible for this change predicts that a decrease in steric bulk above ring A could promote monocyclization.

Heterologous expression of arabidiol synthase in yeast resulted in isolation and characterization of the novel tricyclic triterpene diol arabidiol. Analysis of the arabidiol structure and determination of arabidiol stereochemistry has allowed development of a general rule for water addition in triterpene biosynthesis in which water adds from the distal side of the carbocation for the majority of carbocations, with the notable exception of tetracycles. Mechanistic analysis of product formation in arabidiol synthase explains why deprotonation dominates over water addition in triterpene biosynthesis and why cyclases that produce diols also exhibit olefinic product formation. Analysis of the arabidiol synthase minor products provided another example of a cyclase previously reported as an accurate enzyme but in fact produces numerous minor products. It also supports the suggestion that inaccuracy in triterpene biosynthesis is a default for cyclases. Arabidiol synthase is also the first example of a cyclase that makes all variety of ring systems, from mono- to pentacycles.

Functional characterization of these three oxidosqualene cyclases has made a significant contribution to illuminating triterpene biosynthesis in *Arabidopsis*. This work demonstrates the utility of the genome mining approach in conjunction with heterologous expression and uncovered three enzymes with novel functions.

Some of the questions yet to be answered are what is the complete triterpene product profile of *Arabidopsis*, and what is the function of these compounds in plants. The complete characterization of *Arabidopsis* oxidosqualene cyclases will be finished soon and would address the first question. Application of bioinformatics tools, consideration of already available expression data, and overexpression/knockout experiment with plants could help with establishing biological functions of identified

triterpenes. Insights gained in this way could be extended to analysis of oxidosqualene cyclases from other sources and potentially lead to the discovery of novel natural products.

Appendix A

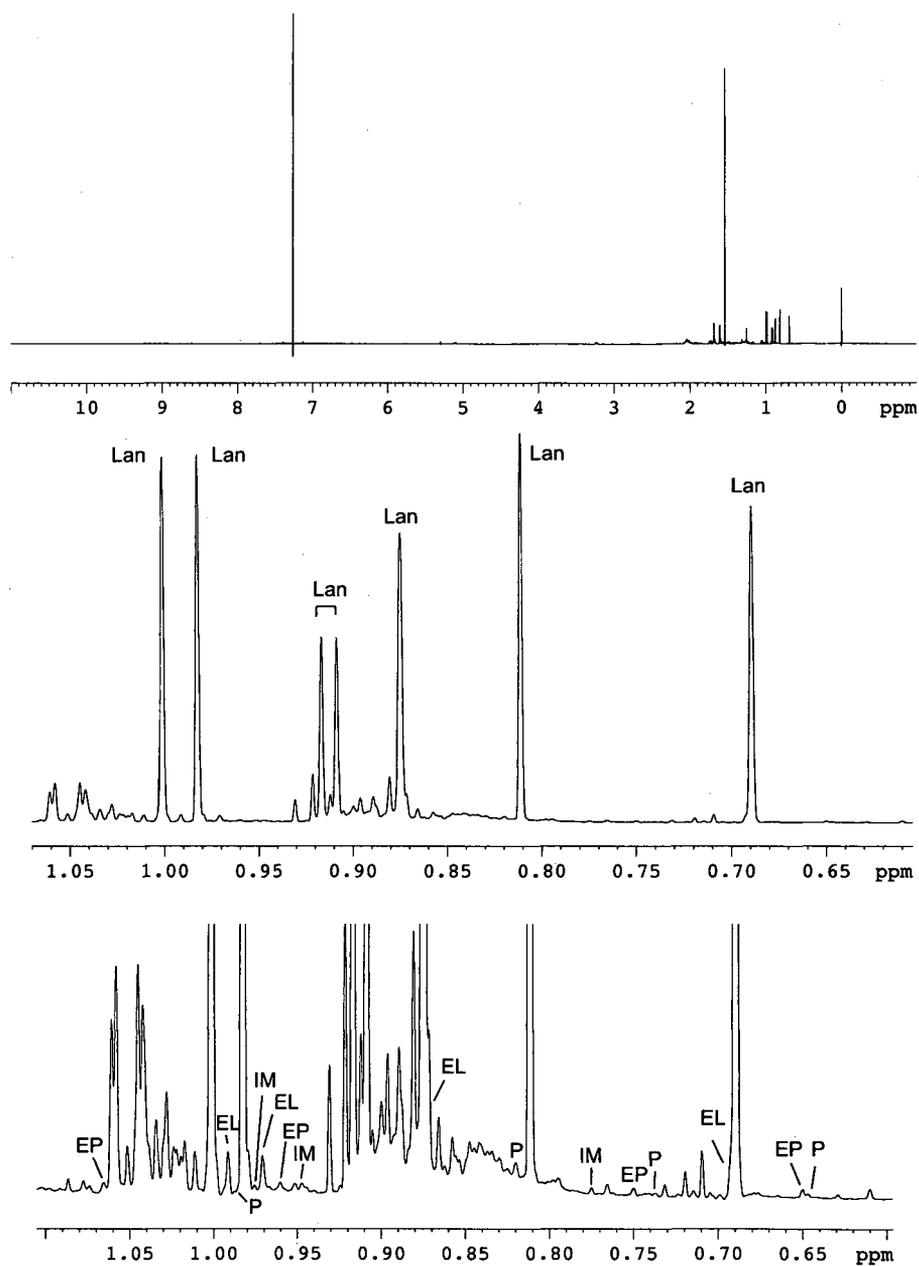


Figure 1. 800 MHz ^1H NMR analysis of the triterpene fraction of RXY6[pLH1.25]. Peaks denoted Lan correspond to lanosterol, EL corresponds to epilanoesterol, P to parkeol, EP epiparkeol, and IM to isomalabaricatrienol.

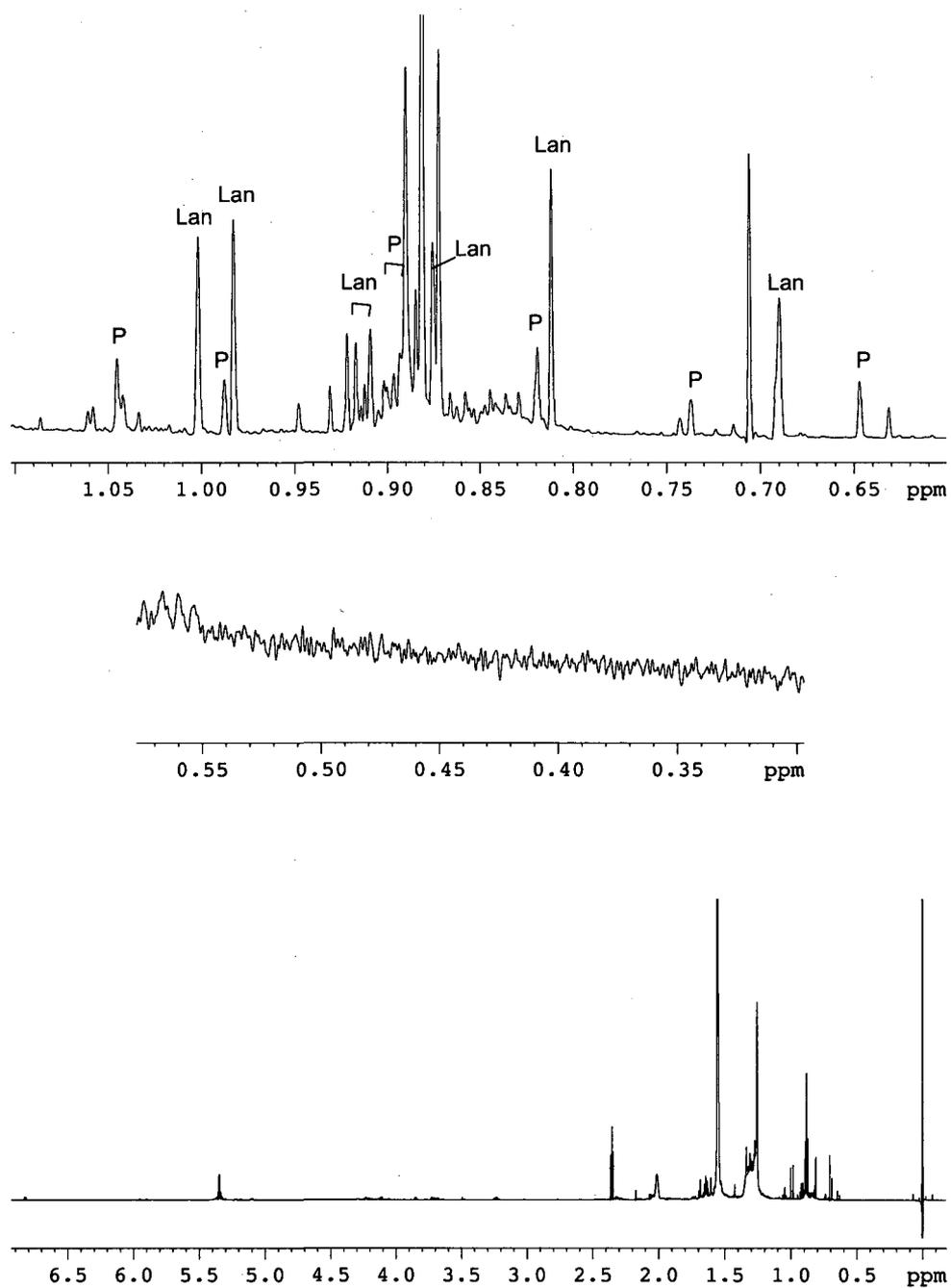


Figure 2. 800 MHz ^1H NMR analysis of the triterpenes produced by AtLSS1 Val481Ile mutant RXY6[pDAL7.0]. Peaks denoted Lan correspond to lanosterol, P to parkeol.

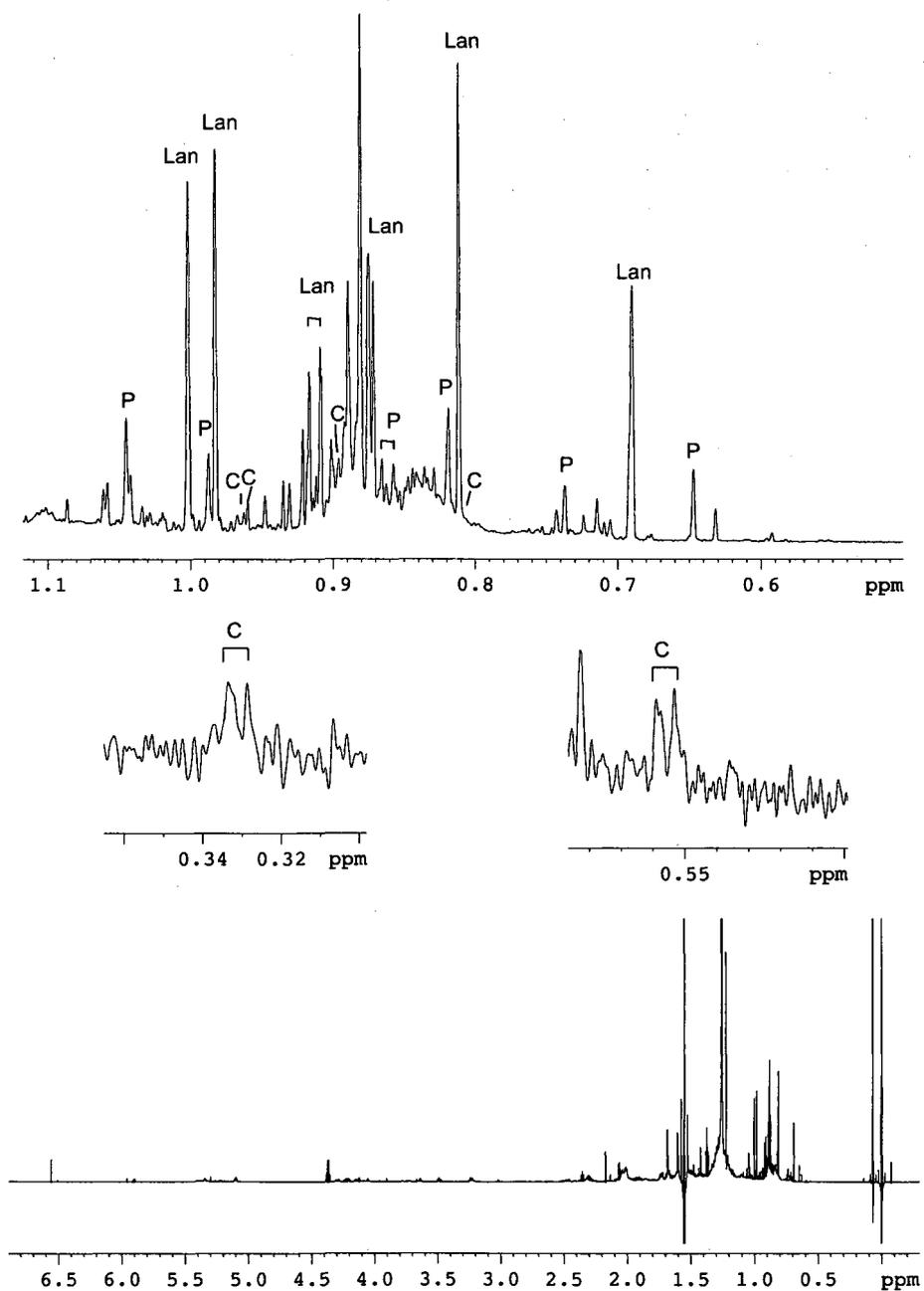


Figure 3. 800 MHz ¹H NMR analysis of the triterpenes produced by AtLSS1 Asn477His mutant RXY6[pDAL6.0]. Peaks denoted Lan correspond to lanosterol, P to parkeol, and C to cycloartenol.

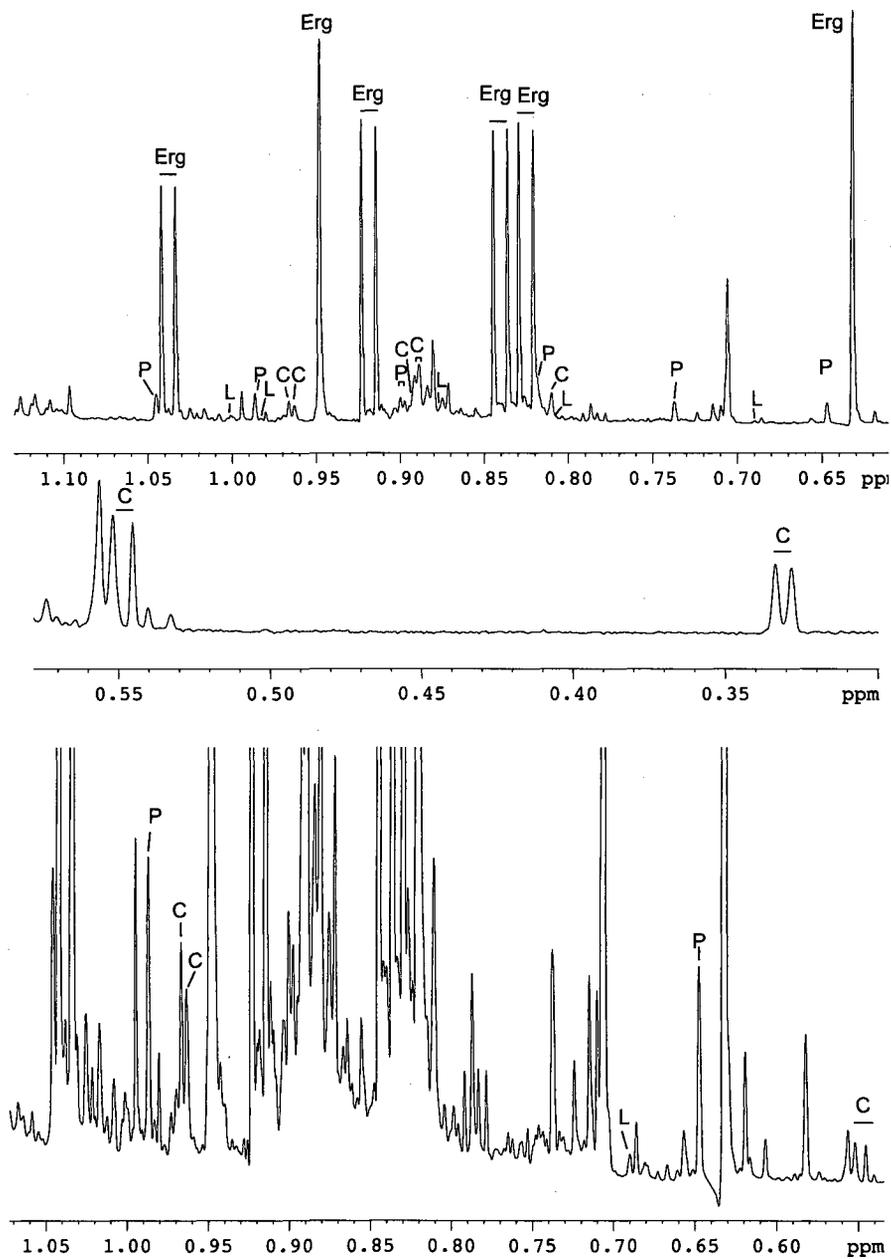


Figure 4. 800 MHz ^1H NMR analysis of the triterpenes produced by AtLSS1 Asn477His/Val481Ile double mutant RXY6[pMDK14.3]. Peaks denoted Lan correspond to lanosterol, P to parkeol, and C to cycloartenol.

Table 1. ^{13}C and ^1H NMR assignments for arabidiol

Carbon	δ_{C}	Hydrogen	δ_{H}	Scalar ^1H - ^1H couplings (Hz)
1	38.67	1 α	1.016	td, 13.1, 4.4
		1 β	1.493	dt, 13.3, 3.5
2	27.14	2 α	1.628	dddd, 13.2, 5.3, 4.4, 3.3
		2 β	1.596	m
3	79.25	3 α	3.200	dd, 11.0, 5.4
4	38.68			
5	56.01	5 α	0.767	dd, 12.2, 2.6
6	19.05	6 α	1.559	m
		6 β	1.488	dddd, 14.0, 12.6, 12.2, 3.5
7	42.44	7 α	1.205	td, 12.7, 4.3
		7 β	2.088	dt, 12.8, 3.4
8	43.89			
9	63.13	9 α	1.067	dd, 13.2, 7.1
10	36.72			
11	19.24	11 α	1.473	dddd, 12.2, 9.7, 7.2, 3.4
		11 β	1.361	br qd, 13, 6.6
12	21.94	12 α	1.678	dtd, 13.3, 9.5, 6.6
		12 β	1.755	dddd, 13.4, 11.4, 10.4, 3.4
13	58.46	13 α	1.381	dd, 10.4, 9.5
14	75.01			
15	43.84	15R	1.367	m
		15S	1.433	m
16	22.85	16R	1.97	m
17	*124.28	17	5.095	t of sextet, 7.0, 1.3
18	135.17			
19	39.68	19	1.97	m
20	26.67	20	2.06	m
21	*124.34	21	5.083	t of septet, 7.0, 1.4
22	131.39			
4 α -Me	28.06	4 α -Me	0.973	s
4 β -Me	15.17	4 β -Me	0.788	s
10-Me	15.78	10-Me	0.859	d, 0.8
8-Me	15.72	8-Me	0.979	d, 0.7
14-Me	25.78	14-Me	1.268	s
18-Me	16.00	18-Me	1.607	dt, 1.5, 0.7
22-Me (E)	25.69	22-Me (E)	1.679	qd, 1.3, 0.3
22-Me (Z)	17.69	22-Me (Z)	1.599	br ddt, 1.4, 0.9, 0.7

^a Data were acquired on 500 or 800 MHz instruments in CDCl_3 solution (<10 mM) at 25 °C. Chemical shifts were corrected for effects of strong coupling. Under the specified conditions of temperature and concentration, reproducibility is ca. ± 0.001 ppm (^1H shieldings given to 3 decimal places), ± 0.03 ppm (^{13}C shieldings), or ± 0.3 Hz (^1H - ^1H couplings). Values marked by an asterisk may be interchanged.

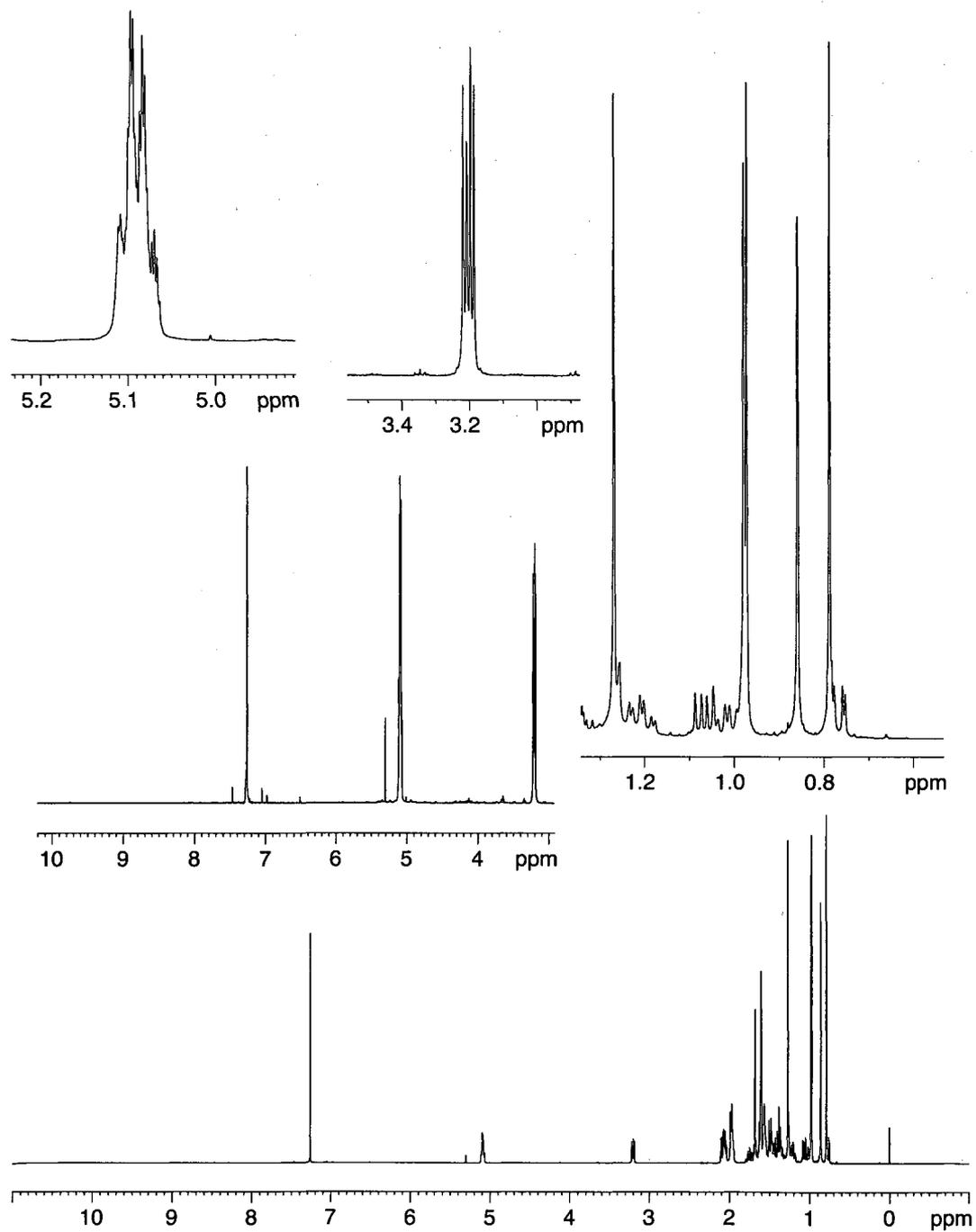


Figure 5. 800 MHz ^1H NMR spectrum of arabidiol.

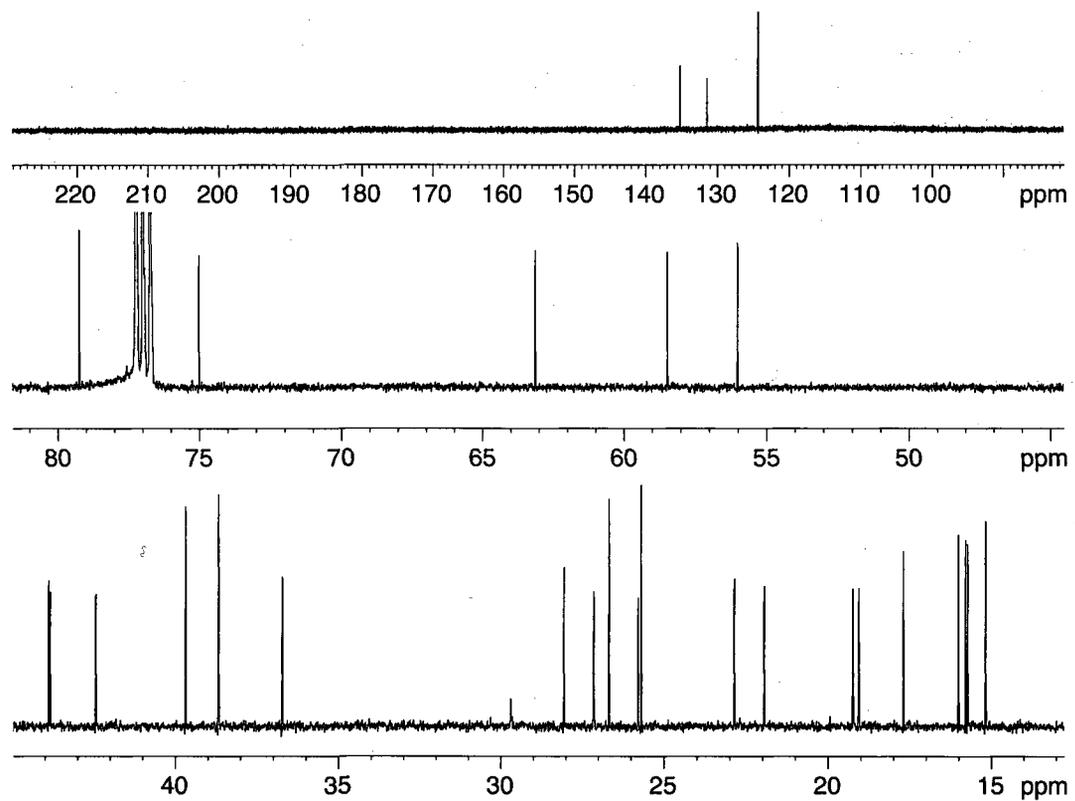
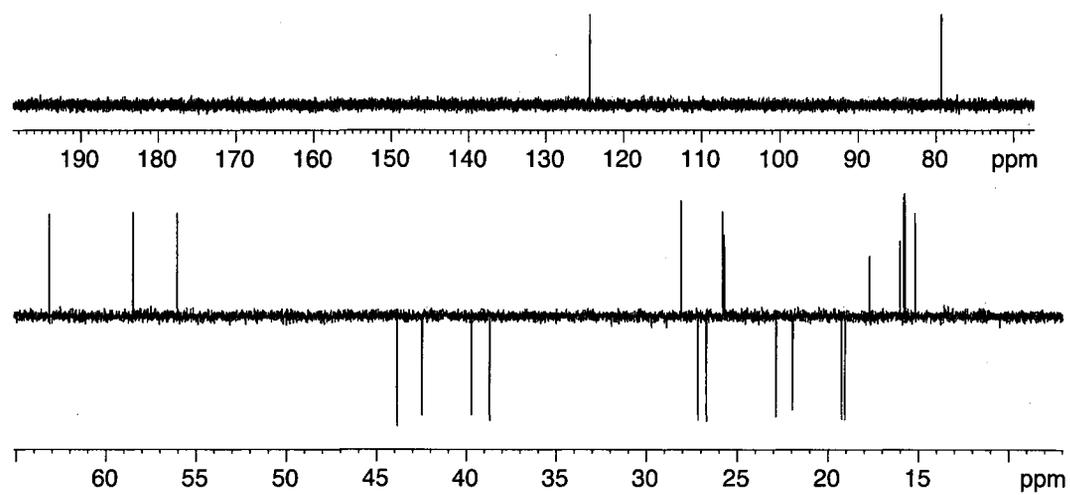
Figure 6. ^{13}C NMR of arabidiol.

Figure 7. DEPT spectra of arabidiol.

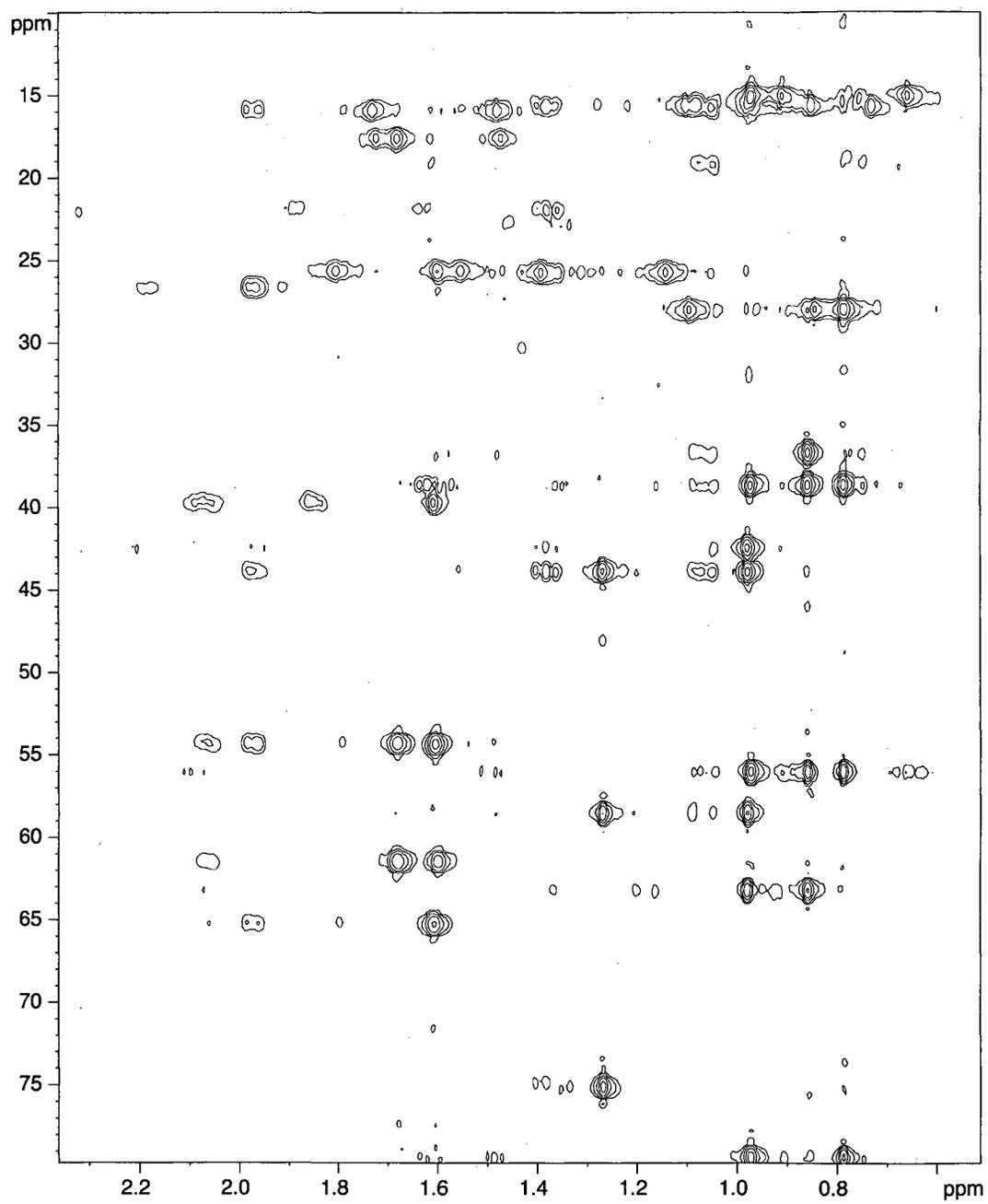


Figure 8. 2D HMBC spectrum of arabioidol.

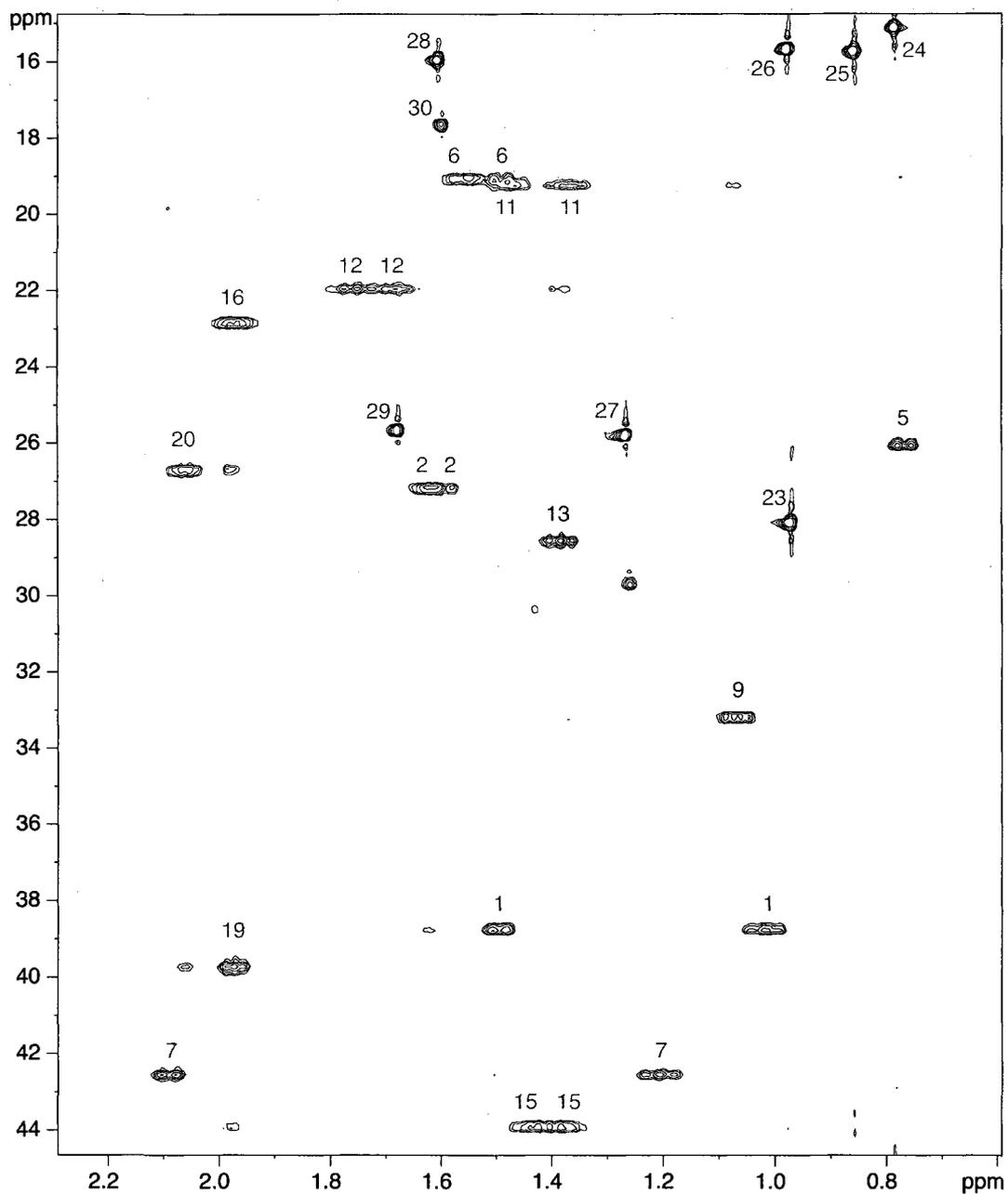


Figure 9. 2D HSQC spectrum of arabioidol.

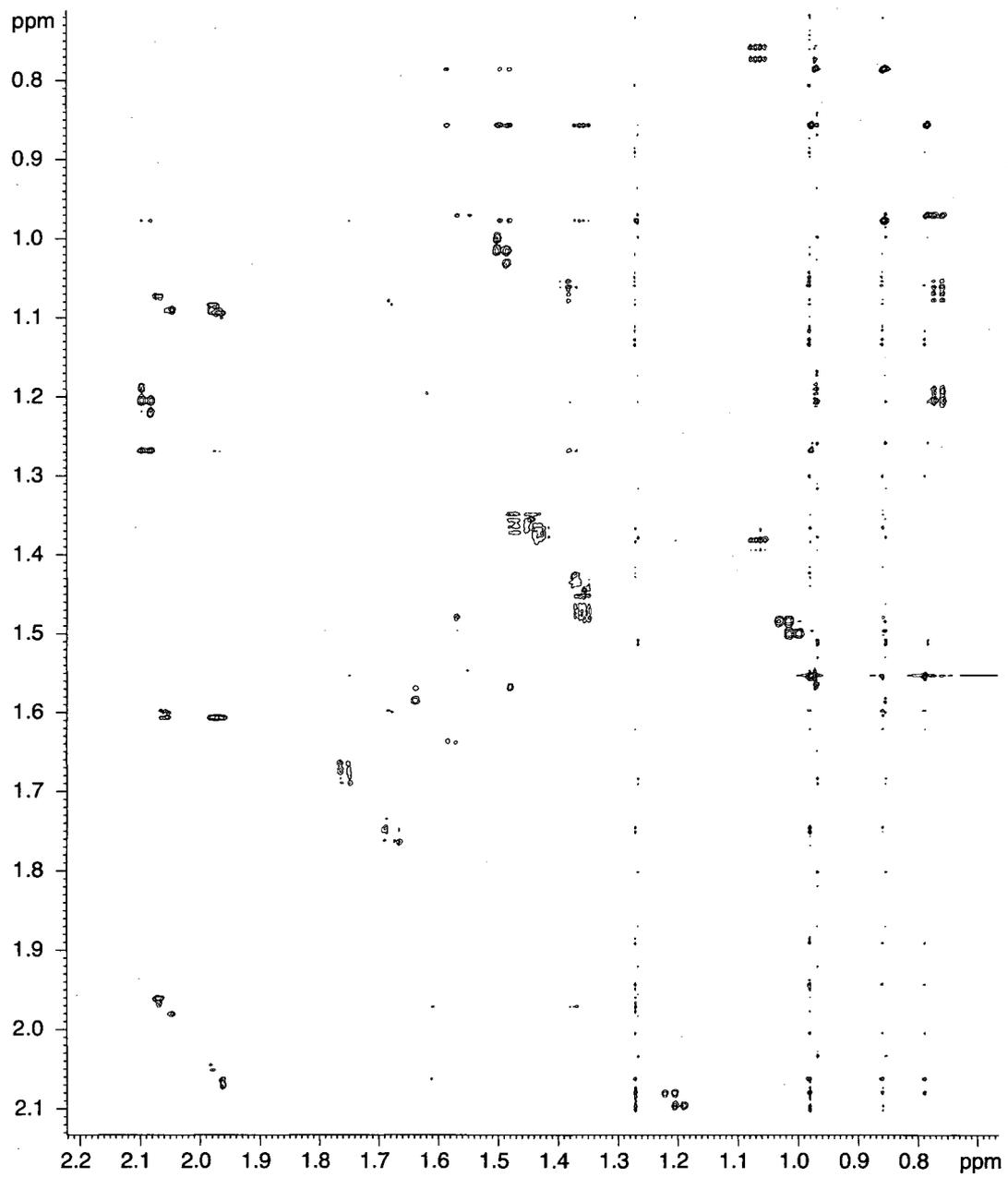


Figure 10. NOESY spectrum of arabidiol.

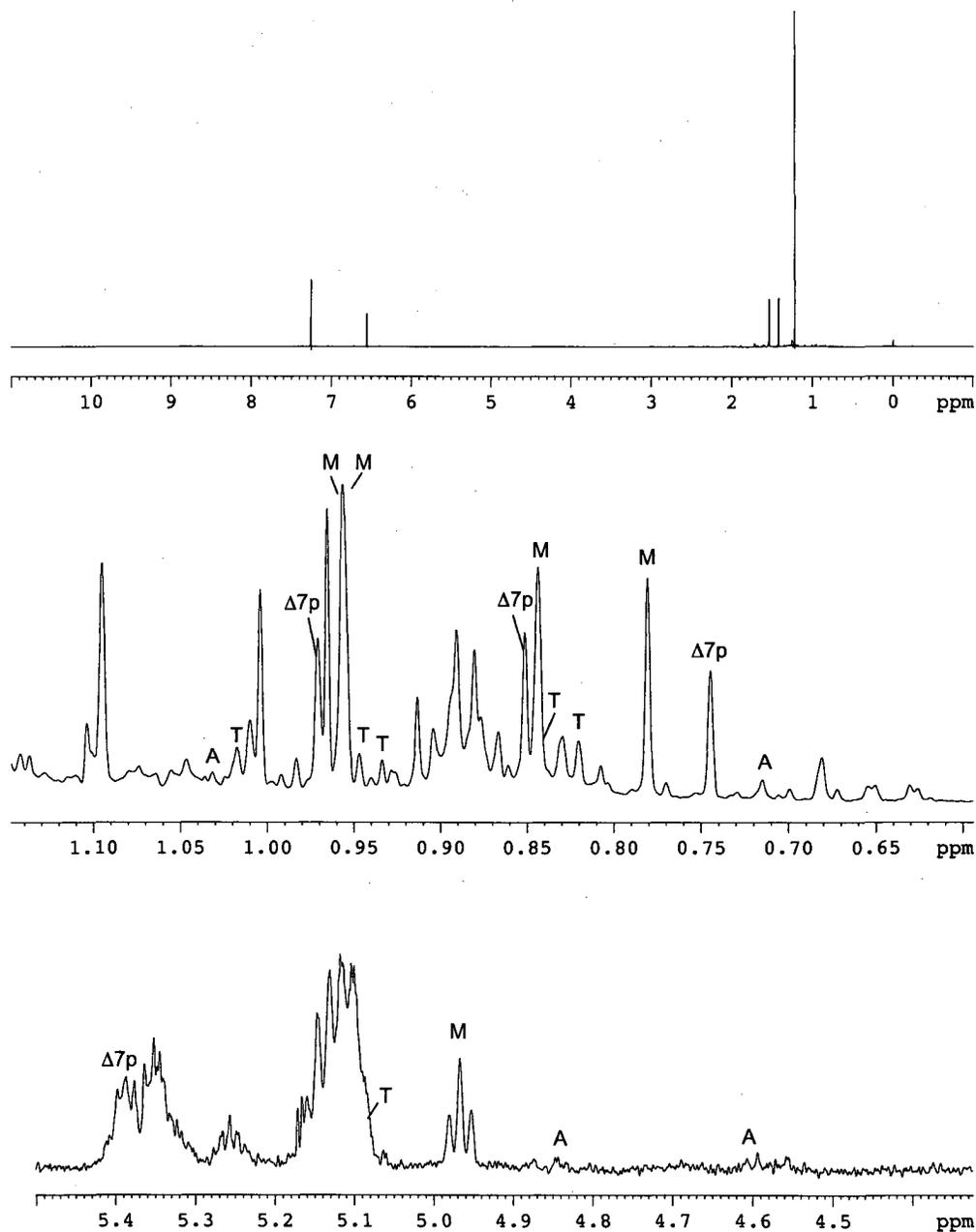


Figure 11. 500 MHz ¹H NMR analysis of column fraction 8 from SMY8[pMDK3.6]. Peaks corresponding to achilleol A are denoted as A, thalianol peaks denoted as T, peaks for Δ7-polypodatertaenol are assigned as Δ7p, and peaks for (13*R*,14*E*,17*E*)-malabarica-14,17,21-trien-3β-ol are shown as M.

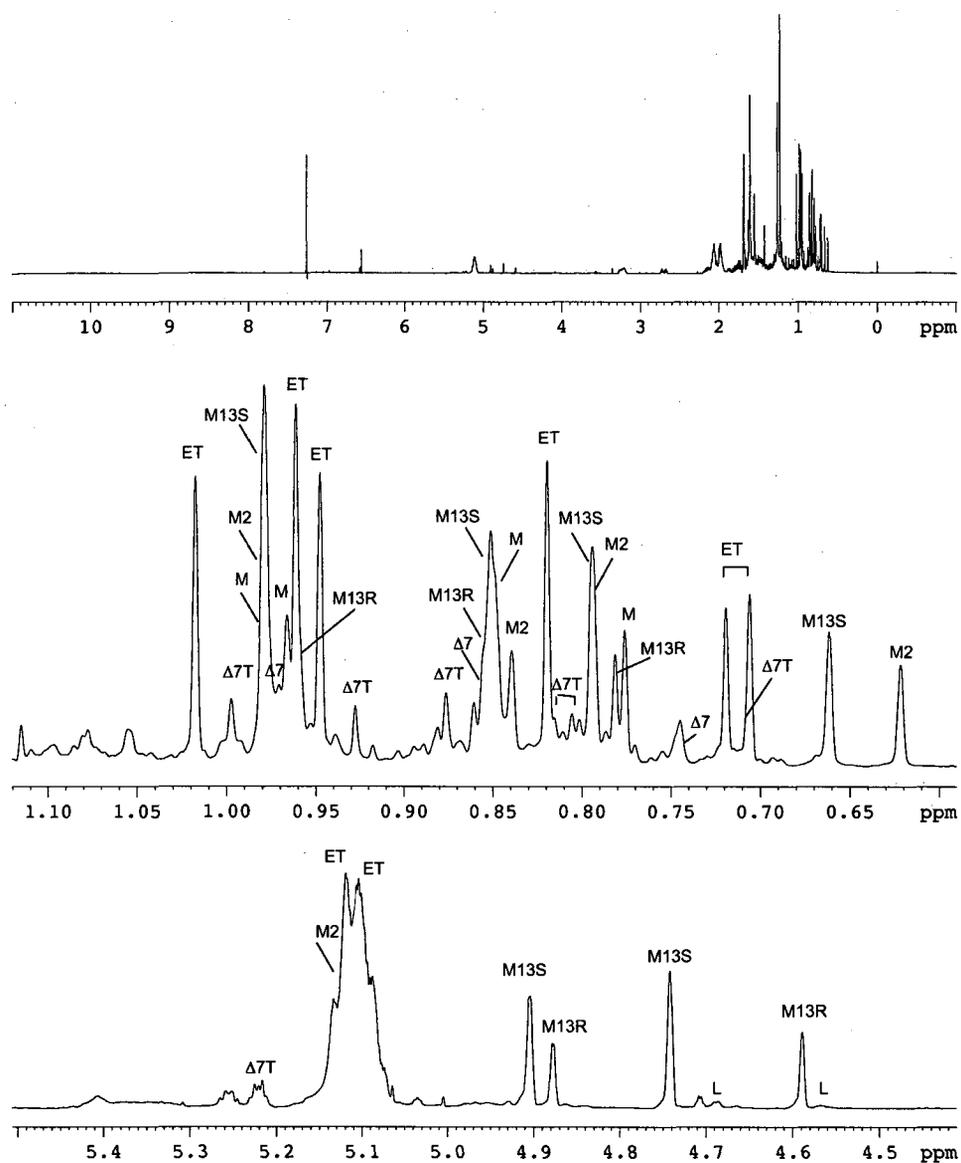


Figure 12. 500 MHz ¹H NMR analysis of column fraction 9 from SMY8[pMDK3.6]. Peaks corresponding to 14-epithalianol are denoted as ET, peaks for Δ7-polypodatertaenol are assigned as Δ7, and peaks for (13*R*,14*E*,17*E*)-malabarica-14,17,21-trien-3β-ol are assigned as M.

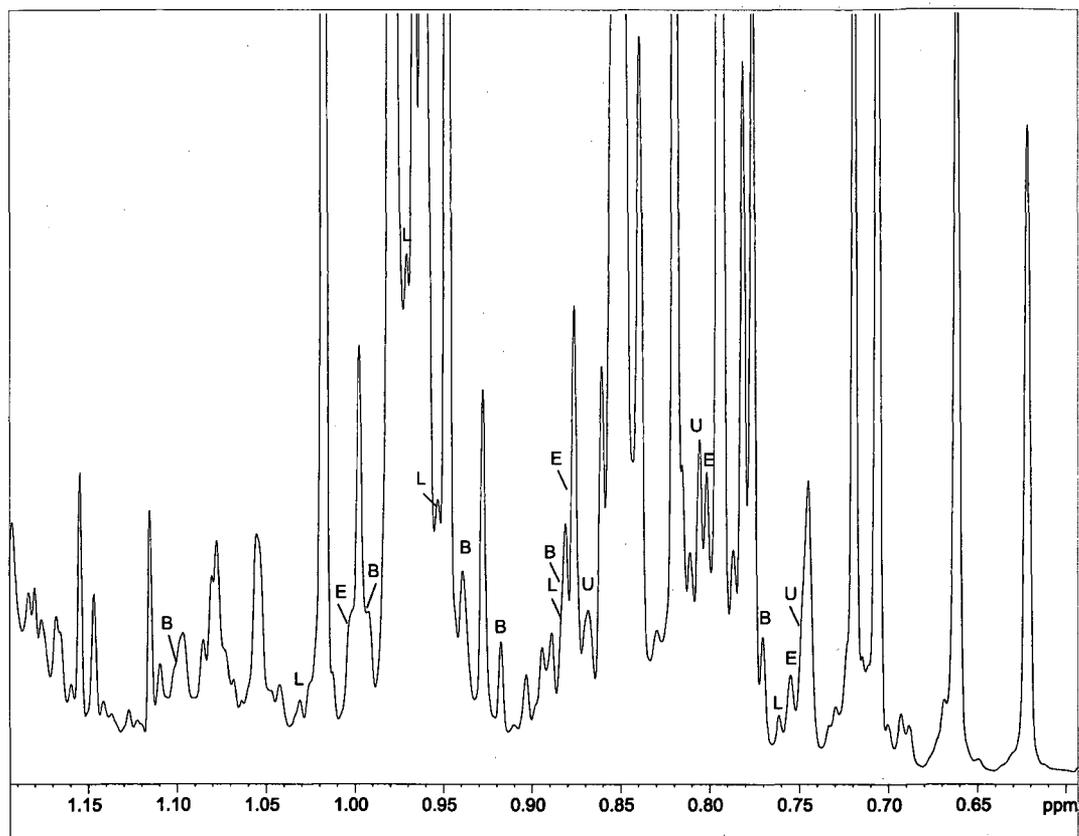


Figure 13. 500 MHz ¹H NMR analysis of column fraction 9 from SMY8[pMDK3.6] (expanded upfield region). Peaks corresponding to lupeol are denoted as L, peaks for euphol are assigned as E, and peaks for butyrospermol are assigned as U.

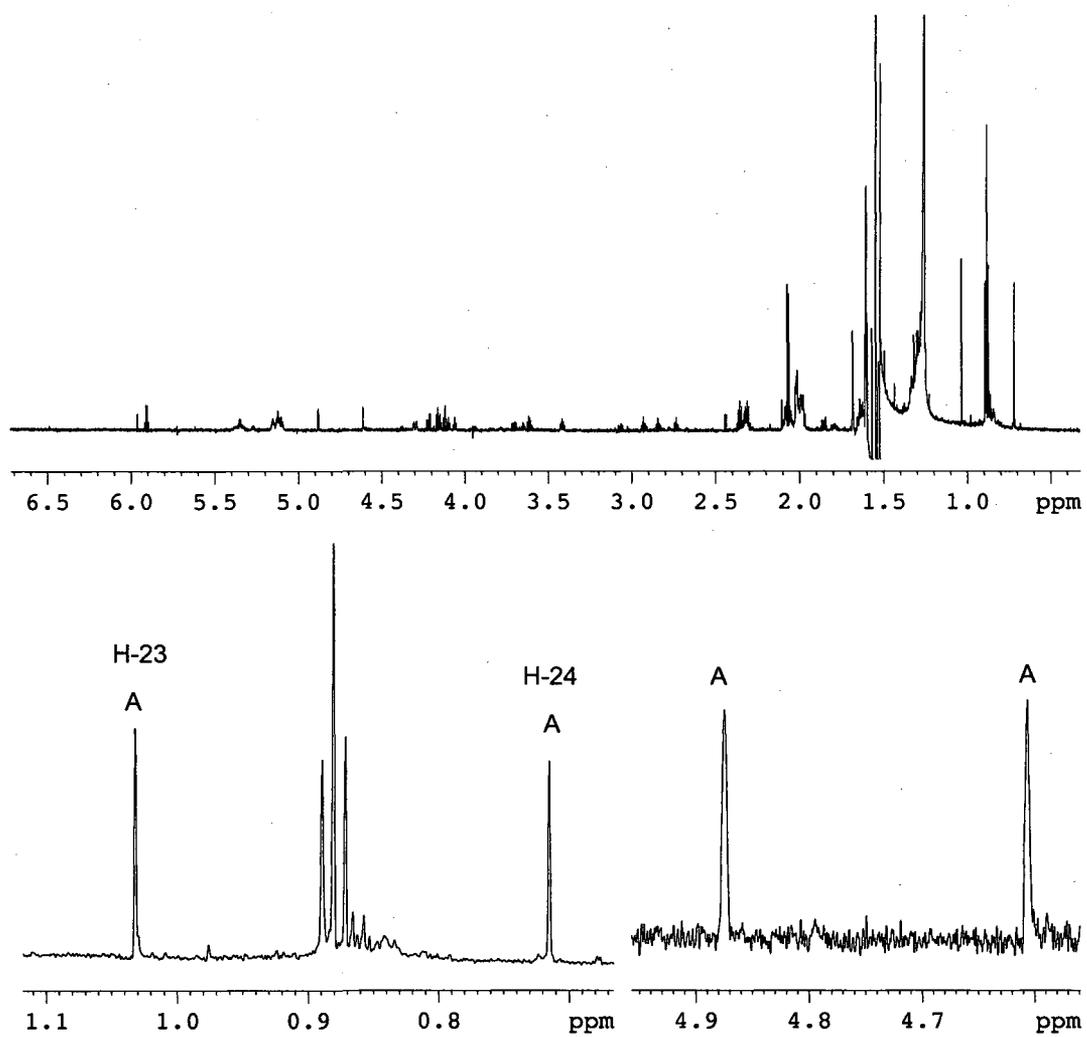


Figure 14. 800 MHz ^1H NMR analysis of column fraction 8 from SMY8[pMDK3.6], HPLC fractions 75-80. Peaks corresponding to achilleol A are assigned as A.

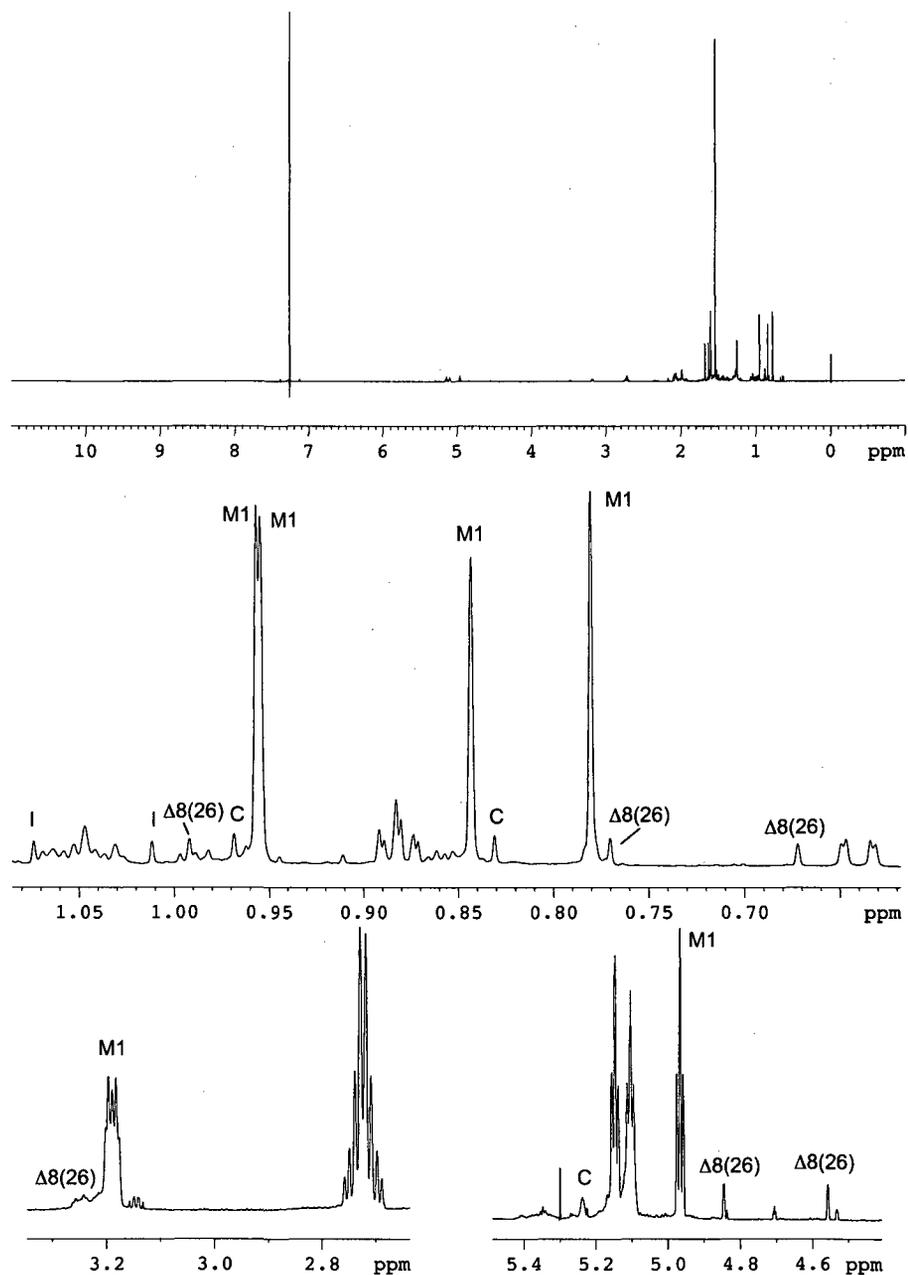


Figure 15. 800 MHz ^1H NMR analysis of column fraction 8 from SMY8[pMDK3.6], HPLC fractions 81-87. Peaks corresponding to camelliol C are assigned as C. Peaks for $\Delta 8(26)$ -polypodatetraenol are shown as $\Delta 8(26)$, and M1 denotes peaks corresponding to a novel tricyclic triterpene (13*R*,14*E*,17*E*)- malabarica-14,17,21-trien-3 β -ol.

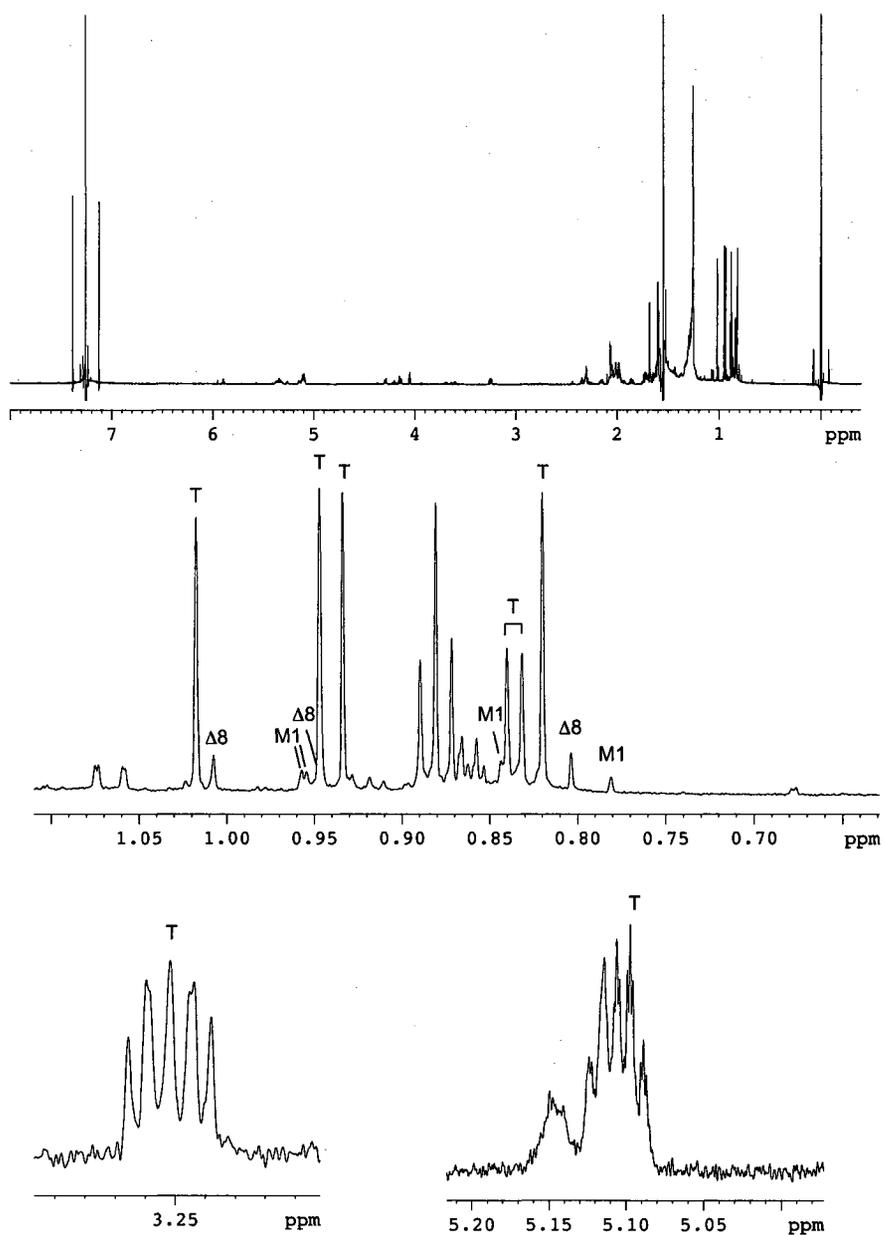


Figure 16. 800 MHz ^1H NMR analysis of column fraction 8 from SMY8[pMDK3.6], HPLC fractions 91-92. Peaks corresponding to thalianol are assigned as T. Peaks for $\Delta 8$ -polypodatetraenol are shown as $\Delta 8$, and M1 denotes peaks corresponding to (13*R*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol.

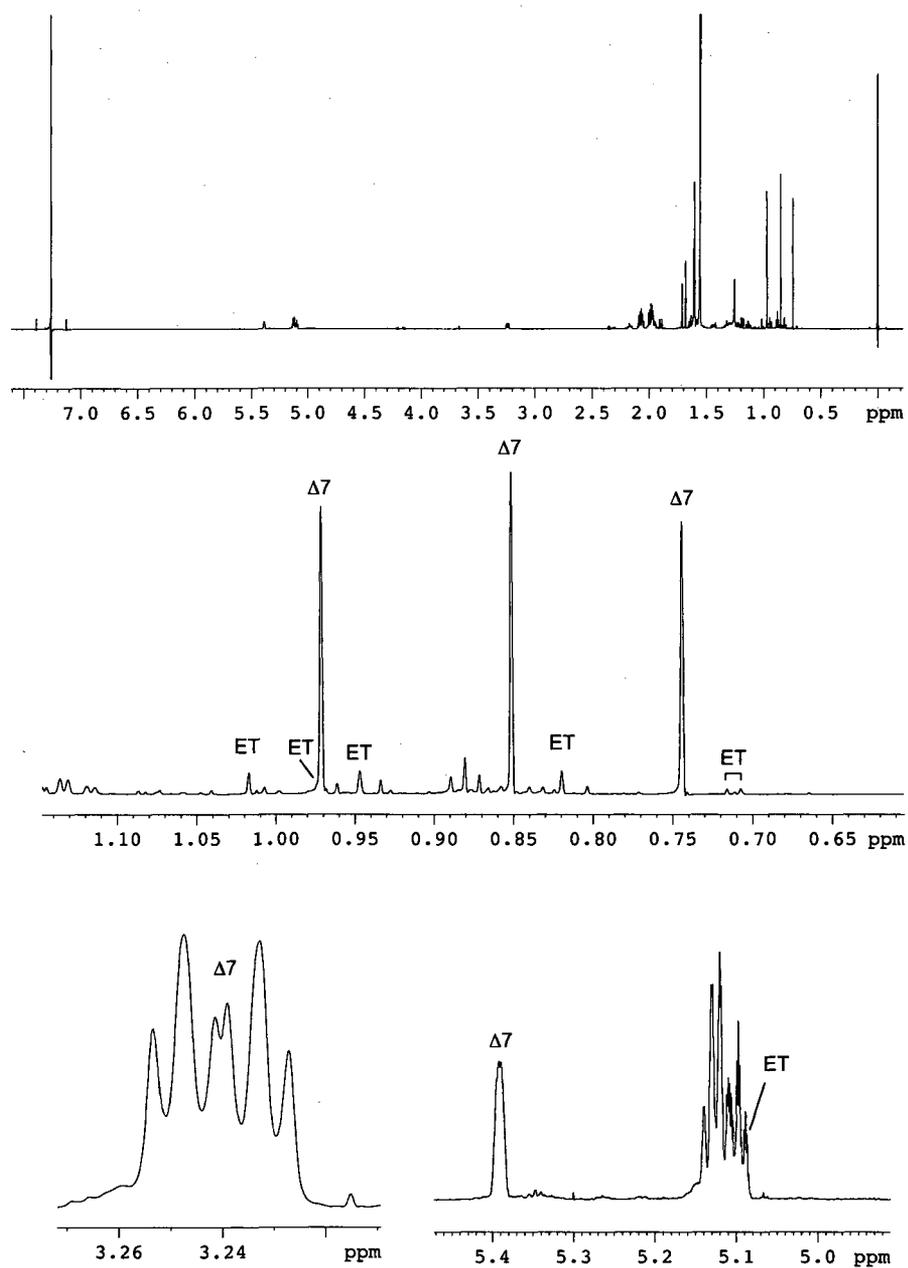


Figure 17. 800 MHz ¹H NMR analysis of column fraction 8 from SMY8[pMDK3.6], HPLC fractions 93-98. Peaks corresponding to 14-epithalianol are assigned as ET. Peaks for Δ7-polypodatetraenol are shown as Δ7.

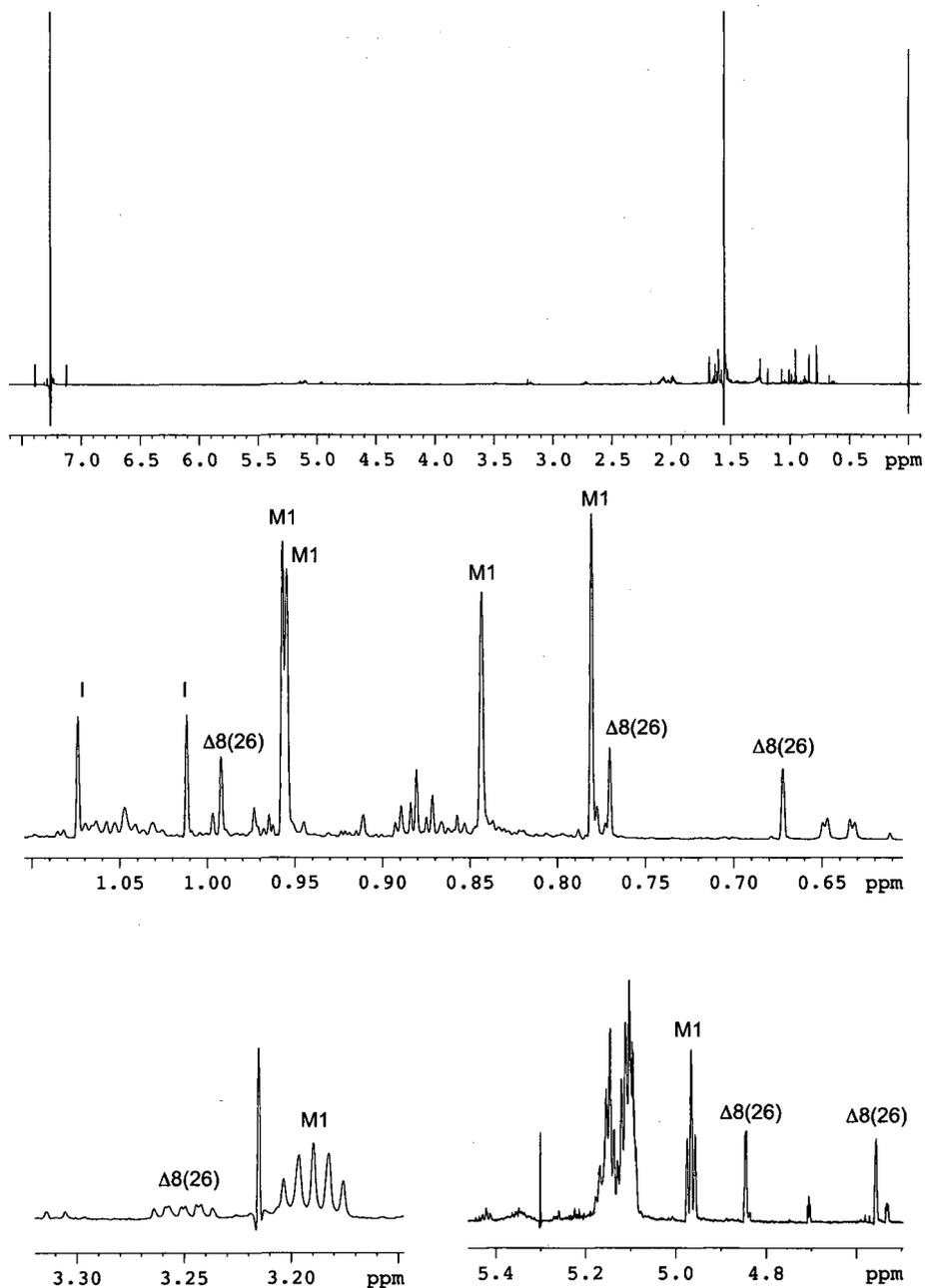


Figure 18. 800 MHz ¹H NMR analysis of column fraction 9 from SMY8[pMDK3.6], HPLC fractions 112-114. Peaks for Δ8(26)-polypodatetraenol are shown as Δ8, and M1 denotes peaks corresponding to (13*R*,14*E*,17*E*)-malabarica-14,17,21-trien-3β-ol.

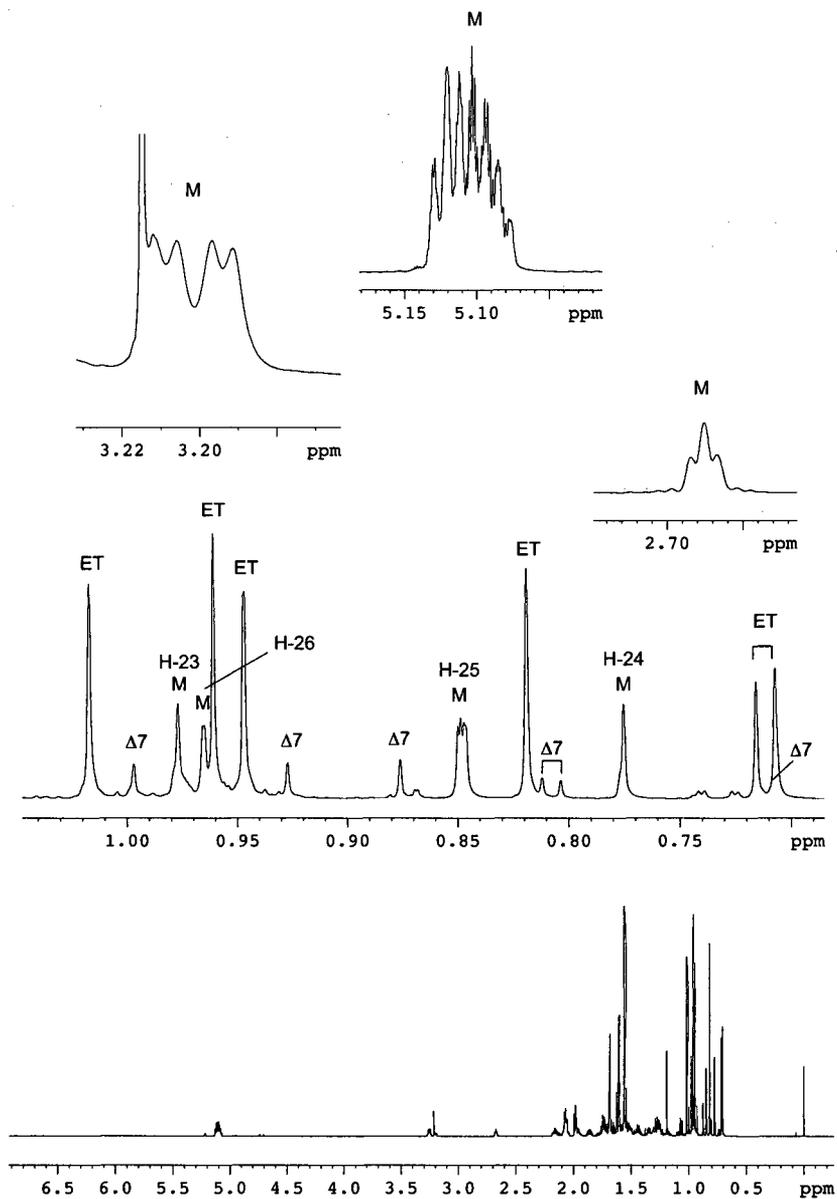


Figure 19. 800 MHz ^1H NMR analysis of column fraction 9 from SMY8[pMDK3.6], HPLC fractions 126-132. Peaks corresponding to 14-epithalianol are assigned as ET. Peaks for (13*R*,14*ξ*,17*E*)-malabarica-7,17,21-trien-3 β -ol are shown as Δ 7, and M indicates peaks corresponding to a novel tricyclic triterpene (13*R*,14*Z*,17*E*)-malabarica-14,17,21-trien-3 β -ol.

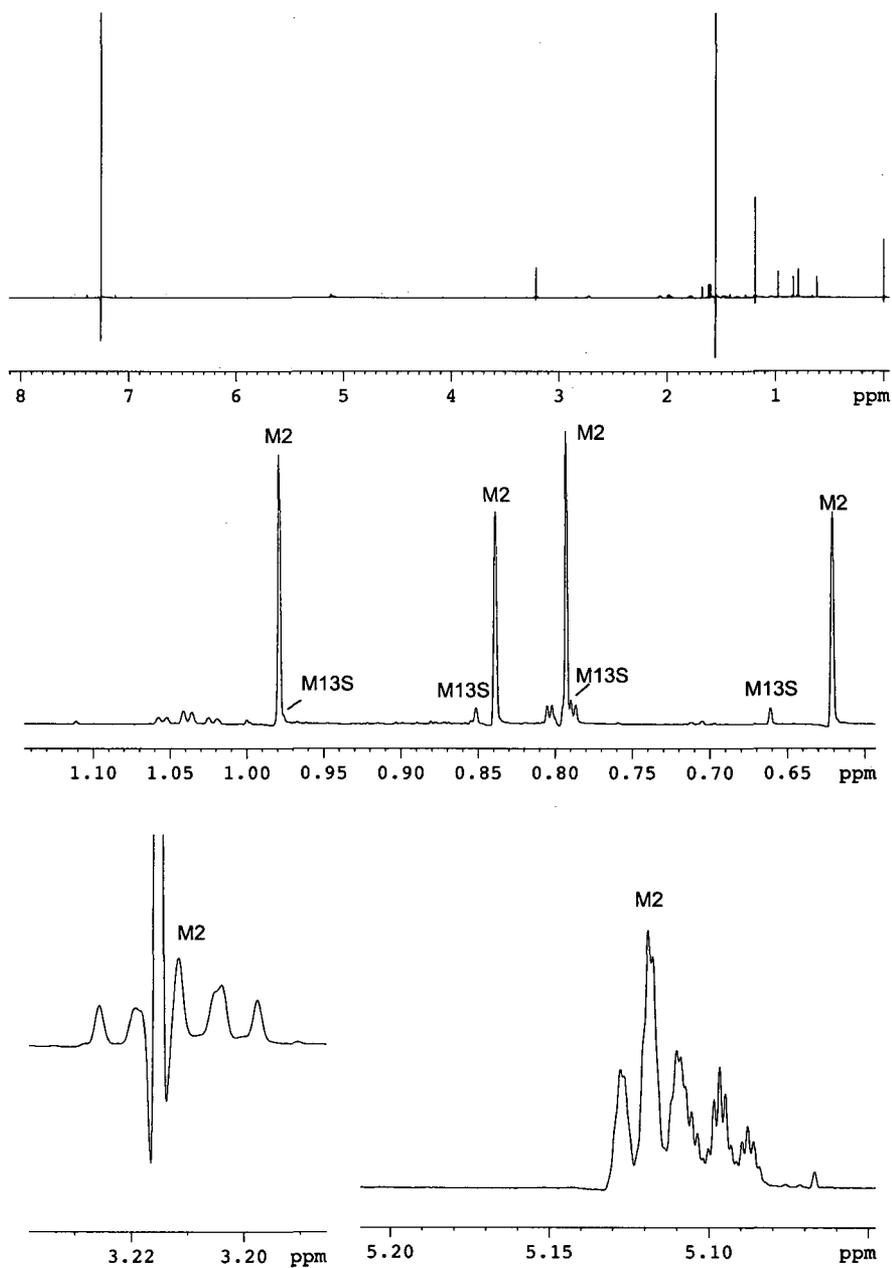


Figure 20. 800 MHz ¹H NMR analysis of column fraction 9 from SMY8[pMDK3.6], HPLC fractions 138-139. M2 indicates peaks corresponding to (13*S*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol. The M13S assigns peaks for (13*S*,17*E*)-malabarica-14(27),17,21-trien-3 β -ol.

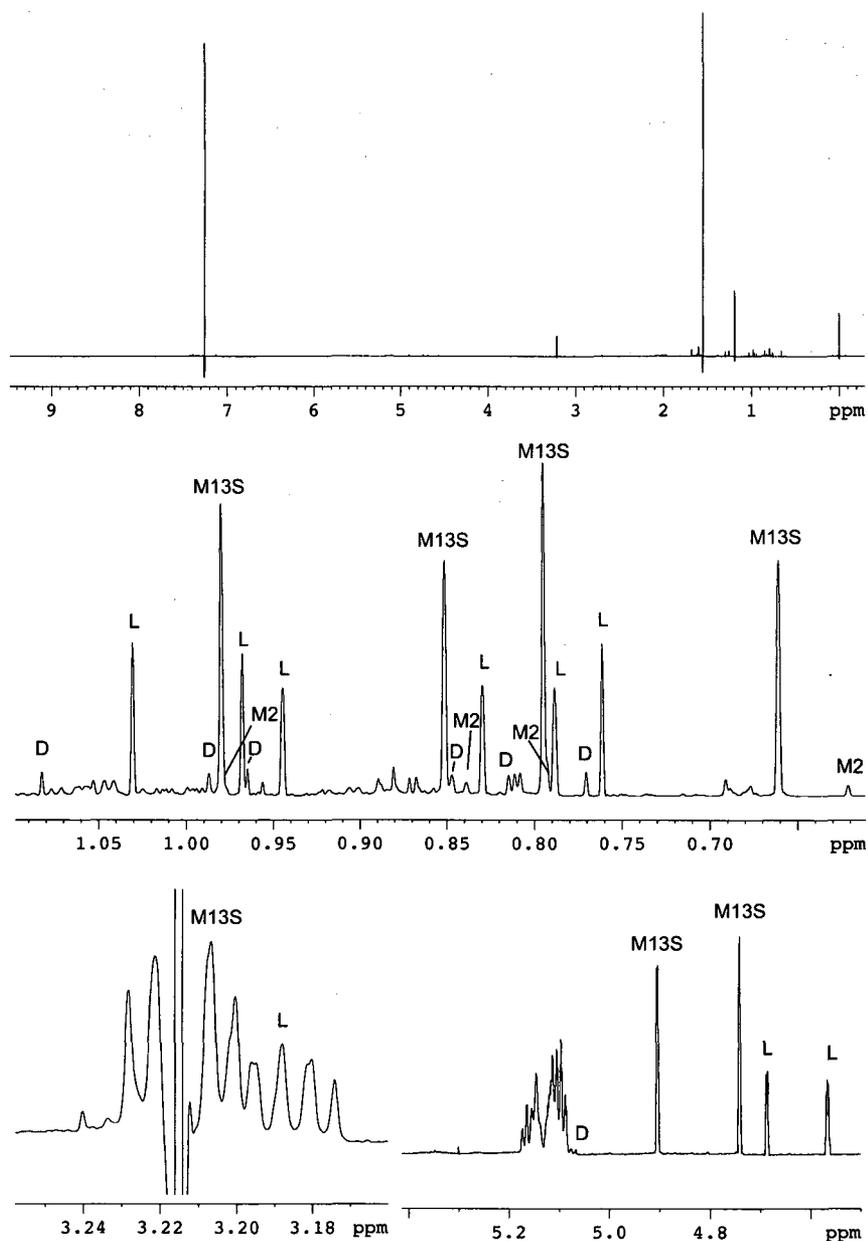


Figure 21. 800 MHz ^1H NMR analysis of column fraction 9 from SMY8[pMDK3.6], HPLC fractions 145-147. M2 indicates peaks corresponding to (13*S*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol. M13S assigns peaks for (13*S*,17*E*)-malabarica-14(27),17,21-trien-3 β -ol. L corresponds to lupeol signals, and the D sign indicates peaks for dammara-13(17),24-dienol.

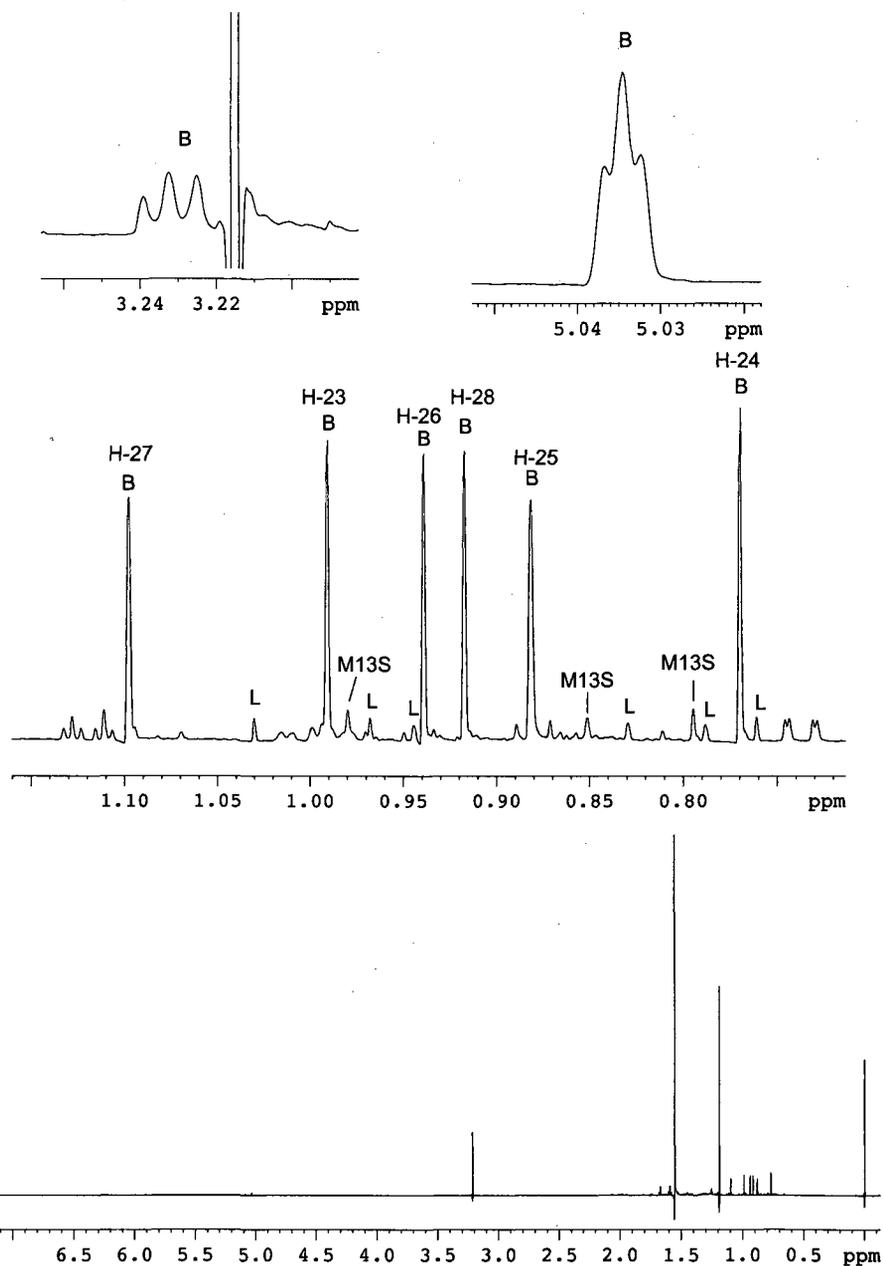


Figure 22. 800 MHz ¹H NMR analysis of column fraction 9 from SMY8[pMDK3.6], HPLC fractions 148-152. M13S assigns peaks for (13*S*,17*E*)-malabarica-14(27),17,21-trien-3 β -ol. The L sign corresponds to lupeol signals. Peaks corresponding to a novel tetracyclic triterpene bacchara-13(18),21-dienol are shown as B.

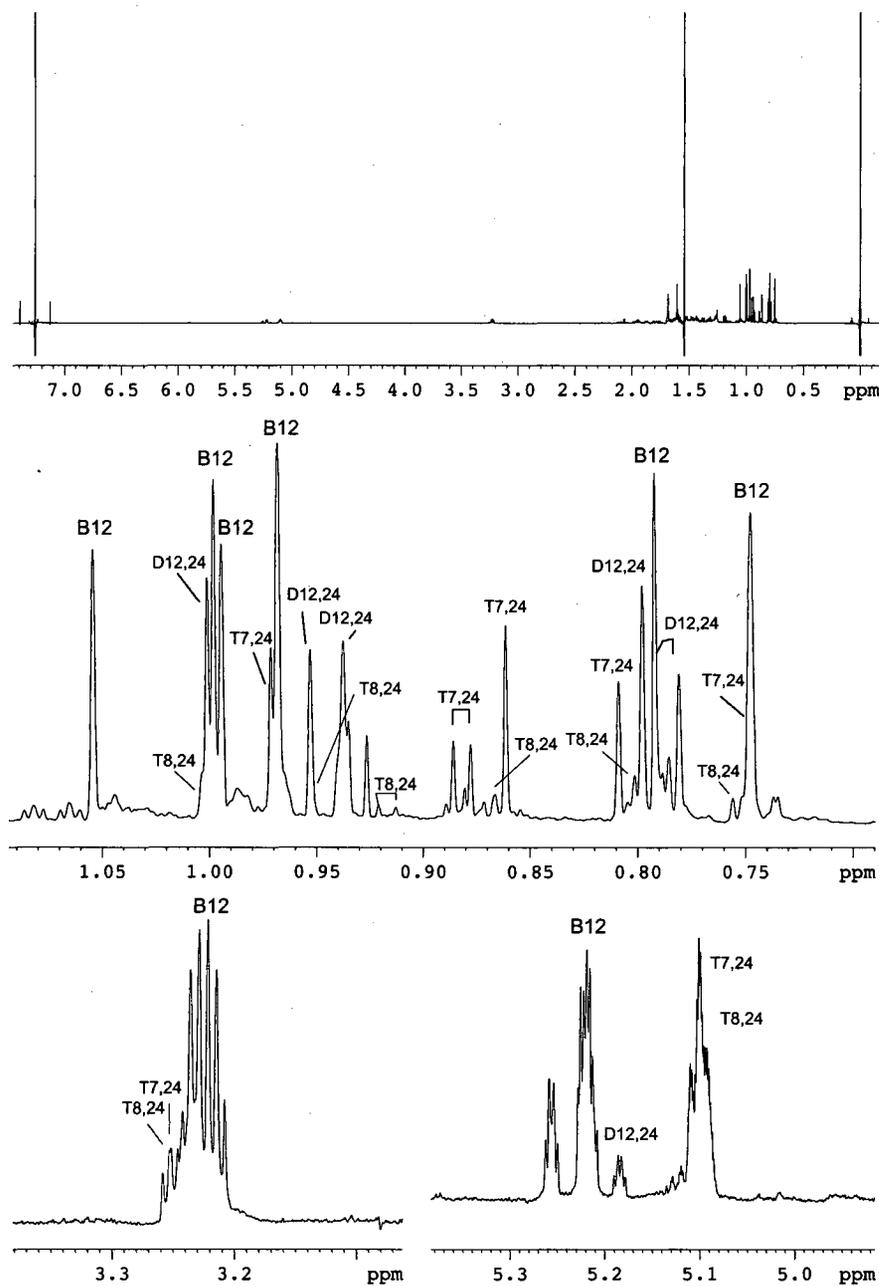


Figure 23. 800 MHz ^1H NMR analysis of column fraction 9 from SMY8[pMDK3.6], HPLC fractions 162-166. Peaks assigned for bacchara-12,21-dienol are shown as B12; for tirucalla-7,24-dienol as T7,24; for tirucalla-8,24-dienol as T8,24; and for dammara-12,24-dienol as D12,24.

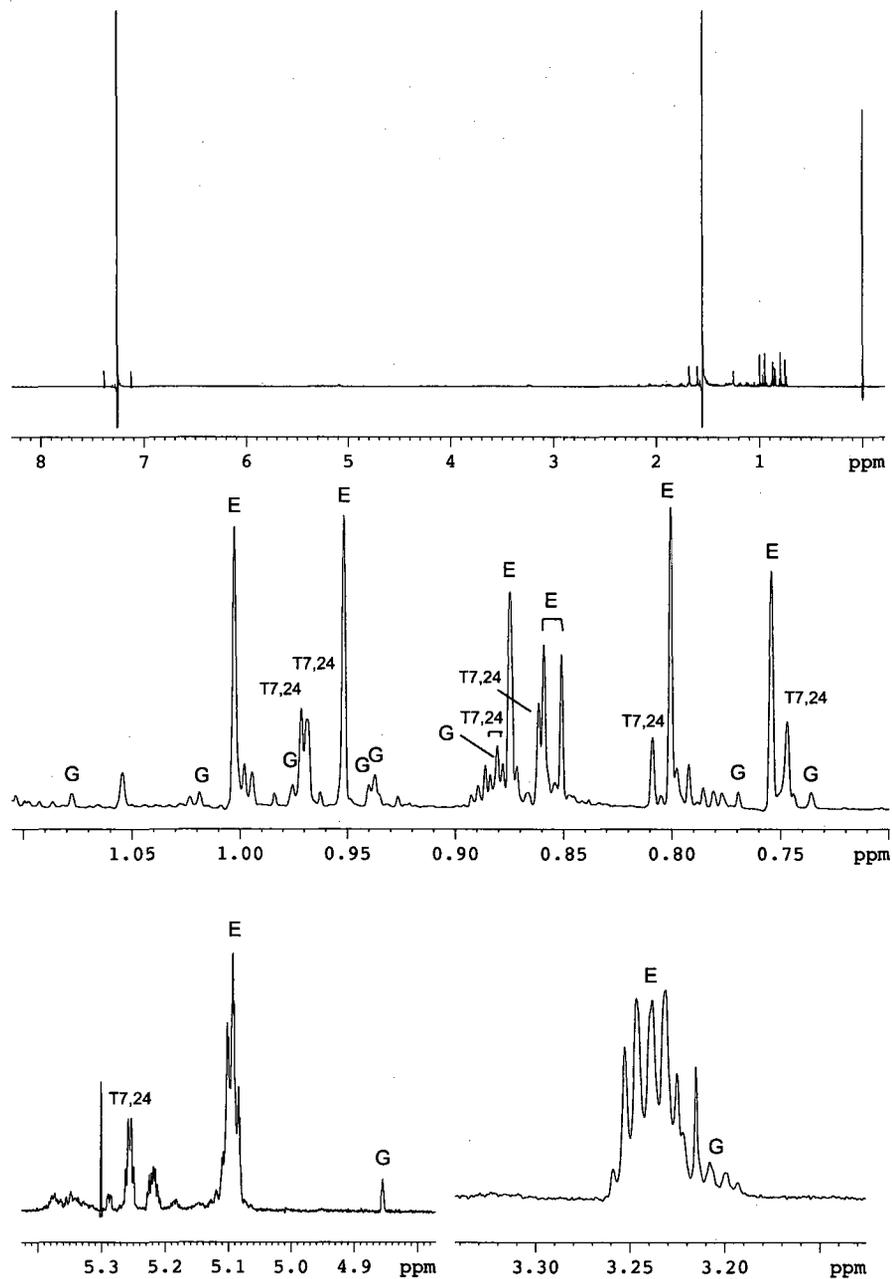


Figure 24. 800 MHz ^1H NMR analysis of column fraction 9 from SMY8[pMDK3.6], HPLC fractions 167-171. Peaks assigned for tirucalla-7,24-dienol are shown as T7,24; for germanicol as G; and for euphol as E.

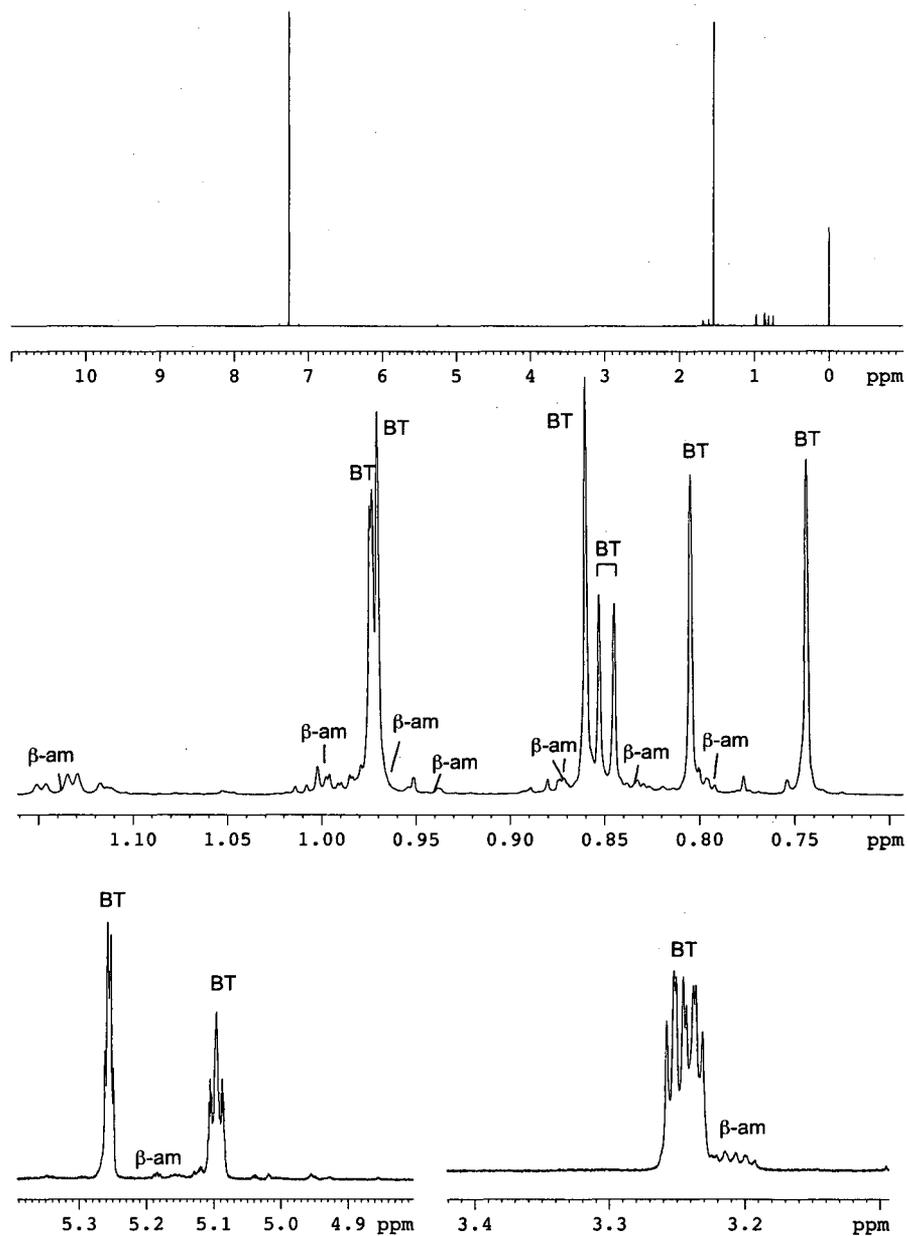


Figure 25. 800 MHz ^1H NMR analysis of column fraction 9 from SMY8[pMDK3.6], HPLC fractions 172-175. Peaks assigned for butyrospermol indicated as BT, peaks for β -amyryn are shown as β -am.

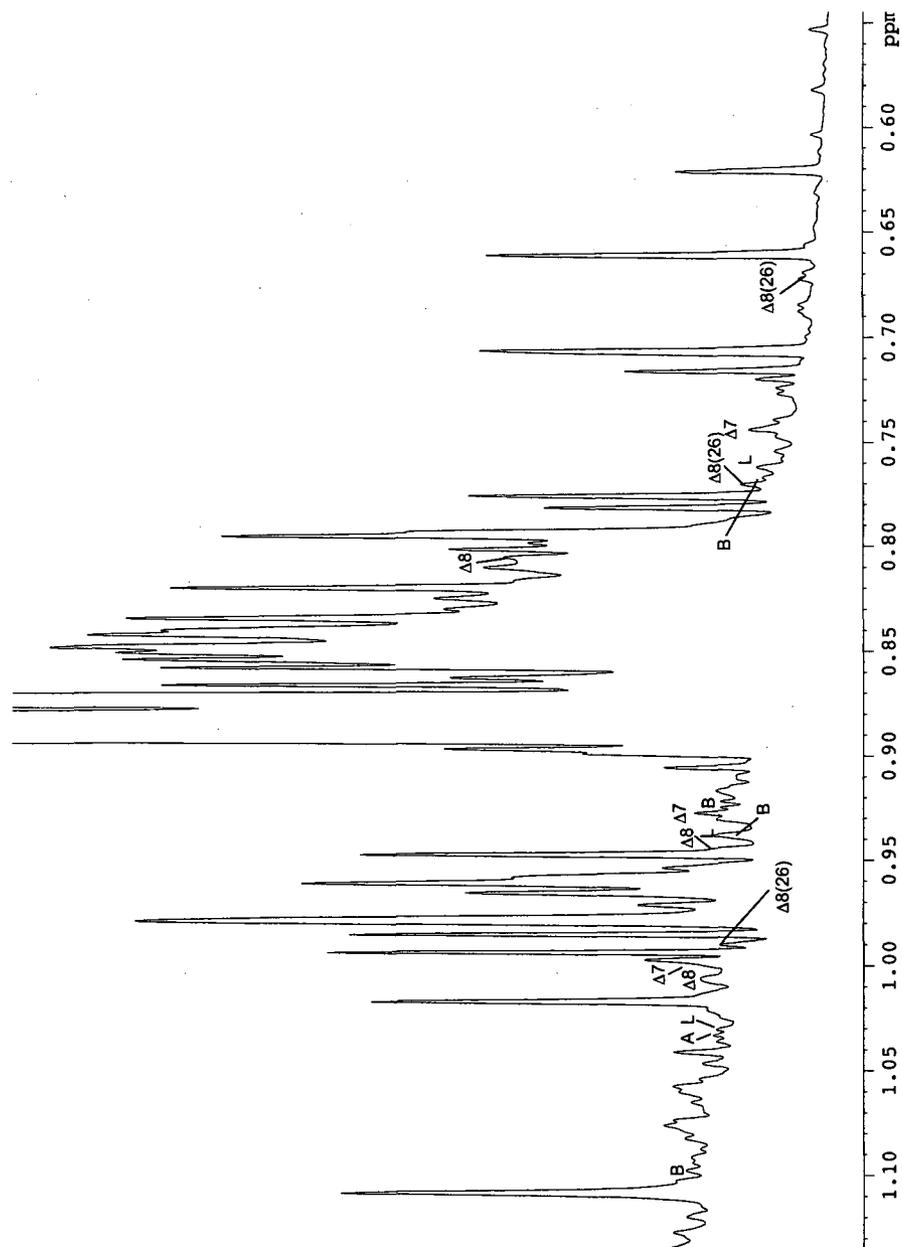


Figure 27. The expanded upfield region of the 800 MHz ^1H NMR spectrum of the preparative TLC fraction of RXY6[pMDK3.6] containing minor compounds. Characteristic peaks for minor compounds are denoted as: $\Delta 7$ for $\Delta 7$ -polypodatetraenol, $\Delta 8$ for $\Delta 8$ -polypodatetraenol, $\Delta 8(26)$ for $\Delta 8(26)$ -polypodatetraenol, L for lupeol, A for achilleol A, and B for bacchara-13(18),21-dienol.

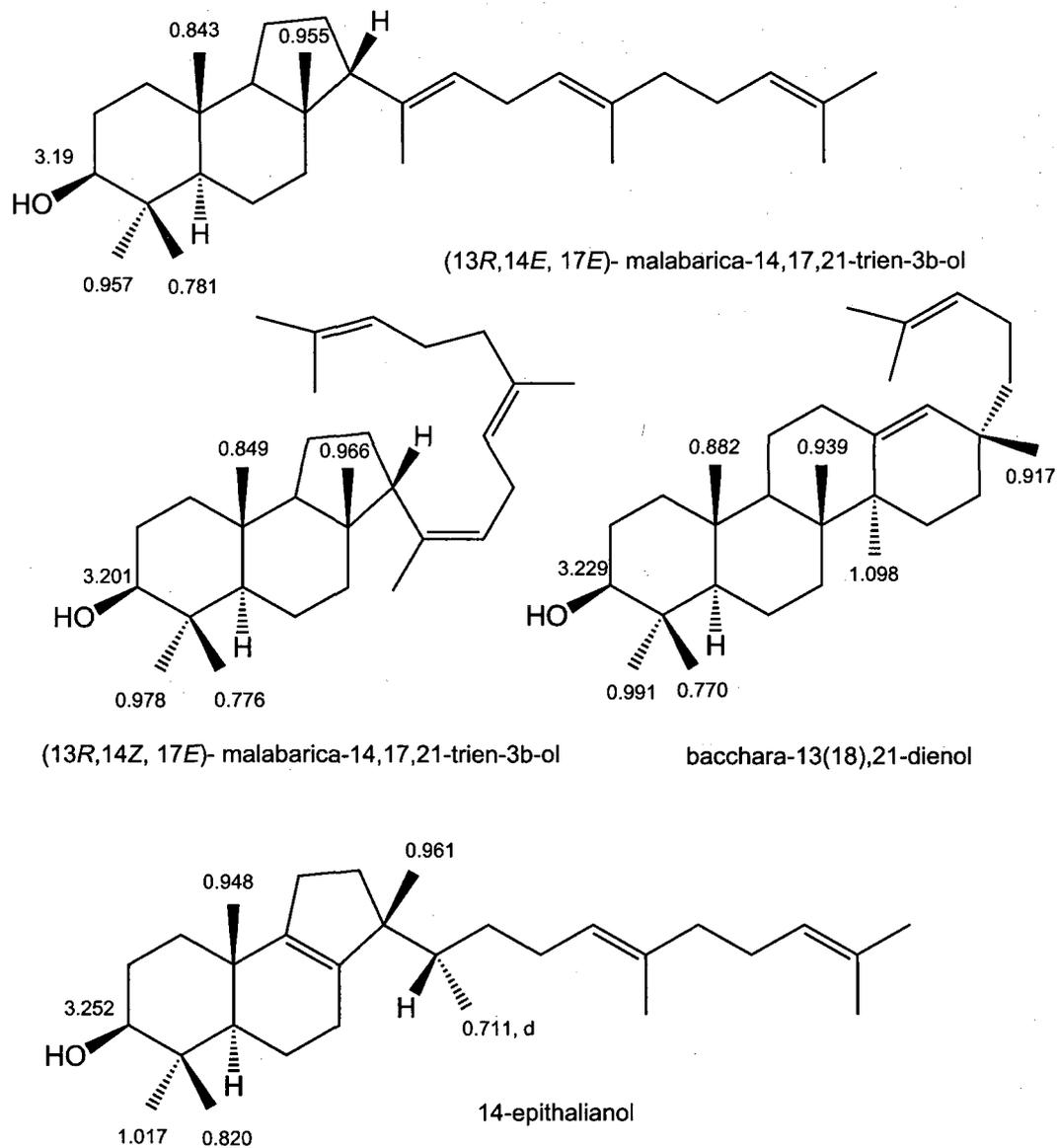


Figure 28. NMR assignments for upfield methyl singlets of 14-epithalianol, bacchara-13(18),21-dienol, (13*R*,14*Z*,17*E*)-malabarica-14,17,21-trien-3 β -ol, and (13*R*,14*E*,17*E*)-malabarica-14,17,21-trien-3 β -ol.

Appendix B

Amp	ampicillin
Asn, N	asparagine
BAM	β -amyrin synthase
BARS1	baruol synthase
BLAST	Basic Local Alignment Search Tool
BSTFA	bis(trimethylsilyl)trifluoroacetamide
CAMS1	camelliol C synthase
CAS	cycloartenol synthase
cDNA	complementary deoxyribonucleic acid
COSYDEC	decoupled ^1H - ^1H correlation NMR spectroscopy
Cys, C	cysteine
D, DEX	dextrose
DEPT	distortionless enhancement by polarization transfer
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphates
DOS	dioxidosqualene
DTT	dithiothreitol
E	ergosterol
EDTA	ethylenediamine tetraacetic acid
EI	electron impact
EtOH	ethanol
FID	flame ionization detector

G, GAL	galactose
GC	gas chromatography
GC-MS	gas chromatography mass spectrometry
GTAE	40mM tris base, 20 mM acetic acid, 1mM EDTA, 1 mM guanosine
HEME	hemin chloride
HMBC	heteronuclear multiple bond correlation
His, H	histidine
HPLC	high-performance liquid chromatography
HSQC	heteronuclear single-quantum coherence
Ile, I	isoleucine
Kan	kanamycin
LB	Luria Bertani broth
Leu, L	leucine
LAS	lanosterol synthase
LSS	lanosterol synthase
MPSS	massively parallel signature sequencing
mqH ₂ O	deionized milli Q water
MRN1	marneral synthase
mRNA	messenger ribonucleic acid
MS	mass spectra
MTBE	methyl <i>tert</i> -butyl ether

NADPH	protonated form of nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NMO	<i>N</i> -methylnmorpholine- <i>N</i> -oxide
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
NOESY	nuclear Overhauser effect spectroscopy
NSL	non-saponifiable lipids
OD	optical density
ORF	open reading frame
OS	(<i>S</i>)-2,3-oxidosqualene
OSC(s)	oxidosqualene cyclase(s)
PCR	polymerase chain reaction
PEG	polyethylene glycol
PMMA	polymethylmethacrylate
Pro, P	proline
PTLC	preparative thin layer chromatography
RNAse	ribonuclease
RT-PCR	reverse transcription polymerase chain reaction
SC	synthetic complete
SDS	sodium dodecyl sulfate
SHC	squalene hopene cyclase
ssDNA	single stranded DNA

TAE	40 mM tris base, 20 mM acetic acid, 1mM EDTA
TE8	10 mM tris-HCl (pH 8), 1mM EDTA
THAS1	thalianol synthase
THF	tetrahydrofuran
Thr, T	threonine
TMS	trimethylsilane (trimethylsilyl)
TPAP	tetrapropylammonium perruthenate
TPS10	maize sesquiterpene synthase
Trp, W	tryptophan
Tyr, Y	tyrosine
TLC	thin layer chromatography
TMS	trimethylsilane/trimethylsilyl
trHMG1	<i>Saccharomyces cerevisiae</i> truncated hydroxymethylglutaryl CoA reductase
Tris-HCl	tris(hydroxymethyl)aminomethane hydrochloride
Ura, U	uracil
UV	ultraviolet
Val, V	valine
YP	yeast extract/peptone

References

- ¹ Xu, R.; Fazio, G. C.; Matsuda, S. P. T. *Phytochemistry* **2004**, *65*, 261-290.
- ² Hoshino, T.; Sato, T. *Chem. Commun. (Camb.)* **2002**, *4*, 291-301.
- ³ Thoma, R.; Schulz-Gasch, T.; D'Arcy, B.; Benz, J.; Aebi, J.; Dehmlow, H.; Henning, M.; Stihle, M.; Ruf, A. *Nature* **2004**, *432*, 118-122.
- ⁴ Van Tamelen, E. E.; Willett, J. D.; Clayton, R. B. *J. Am. Chem. Soc.* **1967**, *89*, 3371-3373.
- ⁵ Willett, J. D.; Sharpless, K. B.; Lord, K. E.; Van Tamelen, E. E.; Clayton, R. B. *J. Biol. Chem.* **1967**, *242*, 4182-4191.
- ⁶ Bentley, H. R.; Henry, J. A.; Irvine, D. S.; Spring, F. S. *J. Chem. Soc.* **1953**, 3673-3678.
- ⁷ Irvine, D. S.; Henry, J. A.; Spring, F. S. *J. Chem. Soc.* **1955**, 1316-1320.
- ⁸ Zincke, A.; Friedrich, A.; Rollett, A. *Monatsh. Chem.* **1920**, *41*, 253-270.
- ⁹ Goad, L. J.; Akihisa, T. *Analysis of Sterols*; Blackie (Chapman & Hall): London, **1997**, appendix 3, section (h).
- ¹⁰ Thoms, H. *Arch. Pharm.* **1897**, *235*, 39-43.
- ¹¹ Connolly, J. D.; Hill, R. A., Triterpenoids. *Nat. Prod. Rep.* **2002**, *19*, 494-513.
- ¹² Connolly, J. D.; Hill, R. A., Triterpenoids. *Nat. Prod. Rep.* **2005**, *22*, 487-503.
- ¹³ Huhman, D. V.; Berhow, M. A.; Sumner, L. W. *J. Agric. Food Chem.* **2005**, *53*, 1914-1920.
- ¹⁴ Kunst, L.; Samuels, A. L. *Prog. Lipid Res.* **2003**, *42*, 51-80.
- ¹⁵ Shan, H.; Wilson, W. K.; Phillips, D. R.; Bartel, B.; Matsuda, S. P. T. *Org. Lett.* **2008**, *10*, 1897-1900.
- ¹⁶ Haralampidis, K.; Bryan, G.; Qi, X.; Papadopoulou, K.; Bakht, S.; Melton, R.;

-
- Osbourn, A. E. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 13431-13436.
- ¹⁷ Qi, X.; Bakht, S.; Leggett, M.; Maxwell, C.; Melton, R.; Osbourn, A. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 8233-8238.
- ¹⁸ Hayashi, H.; Huang, P.; Takada, S.; Obinata, M.; Inoue, K.; Shibuya, M.; Ebizuka, Y. *Biol. Pharm. Bull.* **2004**, *27*, 1086-1092.
- ¹⁹ Chaturvedi, P. K.; Bhui, K.; Shukla, Y. *Cancer Lett.* **2008**, *263*, 1 - 13.
- ²⁰ Husselstein-Muller, T.; Schaller, H.; Benveniste, P. *Plant Mol. Biol.* **2001**, *45*, 75-92.
- ²¹ Jenks, M. A.; Tuttle, H. A.; Eigenbrode, S. D.; Feldmann, K. A. *Plant Phys.* **1995**, *108*, 359-377.
- ²² Ohyama, K.; Suzuki, M.; Masuda, K.; Yoshida, S.; Muranaka, T. *Chem. Pharm. Bull.* **2007**, *55*, 1518-1521.
- ²³ D'Auria, J. C.; Gershenzon, J. *Curr. Opin. Plant Biol.* **2005**, *8*, 308-316.
- ²⁴ Phillips, D. R.; Rasbery, J. M.; Bartel, B.; Matsuda, S. P. T. *Curr. Opin. Plant Biol.* **2006**, *9*, 305-314.
- ²⁵ Bachmann, B. O. *Nat. Chem. Biol.* **2005**, *1*, 244-245.
- ²⁶ www.ncbi.nlm.nih.gov
- ²⁷ Schnee, C.; Kollner, T. G.; Held, M.; Turlings, T. C. J.; Gershenzon, J.; Degenhardt, J. *Proc. Nat. Acad. Sci. USA* **2006**, *103*, 1129-1134.
- ²⁸ *Arabidopsis thaliana*: The Arabidopsis Genome Initiative. *Nature* **2000**, *408*, 796-815;
Oryza sativa: International Rice Genome Sequencing Project. *Nature*, **2005**. *436*, 793-800; *Populus trichocarpa*: Tuskan, G. A.; DiFazio, S.; Jansson, S.; Bohlmann, J.; Grigoriev, I.; Hellsten, U.; Putnam, N.; Ralph, S.; Rombauts, S.; Salamov, A.; Schein, J.; Sterck, L.; Aerts, A.; Bhalerao, R. R.; Bhalerao, R. P.; Blaudez, D.; Boerjan, W.;

Brun, A.; Brunner, A.; Busov, V.; Campbell, M.; Carlson, J.; Chalot, M.; Chapman, J.; Chen, G. L.; Cooper, D.; Coutinho, P. M.; Couturier, J.; Covert, S.; Cronk, Q.; Cunningham, R.; Davis, J.; Degroeve, S.; DeJardin, A.; dePamphilis, C.; Detter, J.; Dirks, B.; Dubchak, I.; Duplessis, S.; Ehlting, J.; Ellis, B.; Gendler, K.; Goodstein, D.; Gribskov, M.; Grimwood, J.; Groover, A.; Gunter, L.; Hamberger, B.; Heinze, B.; Helariutta, Y.; Henrissat, B.; Holligan, D.; Holt, R.; Huang, W.; Islam-Faridi, N.; Jones, S.; Jones-Rhoades, M.; Jorgensen, R.; Joshi, C.; Kangasjarvi, J.; Karlsson, J.; Kelleher, C.; Kirkpatrick, R.; Kirst, M.; Kohler, A.; Kalluri, U.; Larimer, F.; Leebens-Mack, J.; Leple, J. C.; Locascio, P.; Lou, Y.; Lucas, S.; Martin, F.; Montanini, B.; Napoli, C.; Nelson, D. R.; Nelson, C.; Nieminen, K.; Nilsson, O.; Pereda, V.; Peter, G.; Philippe, R.; Pilate, G.; Poliakov, A.; Razumovskaya, J.; Richardson, P.; Rinaldi, C.; Ritland, K.; Rouze, P.; Ryaboy, D.; Schmutz, J.; Schrader, J.; Segerman, B.; Shin, H.; Siddiqui, A.; Sterky, F.; Terry, A.; Tsai, C. J.; Uberbacher, E.; Unneberg, P. *Science* **2006**, *313*, 1596-1604.

²⁹ Abe, I. *Nat. Prod. Rep.* **2007**, *24*, 1311-1331.

³⁰ Tchen, T. T.; Bloch, K. J. *J. Am. Chem. Soc.* **1955**, *77*, 6085-6086.

³¹ Heintz, R.; Benveniste, P. *J. Biol. Chem.* **1974**, *249*, 4267-4274.

³² Ehrhardt, J. D.; Hirth, L.; Ourisson, G. *Phytochemistry* **1967**, *6*, 815-821.

³³ Alcaide, A.; Devys, M.; Barbier, M. *Phytochemistry* **1968**, *7*, 329-330.

³⁴ Williams, B. L.; Goodwin, T. W. *Phytochemistry* **1965**, *4*, 81-88.

³⁵ Giner, J. L.; Berkowitz, J. D.; Andersson, T. *J. Nat. Prod.* **2000**, *63*, 267-269.

³⁶ Gibbons, G. F.; Goad, L. J.; Goodwin, T. W.; Nes, W. R. *J. Biol. Chem.* **1971**, *246*, 3967-3976.

-
- ³⁷ Hall, J.; Smith, A. R.; Goad, L. J.; Goodwin, T. W. *Biochem. J.* **1969**, *112*, 129-130.
- ³⁸ Raederstorff, D.; Rohmer, M. *Biochem. J.* **1985**, *231*, 609-615.
- ³⁹ Raederstorff, D.; Rohmer, M. *Eur. J. Biochem.* **1987**, *164*, 421-426.
- ⁴⁰ Anding, C.; Brandt, R. D.; Ourisson, G. *Eur. J. Biochem.* **1971**, *24*, 259-263.
- ⁴¹ Tsai, L. B.; Patterson, G. W. *Phytochemistry* **1976**, *15*, 1131-1133.
- ⁴² Hewlins, M. J.; Ehrhardt, J. D.; Hirth, L.; Ourisson, G. *Eur. J. Biochem.* **1969**, *8*, 184-188.
- ⁴³ Russell, P. T.; Van Aller, R. T.; Nes, W. R. *J. Biol. Chem.* **1967**, *242*, 5802-5806.
- ⁴⁴ Baisted, D. J.; Gardner, R. L.; McReynolds, L. A. *Phytochemistry* **1968**, *7*, 945-949.
- ⁴⁵ Giner, J. L.; Djerassi, C. *Phytochemistry* **1995**, *39*, 333-335.
- ⁴⁶ Mimaki, Y.; Ori, K.; Sashida, Y.; Nikaido, T.; Song, L. G.; Ohmoto, T. *Chem. Lett.* **1992**, *10*, 1999-2000.
- ⁴⁷ Mimaki, Y.; Kubo, S.; Kinoshita, Y.; Sashida, Y.; Song, L. G.; Nikaido, T.; Ohmoto, T. *Phytochemistry* **1993**, *34*, 791-797.
- ⁴⁸ Kuroda, M.; Mimaki, Y.; Ori, K.; Koshino, H.; Nukada, T.; Sakagami, H.; Sashida, Y. *Tetrahedron* **2002**, *58*, 6735-6740.
- ⁴⁹ Corey, E. J.; Matsuda, S. P. T.; Bartel, B. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11628-11632.
- ⁵⁰ Corey, E. J.; Matsuda, S. P. T.; Bartel, B. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2211-2215.
- ⁵¹ Shi, Z.; Buntel, C. J.; Griffin, J. H. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 7370-7374.
- ⁵² Meyer, M. M.; Xu, R.; Matsuda, S. P. T. *Org. Lett.* **2002**, *4*, 1395-1398.
- ⁵³ Hoshino, T.; Sato, T. *Chem. Commun.* **1999**, *1999*, 2005-2006.

-
- ⁵⁴ Corey, E. J.; Cheng, H.; Baker, C. H.; Matsuda, S. P. T.; Li, D.; Song, X. *J. Am. Chem. Soc.* **1997**, *119*, 1277-1288.
- ⁵⁵ Corey, E. J.; Cheng, H.; Baker, C. H.; Matsuda, S. P. T.; Li, D.; Song, X. *J. Am. Chem. Soc.* **1997**, *119*, 1289-1296.
- ⁵⁶ Herrera, J. B. R.; Wilson, W. K.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2000**, *122*, 6765-6766.
- ⁵⁷ Segura, M. J. R.; Jackson, B. E.; Matsuda, S. P. T. *Nat. Prod. Rep.* **2003**, *20*, 304-317.
- ⁵⁸ Lodeiro, S. Ph.D. thesis, Rice University, **2000**.
- ⁵⁹ A crystal structure of human lanosterol synthase from Protein Data Bank (1W6K)^{R21} was used for this illustration, visualized with MacPyMol software for educational use (Delano Scientific LLC).
- ⁶⁰ Lodeiro, S.; Segura, M. J. R.; Stahl, M.; Schulz-Gasch, T.; Matsuda, S. P. T. *ChemBioChem* **2004**, *5*, 1581-1585.
- ⁶¹ Matsuda, S. P.; Darr, L. B.; Hart, E. A.; Herrera, J. B.; McCann, K. E.; Meyer, M. M.; Pang, J.; Schepmann, H. G. *Org. Lett.* **2000**, *2*, 2261-2263.
- ⁶² Hart, E. A.; Hua, L.; Darr, L. B.; Wilson, W. K.; Pang, J.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **1999**, *121*, 9887-9888.
- ⁶³ Segura, M. J. R.; Lodeiro, S.; Meyer, M. M.; Patel, A. J.; Matsuda, S. P. T. *Org. Lett.* **2002**, *4*, 4459-4462.
- ⁶⁴ Lodeiro, S.; Schulz-Gasch, T.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2005**, *127*, 14132-14133.
- ⁶⁵ Joubert, B. M.; Hua, L.; Matsuda, S. P. T. *Org. Lett.* **2000**, *2*, 339-341.
- ⁶⁶ Meyer, M. M.; Segura, M. J. R.; Wilson, W. K.; Matsuda, S. P. T. *Angew. Chem. Int.*

Ed. **2000**, *39*, 4090-4092.

- ⁶⁷ Lodeiro, S.; Wilson, W. K.; Shan, H.; Matsuda, S. P. T. *Org. Lett.* **2006**, *8*, 439-442.
- ⁶⁸ (a) Shibuya, M.; Adachi, S.; Ebizuka, Y. *Tetrahedron* **2004**, *60*, 6995-7003. (b)
Protostadienol synthase is another protosteryl-type cyclase that would be involved in secondary metabolism.
- ⁶⁹ Xiong, Q.; Wilson, W. K.; Matsuda, S. P. T. *Angew. Chem. Int. Ed.* **2006**, *45*, 1285-1288.
- ⁷⁰ Fazio, G. C.; Xu, R.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2004**, *126*, 5678-5679.
- ⁷¹ Xiong, Q.; Rocco, F.; Wilson, W. K.; Xu, R.; Ceruti, M.; Matsuda, S. P. T. *J. Org. Chem.* **2005**, *70*, 5362-5375.
- ⁷² Fischer, W. W.; Pearson, A. *Geobiology* **2007**, *5*, 19-34.
- ⁷³ Reference ²¹ provides a review of plant oxidosqualene cyclases and contains a phylogenetic tree with most oxidosqualene cyclases.
- ⁷⁴ Herrera, J. B. R.; Bartel, B.; Wilson, W. K.; Matsuda, S. P. T. *Phytochemistry* **1998**, *49*, 1905-1911.
- ⁷⁵ Shibuya, M.; Zhang, H.; Endo, A.; Shishikura, K.; Kushiro, T.; Ebizuka, Y. *Eur. J. Biochem.* **1999**, *266*, 302-307.
- ⁷⁶ Morita, M.; Shibuya, M.; Kushiro, T.; Masuda, K.; Ebizuka, Y. *Eur. J. Biochem.* **2000**, *267*, 3453-3460.
- ⁷⁷ Iturbe-Ormaetxe, I.; Haralampidis, K.; Papadopoulou, K.; Osbourn, A. E. *Plant Mol. Biol.* **2003**, *51*, 731-743.
- ⁷⁸ Kawano, N.; Ichinose, K.; Ebizuka, Y. *Biol. Pharm. Bull.* **2002**, *25*, 477-482.

-
- ⁷⁹ Basyuni, M.; Oku, H.; Inafuku, M.; Baba, S.; Iwasaki, H.; Oshiro, K.; Okabe, T.; Shibuya, M.; Ebizuka, Y. *Phytochemistry* **2006**, *67*, 2517-2524.
- ⁸⁰ Hayashi, H.; Huang, P.; Inoue, K.; Hiraoka, N.; Ikeshiro, Y.; Yazaki, K.; Tanaka, S.; Kushiro, T.; Shibuya, M.; Ebizuka, Y. *Eur. J. Biochem.* **2001**, *268*, 6311-6317.
- ⁸¹ Tansakul, P.; Shibuya, M.; Kushiro, T.; Ebizuka, Y. *FEBS Lett.* **2006**, *580*, 5143-5149.
- ⁸² Segura, M. J. R.; Meyer, M. M.; Matsuda, S. P. T. *Org. Lett.* **2000**, *2*, 2257-2259.
- ⁸³ Kushiro, T.; Hoshino, M.; Tsutsumi, T.; Kawai, K. I.; Shiro, M.; Shibuya, M.; Ebizuka, Y. *Org. Lett.* **2006**, *8*, 5589-5592.
- ⁸⁴ Kushiro, T.; Shibuya, M.; Masuda, K.; Ebizuka, Y. *Tetrahedron Lett.* **2000**, *41*, 7705-7710.
- ⁸⁵ Ebizuka, Y.; Katsube, Y.; Tsutsumi, T.; Kushiro, T.; Shibuya, M. *Pure Appl. Chem.* **2003**, *75*, 369-374.
- ⁸⁶ Liu, H.; Krizek, J.; Bretscher, A. *Genetics* **1992**, *132*, 665-673.
- ⁸⁷ Nadeau, R. G.; Hanzlik, R. P. *Methods Enzymol.* **1968**, *15*, 346-351.
- ⁸⁸ Corey, E. J.; Matsuda, S. P. T.; Baker, C. H.; Ting, A. Y.; Cheng, H. *Biochem. Biophys. Res. Commun.* **1996**, *219*, 327-331.
- ⁸⁹ Hart, E. A.; Hua, L.; Darr, L. B.; Wilson, W. K.; Pang, J.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **1999**, *121*, 9887-9888.
- ⁹⁰ Joubert B. M.; Hua L.; Matsuda S. P. T. *Org. Lett.* **2000**, *2*, 339-341.
- ⁹¹ Fazio, G. C.; Xu, R.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **2004**, *126*, 5678-5679.
- ⁹² Hua, L. Ph.D. thesis, Rice University, **2000**, p. 96.
- ⁹³ Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**,

-
- 215, 403-410.
- ⁹⁴ The Arabidopsis Genome Initiative, *Nature*. **2000**, *408*, 796-815.
- ⁹⁵ Minet, M.; Dufour, M.-E.; Lacroute, F. *Plant J.* **1992**, *2*, 417-422.
- ⁹⁶ Parts of this chapter were previously published in: Kolesnikova, M. D.; Xiong, Q.; Lodeiro, S.; Hua, L.; Matsuda, S. P. T. *Arch. Biochem. Biophys.*, **2006**, *447*, 87-95.
- ⁹⁷ Emmons, G. T.; Wilson, W. K.; Schroepfer, G. J. J. *Magn. Res. Chem.* **1989**, *27*, 1012-1024.
- ⁹⁸ Adler, J. H.; Young, M.; Nes, W. R. *Lipids* **1977**, *12*, 364-366.
- ⁹⁹ Milon, A., Nakatani, A., Kintzinger, J.P., Ourisson, G. *Helv. Chim. Acta* **1989**, *72*, 1-11.
- ¹⁰⁰ Hart, E. A.; Hua, L.; Darr, L. B.; Wilson, W. K.; Pang, J.; Matsuda, S. P. T. *J. Am. Chem. Soc.* **1999**, *121*, 9887-9888.
- ¹⁰¹ Barrero, A. F.; Alvarez-Manzaneda, E. J.; Alvarez-Manzaneda, R. *Tetrahedron Lett.* **1989**, *30*, 3351-3352.
- ¹⁰² Akihisa, T.; Arai, K.; Kimura, Y.; Koike, K.; Kokke, W. C. M. C.; Shibata, T.; Nikaido, T. *J. Nat. Prod.* **1999**, *62*, 265-268.
- ¹⁰³ Hayashi, H.; Hiraoka, N.; Ikeshiro, Y.; Yazaki, K.; Tanaka, S.; Kushiro, T.; Shibuya, M.; Ebizuka, Y. *Plant Physiol.* **1999**, *121*, 1384.
- ¹⁰⁴ Zhang, H.; Shibuya, M.; Yokota, S.; Ebizuka, Y. *Biol. Pharm. Bull.* **2003**, *26*, 642-650.
- ¹⁰⁵ Morita, M.; Shibuya, M.; Lee, M. S.; Sankawa, U.; Ebizuka, Y. *Biol. Pharm. Bull.* **1997**, *20*, 770-775.

-
- ¹⁰⁶ Hayashi, H.; Hiraoka, N.; Ikeshiro, Y.; Kushiro, T.; Morita, M.; Shibuya, M.; Ebizuka, Y. *Biol. Pharm. Bull.* **2000**, *23*, 231-234.
- ¹⁰⁷ Kushiro, T.; Shibuya, M.; Ebizuka, Y. *Eur. J. Biochem.* **1998**, *256*, 238-44.
- ¹⁰⁸ Kushiro, T.; Shibuya, M.; Ebizuka, Y. In *Towards Natural Medicine Research in the 21st Century, Excerpta Medica International Congress Series 1157*; Ageta, H., Aimi, N., Ebizuka, Y., Fujita, T., Honda, G., Eds.; Elsevier Science: Amsterdam, 1998, p 421-428.
- ¹⁰⁹ Joubert, B. M.; Buckner, F. S.; Matsuda, S. P. T. *Org. Lett.* **2001**, *3*, 1957-1960.
- ¹¹⁰ Schwenk, E.; Alexander, G. J. *Arch. Biochem. Biophys.* **1958**, *76*, 65-74.
- ¹¹¹ Benveniste, P.; Hirth, L.; Ourisson, G. *Phytochemistry* **1966**, *5*, 45-58.
- ¹¹² Goad, L. J.; Hammam, A. S.; Dennis, A.; Goodwin, T. W. *Nature* **1966**, *210*, 1322-1324.
- ¹¹³ Expression data for *Arabidopsis* available at www.genevestigator.ethz.ch
- ¹¹⁴ Suzuki, M.; Xiang, T.; Ohyama, K.; Seki, H.; Saito, K.; Muranaka, T.; Hayashi, H.; Katsube, Y.; Kushiro, T.; Shibuya, M.; Ebizuka, Y. *Plant Cell Physiol.* **2006**, *47*, 565-571.
- ¹¹⁵ Sawai, S.; Akashi, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Ayabe, S.-i.; Aoki, T. *Plant Cell Physiol.* **2006**, *47*, 673-677.
- ¹¹⁶ Parts of this chapter have been previously published in: Kolesnikova, M. D.; Wilson, W. K.; Lynch, D. A.; Obermeyer, A. C.; Matsuda, S. P. T. *Org. Lett.* **2007**, *9*, 5223-5226.
- ¹¹⁷ Zhang, H.; Shibuya, M.; Yokota, S.; Ebizuka, Y. *Biol. Pharm. Bull.* **2003**, *26*, 642-650.

-
- ¹¹⁸ Sawai, S.; Shindo, T.; Sato, S.; Kaneko, T.; Tabata, S.; Ayabe, S.-i.; Aoki, T. *Plant Sci.* **2006**, *170*, 247-257.
- ¹¹⁹ Akihisa, T.; Koike, K.; Kimura, Y.; Sashida, N.; Matsumoto, T.; Ukiya, M.; Nikaido, T. *Lipids* **1999**, *34*, 1151-1157.
- ¹²⁰ Akihisa, T.; Wijeratne, E. M. K.; Tokuda, H.; Enjo, F.; Toriumi, M.; Kimura, Y.; Koike, K.; Nikaido, T.; Tezuka, Y.; Nishino, H. *J. Nat. Prod.* **2002**, *65*, 158-162.
- ¹²¹ Rukachaisirikul, V.; Pailee, P.; Hiranrat, A.; Tuchinda, P.; Yoosook, C.; Kasisit, J.; Taylor, W. C.; Reutrakul, V. *Planta Med.* **2003**, *69*, 1141-1146.
- ¹²² Arai, Y.; Hirohara, M.; Ageta, H.; Hsu, H. Y. *Tetrahedron Lett.* **1992**, *33*, 1325-1328.
- ¹²³ Barrero, A. F.; Haidour, A.; Munoz-Dorado, M.; Akssira, M.; Sedqui, A.; Mansour, I. *Phytochemistry* **1998**, *48*, 1237-1240.
- ¹²⁴ Barrero, A. F.; Alvarez-Manzaneda, E. J.; Herrador, M. M.; Alvarez-Manzaneda, R.; Quilez, J.; Chahboun, R.; Linares, P.; Rivas, A. *Tetrahedron Lett.* **1999**, *40*, 8273-8276.
- ¹²⁵ Data available from: <http://mpss.udel.edu/at/>
- ¹²⁶ The mRNA was kindly provided by Jeanne Rasbery and Professor Bonnie Bartel (Rice University).
- ¹²⁷ Part of this chapter was previously published in: Kolesnikova, M. D.; Obermeyer, A. C.; Wilson, W. K.; Lynch, D. A.; Xiong, Q.; Matsuda, S. P. T. *Org. Lett.* **2007**, *9*, 2183-2186.
- ¹²⁸ The name and structure of arabidiol without the C-14 stereochemistry determined was published in the paper: Xiang, T.; Shibuya, M.; Katsube, Y.; Tsutsumi, T.; Otsuka, M.; Zhang, H.; Masuda, K.; Ebizuka, Y. *Org. Lett.* **2006**, *8*, 2835-2838. The authors

did not report any minor products (except arabidiol epoxide, a product of dioxidosqualene cyclization).

- ¹²⁹ Lodeiro, S.; Xiong, Q.; Wilson, W. K.; Kolesnikova, M. D.; Onak, C. S.; Matsuda, S. *P. T. J. Am. Chem. Soc.* **2007**, *129*, 11213-11222.
- ¹³⁰ Kishi, M.; Kato, T.; Kitahara, Y. *Chem. Pharm. Bull.* **1967**, *15*, 1071-1073.
- ¹³¹ Kitahara, Y.; Kato, T.; Kishi, M. *Chem. Pharm. Bull.* **1968**, *16*, 2216-2222.
- ¹³² Krief, A.; Pasau, P.; Guittet, E.; Shan, Y. Y.; Herin, M. *Bioorg. Med. Chem.* **1993**, *3*, 365-368.
- ¹³³ Hoshino, T.; Shimizu, K.; Sato, T. *Angew. Chem., Int. Ed.* **2004**, *43*, 6700-6703.
- ¹³⁴ Abe, I.; Rohmer, M. *J. Chem. Soc., Perkin Trans. 1* **1994**, 783-791.
- ¹³⁵ Nguyen, L. H. D.; Harrison, L. J. *Phytochemistry* **1998**, *50*, 471-476.
- ¹³⁶ Bennett, G. J.; Harrison, L. J.; Sia, G.-L.; Sim, K.-Y. *Phytochemistry* **1993**, *32*, 3822-3823.
- ¹³⁷ Justicia, J.; Rosales, A.; Bunuel, E.; Oller-Lopez Juan, L.; Valdivia, M.; Haidour, A.; Oltra, J. E.; Barrero Alejandro, F.; Cardenas Diego, J.; Cuerva Juan, M. *Chem. Eur. J.* **2004**, *10*, 1778-1788.
- ¹³⁸ Hoshino, T.; Kouda, M.; Abe, T.; Sato, T. *Chem. Commun.* **2000**, 1485-1486.
- ¹³⁹ Arai, Y.; Hirohara, M.; Ageta, H. *Tetrahedron Lett.* **1989**, *30*, 7209-7212.
- ¹⁴⁰ Itoh, T.; Tamura, T.; Matsumoto, T. *Lipids* **1976**, *11*, 434-441.
- ¹⁴¹ Polonsky, J.; Varon, Z.; Rabanal, R. M.; Jacquemin, H. *Israel J. Chem.* **1977**, *16*, 16-19.
- ¹⁴² Akihisa, T.; Kimura, Y.; Tamura, T. *Phytochemistry* **1994**, *37*, 1413-1415.

-
- ¹⁴³ Abe, I.; Rohmer, M. *J. Chem. Soc., Perkin Trans.* **1994**, *1*, 783-791.
- ¹⁴⁴ Pale-Grosdemange, C.; Feil, C.; Rohmer, M.; Poralla, K. *Angew. Chem. Int. Ed.* **1998**, *37*, 2237-2240.
- ¹⁴⁵ Kajikawa, M.; Yamato, K. T.; Fukuzawa, H.; Sakai, Y.; Uchida, H.; Ohyama, K. *Phytochemistry* **2005**, *66*, 1759-1766.
- ¹⁴⁶ Molecular modeling was done with Gaussian software: Frisch, M. et al. *Gaussian 03*, revision D.01; Gaussian, Inc.: Wallingford, CT, 2005.
- ¹⁴⁷ Hoshino, T.; Nakano, S.; Kondo, T.; Sato, T.; Miyoshi, A. *Org. Biomol. Chem.* **2004**, *2*, 1456-1470.
- ¹⁴⁸ Hoshino, T.; Sato, T. *Chem. Commun.* **2002**, 291-301.