ELUCIDATION OF THE TETRATERPENE HYDROCARBON BIOSYNTHETIC PATHWAY IN THE GREEN MICROALGA *BOTRYOCOCCUS BRAUNII* RACE L

A Dissertation

by

HEM RAJ THAPA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Timothy P. Devarenne
Committee Members,	Arum Han
	John E. Mullet
	James C. Sacchettini
Head of Department,	Gregory D. Reinhart

August 2017

Major Subject: Biochemistry

Copyright 2017 Hem Raj Thapa

ABSTRACT

The colony-forming green microalga *Botryococcus braunii* is a potential source of biofuel feedstocks as it produces large amount of liquid hydrocarbon oils that can be converted into combustion engine fuels. There are three different races of *B. braunii* based on the hydrocarbons it synthesizes. Race A produces fatty acid derived alkadienes and alkatrienes, race B produces the triterpenoid hydrocarbons tetramethylsqualene and botryococcenes, and race L, the focus of this study, produces the C_{40} tetraterpenoid hydrocarbon lycopadiene via a previously uncharacterized biosynthetic pathway. Structural similarities suggest this pathway follows a biosynthetic mechanism analogous to that of C_{30} squalene. Confirming this hypothesis, the studies presented here identified C_{20} geranylgeranyl diphosphate (GGPP) as a precursor for lycopaoctaene biosynthesis, the first committed intermediate in the production of lycopadiene. Two squalene synthase (SS)-like cDNAs were identified in race L with one encoding a true SS, and the other an enzyme with lycopaoctaene synthase (LOS) activity. Interestingly, LOS utilizes alternative C_{15} and C_{20} prenyl diphosphate substrates to produce combinatorial hybrid hydrocarbons, but almost exclusively utilizes GGPP *in vivo*. This discovery highlights how SS enzyme diversification resulted in the production of specialized tetraterpenoid oils in race L of *B. braunii*.

To understand LOS substrate and product specificity, rational mutagenesis experiments were conducted based on sequence alignments with several SS proteins as

well as a structural comparison with the human SS (HSS) crystal structure. Characterization of the LOS mutants *in vitro* identified Ser276 and Ala288 in the LOS active site as key amino acids responsible for controlling substrate binding, and thus the promiscuity of this enzyme. Mutating these residues to those found in HSS largely converted LOS from lycopaoctaene production to C_{30} squalene production. Furthermore, these studies were confirmed *in vivo* by expressing LOS in *E. coli* cells metabolically engineered to produce high FPP and GGPP levels. These studies also offer insights into tetraterpenoid hydrocarbon metabolism in *B. braunii* and provide a foundation for engineering LOS for robust production of specific hydrocarbons of a desired chain length.

DEDICATION

To my beloved parents (Bharat and Binda Devi Thapa) and my wonderful wife (Merina Dhakal). This work would not have been possible without their love, patience and invaluable support.

To my siblings: Bhesh Raj, Indu, Kalpana and Badal for their love and support.

ACKNOWLEDGEMENTS

My sincere thanks to my PhD advisor Dr. Tim Devarenne for his mentorship. I am grateful for his support and especially to the freedom he has given me to carry out my research projects. I joined his lab with limited knowledge and research experience. His mentorship has changed the way I think about science and has inspired me to become a good scientist.

I would like to thank my committee, Dr. Arum Han, Dr. John Mullet and Dr. James Sacchettini and Dr. Andreas Holzenburg for their time, guidance and support throughout my graduate career.

I am grateful to Dr. Shigeru Okada at The University of Tokyo for his great insights, suggestions and support for my thesis project. I would like to thank Dr. Arum Han and Dr. Hyun Soo Kim for giving me the opportunities to work on their microfluidics projects related to algal research. Also, I would like to thank Dr. James Sacchettini and Su Tang for their collaboration on protein crystallization projects.

I would like to thank my undergraduate research students, Victoria Yell and Tajma Francis for their time and effort on my several research projects. Finally, to past and current members of Devarenne lab – Dr. Joel Gray, Dongyin Su, Dan Browne, Mehmet Tatli, Incheol Yeo and Dr. Edmundo Lozoya-Gloria for their helpful discussions.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professors Dr. Timothy P. Devarenne [advisor], Dr. John E. Mullet and Dr. James C. Sacchettini of the Department of Biochemistry and Biophysics, and Professor Dr. Arum Han of the Department of Electrical & Computer Engineering and the Department of Biomedical Engineering.

Most of the work for the dissertation was completed by the student, under the advisement of Dr. Timothy P. Devarenne of the Department of Biochemistry & Biophysics. The names of the collaborators for chapter II and their contributions are as follows: Dr. Mandar T. Naik of the Department of Biochemistry and Biophysics at Texas A&M University did the NMR analysis on hydrocarbons, Dr. Shigeru Okada and Dr. Kentaro Takada of the School of Agricultural and Life Sciences at The University of Tokyo did the ozonolysis experiments of hydrocarbons, Dr. Istvan Molnar and Dr. Yuquan Xu of the School of Natural Resource and Environment at The University of Arizona assembled the transcriptome of race L and identified the LSS and LOS gene sequences, and Daniel R. Browne of the Department of Biochemistry and Biophysics at Texas A&M University computationally searched for GGR-like sequences and identified GGR-1 and GGR-2 gene sequences from the race L transcriptome database. Collaborators for chapter III are from the Department of Biochemistry and Biophysics at Texas A&M University and their contributions are as follows: Su Tang proposed

mutations of LOS residues and conducted screening of crystallization conditions for the LOS protein, and Dr. James C. Sacchettini advised on the purification and crystallization experiments of LOS enzyme.

Funding Sources

This work was supported by grants NSF-EFRI-PSBR #1240478 and DOE-ARPA-E-PETRO #DE-AR0000203 to Dr. Timothy P. Devarenne; by contract DE-EE0003046 awarded to the National Alliance for Advanced Biofuels and Bioproducts (NAABB) from the U.S. Department of Energy, of which Dr. Timothy P. Devarenne was a member.

TABLE OF CONTENTS

Page

Page

Page

LIST OF FIGURES

Page

Page

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

1.1 Microalgae as a source of biofuel

Microalgae are a promising next-generation resource for biofuel feedstock production, which could ultimately serve as an alternative to petroleum-based transportation fuels[1]. Microalgae are considered superior to oil crops for biofuel production as they provide higher biomass and oil productivity, do not compete with food sources, have shorter life cycle, do not require rich nutrients for growth, and have potential to meet fuel demand without land resource limitation as they can be easily cultivated on a non-arable land[1, 2]. Microalgae can be used to produce fuels in several distinct forms such as conversion of algal biomass by biochemical and thermochemical process to produce bioethanol, biomethane and biohydrogen; use of biomass as a solid fuel to produce heat, steam and electricity; transesterification of crude oils to produce biodiesel; and use of liquid hydrocarbons as feedstocks to produce transportation fuels such as gasoline, diesel and kerosene[2-4]. Depending on the microalgae species, oil types vary from triacylglycerols (TAGs) to hydrocarbons[2, 5, 6]. These oils are carbon rich, their energy content is comparable to that of petroleum, and the fuels derived from these oils have been successfully tested in combustion engines making them a direct substitute for petroleum[2, 7, 8].

This dissertation follows the style of the journal *Biochemistry*

1.2 Challenges to commercial production of algal biofuels

Despite the great potential of microalgae as a source of biofuel, and the extensive amount of research conducted at the laboratory scale, key challenges remain for the sustainable large scale production and commercialization of algal biofuels[2, 9]. Microalgae are photosynthetic organisms that utilize sunlight, water, carbon dioxide (CO_2) and minerals to produce biomass and biomolecules[2, 9]. This means a steady and sufficient supply of water, CO_2 and minerals is required when photosynthetic organisms such as microalgae are used as a production host for biofuel feedstocks[2, 9]. Freely available concentrated CO_2 is required for feasible production of algal biofuel[10]. However, current pilot scale experiments conducted in the laboratory utilizes purchased CO_2[10]. Although researchers have suggested the use of concentrated CO_2 produced from coal plants and cement factories for culturing algae, the amount of available CO_2 from these industries is not sufficient to grow enough algae to meet the demand of current transportation fuels[2]. In case of minerals, algae utilize phosphorus and nitrogen as the principal nutrients for their growth[2]. The current supply of phosphorus is limited, and the nitrogen fixation process from the atmosphere requires large amounts of energy which at present accounts for 1.2% of global energy consumption for making nitrogen fertilizers[2, 11, 12]. The use of existing nitrogen and phosphorus supplies for large scale production of algal biofuel will divert these resources from agricultural crops raising a concerns of food versus fuel[2].

Another major challenge for algal biofuel production is the supply of freshwater[2]. Although the use of marine algae are suggested for large scale algal culture, freshwater

is still required to compensate for evaporative losses and for washing salts from the biomass prior to oil extraction from these cultures[2]. Large scale production of algal biomass to replace petroleum suffers from technology development as currently utilized high scale culture platforms such as raceway ponds have very low productivity[10]. Closed culture systems such as photobioreactors have potential to overcome this problem as they have much higher productivity for algal biomass than raceway ponds, but are costly and require large amounts of energy for operation[2, 13]. This suggests that improved, low cost and energy efficient biomass production technology needs to be developed in the near future[2, 13]. Fossil fuels are cheap and their reserve may not run out anytime soon[6]. Assuming the carbon footprint (the measure of amount of carbon released during production and burning of material) of algal derived biofuel is smaller than petroleum fuel, the price of algal fuel still needs to be comparable with fossil fuels, to make it commercially viable[2, 14]. Unfortunately, the major impediments for commercialization of algal biofuel at present are the sustainable production of a large quantity of algal biomass and at a cheaper price than conventianl petroleum fuels[2]. Despite these problems, the quest for sustainable production of algal biofuel must continue and the challenges must be overcome by continued research[2, 6, 10].

1.3 Approaches for sustainable production of algal biofuels

Sustainable and commercially viable production of algal biofuel may be within reach in the near future assuming algal biology can be exploited to its full potential[2, 9]. To achieve this goal, extensive research on the biology of microalgae is desired as many of the challenges in the field of algal biofuel can be overcome by utilizing the tools of

molecular biology, synthetic biology, and genetic and metabolic engineering[2, 9]. Some of the proof-of-concept experiments have already been tested and many more remain to be conducted[9]. The challenge of increasing algal biomass productivity may be resolved by utilizing our knowledge from agricultural productivity[9]. As we know, the ancestors of present day agricultural crops had very low yields, but this problem has been overcome by millennia of crop plant domestication[9]. Domestication of crops included several factors such as the identification and characterization of genes, understanding the interaction and regulation of genetic networks, and selection of specific traits over time to produce sets of genes and genetic networks that resulted in a desired phenotype[9]. Over a few decades, we have observed a tremendous increase in agricultural productivity such that we might able to sustain the current global population[9]. Although algae are ancestral organisms when compared to crop plants, very little knowledge is available about their domestication process[9]. There is a great need for accelerating this process so the high-biomass-yielding algae species, which has the ability to make meaningful impact on the global carbon cycle and energy production, can be achieved in the near future[9]. Targets of the domestication process can include several aspects of algal biology but not limited to manipulation of genes, genetic networks and pathways involved in production of primary and secondary metabolites, photosynthesis, carbon capture, light capture, stress tolerance and nutrients absorption[9, 15, 16].

Rapid advancements in genome sequencing technology have accelerated sequencing and assembly of algal genomes from multiple algal species[9]. This is important as it will unravel the genetic basis of unique traits observed in individual algae

species[9]. Further understanding of genes and genetic networks responsible for specific traits or phenotypes will be required to improve the productivity of desired algal species[9]. The current approach for increasing the production of lipids in microalgae depends on the redirection of metabolic flux to TAG biosynthesis, and is achieved by applying physiological stress such as nitrogen limitation, temperature change or salinity increase to the algal cultures[2, 17-19]. Lipids found in algae species are not limited to only TAGs but includes diverse carbon rich molecules such as terpenoid hydrocarbons[2, 5]. Detailed characterization and elucidation of various lipid biosynthetic pathways must be pursued as it provides the foundation for manipulation of these pathways to enhance lipid production[2, 17, 20]. For example, engineered versions of lipid biosynthetic enzymes with better catalytic properties can be used, and the enzymes involved in rate limiting steps of pathways can be overexpressed for robust production of the desired lipid molecule[2, 17, 21].

Carbon dioxide supplementation in culture medium increases the biomass production in algae, but sufficient supply of free CO_2 to algae cultures still remains a challenge and may be partially overcome by genetic modifications[2]. Organisms such as cyanobacteria are known to have the mechanism that allows concentrating CO_2 into the cell from their culture medium providing us with the opportunity to engineer this mechanism into the desired algae species[22, 23,24]. Although carbon capture by algae can be enhanced using genetic modifications, technology for efficient transfer of CO_2 from the atmosphere into the culture medium must be developed to increase productivity[2, 25]. Like terrestrial crops, biomass productivity in algae can be enhanced by improving

photosynthetic efficiency[21, 26]. Rubisco and other enzymes of the Calvin cycle that are superior in enzymatic properties can be first explored from several different organisms, and then overexpressed in the desired algal host to improve the photosynthetic conversion efficiency[2, 21]. Another genetic engineering strategy to boost productivity in microalgae can be done by improving light-harvesting efficiency, which means increasing the rate of conversion of light energy into chemical energy[27]. Most microalgae use chlorophyll *a* and *b* for light capture and are able to only absorb the visible portion of light (400-700 nm). This light absorption range can be expanded to 750 nm by coexpression of chlorophyll *d* along with chlorophyll *a* and *b*, thus increasing the number of photons available for photosynthesis[27].

Nutrient limitation for large scale algal culture may be resolved by recycling processes[6]. Since the target biofuel feedstocks are carbon rich molecules, major nutrients such as phosphorus and nitrogen can be recycled once the biomass has been used for extraction of oils[6]. There are other alternative ways to supply the demand of nitrogen and phosphorus[2]. An ideal and sustainable approach to provide nitrogen to microalgae is to engineer them with the capacity to fix atmospheric nitrogen[2]. This is possible and worth pursuing in microalgae as there are examples of nitrogen fixing photosynthetic organism such as cyanobacteria species[28]. Furthermore, nitrogen fixation efficiency in cyanobacteria has been successfully enhanced by utilizing genetic engineering approaches[29]. The current supply of phosphorus can be enhanced by its recovery from animal fecal matter, which means more phosphorus available for algae culture[2]. Recovery of phosphorus and nitrogen from municipal waste water system can also be an

important source of nutrients for algal biomass production[30, 31]. Despite many challenges, economically viable production of algal biofuel could be possible when disciplines such as genetics, molecular biology, synthetic biology, metabolic engineering, technology development and many others are integrated to address the challenges we face today in realizing sustainable production of algal biofuel.

1.4 Biofuel feedstocks from microalgae

Oils produced by microalgae range from TAGs to hydrocarbon molecules, with most of the species accumulating TAGs at a high concentration[32]. The majority of algal biofuel research over the past few decades has focused extensively on the selection of microalgae species accumulating TAGs molecules[32]. TAGs are formed as a result of an ester bond formation between three fatty acids (composition could vary) and a glycerol molecule, and are widespread in plants and animal as they are used as storage lipids by these organisms[32]. In the case of algae, TAG accumulation can be further enhanced when grown under nitrogen starvation conditions[17, 32]. TAGs are transesterified with methanol in the presence of an alkali catalyst to produce fatty acid methyl esters (FAMEs) and glycerol as a byproduct of the reaction[3]. FAMEs are also commonly known as biodiesel. Fatty acid composition affects the oil quality of FAMEs as the properties of saturated and unsaturated fatty acids at room temperature are solid and liquid, respectively[32]. Fluidity is desired for biodiesel, and can be achieved by changing the fatty acid composition of TAGs at the cellular level[32, 33]. Unfortunately, unsaturated fatty acids are prone to oxidization, can form viscous polymers and produce oxygen radicals[33]. Due to the inherent problems of low oxidative stability and cold weather

performance, FAMEs are not considered an ideal substitute for fossil derived diesel fuels[32]. Hydrocarbon fuels are preferred over biodiesel and other biofuels as they are highly compatible with existing petroleum infrastructures and possess superior fuel properties such as high-energy content, low freezing temperature and high ignition stability[32, 34]. In recent years, researchers have shown renewed interest on microalgae species producing hydrocarbons, or perform metabolic engineering in microorganisms for the production of hydrocarbon fuels[34]. In the case of microbial hydrocarbon fuel production, researchers have successfully manipulated fatty acid and isoprenoid pathways to produce short-chain (C_5-C_{12}) and long-chain (C_{10}-C_{23}) hydrocarbons[34]. These are proof-of-concept experiments and large scale feasibility still needs to tested[34]. The approach of using microorganisms such as *E. coli* and yeast as hydrocarbon biofuel production hosts is flawed as these organisms do not have CO_2 fixation capacity and rely on organic carbon for their growth. Thus, renewable production of hydrocarbon biofuel must be pursued in photosynthetic organisms such as microalgae, which either naturally produce hydrocarbons or use metabolic engineering techniques to engineer hydrocarbon production pathways into microalgal host that have the capacity to accumulate large amounts of lipids. As mentioned earlier, only a few hydrocarbon producing algae species have been reported[32]. Green algae such as *Chlorella kessleri*, *Chlorella vulgaris*, *Scenedesmus acutus*, *Scenedesmus acuminatus*, *Scenedesmus obliquus* were reported to produce hydrocarbons such as heptadecane, hexacosane and 1-heptadecene[35]. However, the total hydrocarbon content in these species is very low, ranging from 0.001 to 0.31 % of total dry biomass, making them unrealistic hosts for biofuel production[32, 35]. The

green microalga *Botryococcus braunii* is a promising candidate for the production of renewable fuels mainly due to its ability to produce and accumulate large amounts of liquid hydrocarbon oils[5, 32, 36].

1.5 *Botryococcus braunii*

Botryococcus braunii is a colony forming green microalga where individual cells of the colony are embedded in an extracellular matrix (Figure 1)[37, 38]. It is cosmopolitan in nature and is found in fresh water, brackish lakes, reservoirs, and ponds of all continents except in Antarctica[36]. The climatic regions where *B. braunii* is found include continental, temperate, tropical and alpine zones[36]. Based on the nature of the hydrocarbons synthesized, *B. braunii* is classified into three different races A, B and L (Figure 2)[36]. Race A produces fatty acid derived C_{23}-C_{33} alkadienes and alkatrienes; race B produces triterpenoid hydrocarbons C_{34} tetramethylsqualene as a minor product and C_{30}-C_{37} botryococcenes as a major product; race L, the focus of this study, produces the C_{40} tetraterpene lycopadiene as a predominant hydrocarbon (Figure 2)[5, 39-43]. Recent ultrastructure studies of *B. braunii* show several unique features of this alga when compared to other eukaryotic cells[37, 44]. About 90-95% of the hydrocarbons produced by this alga are secreted into the extracellular matrix that holds the cells of the colony together, and the rest of the hydrocarbons are stored as intracellular oil bodies (Figures 1A & 1B)[44]. The extracellular matrix is a resistant biopolymer made of ether linked hydrocarbon polymers specific to each race, and serves as a depository for the secreted hydrocarbons (Figure 1A)[45-47]. This unique feature may serve a special purpose

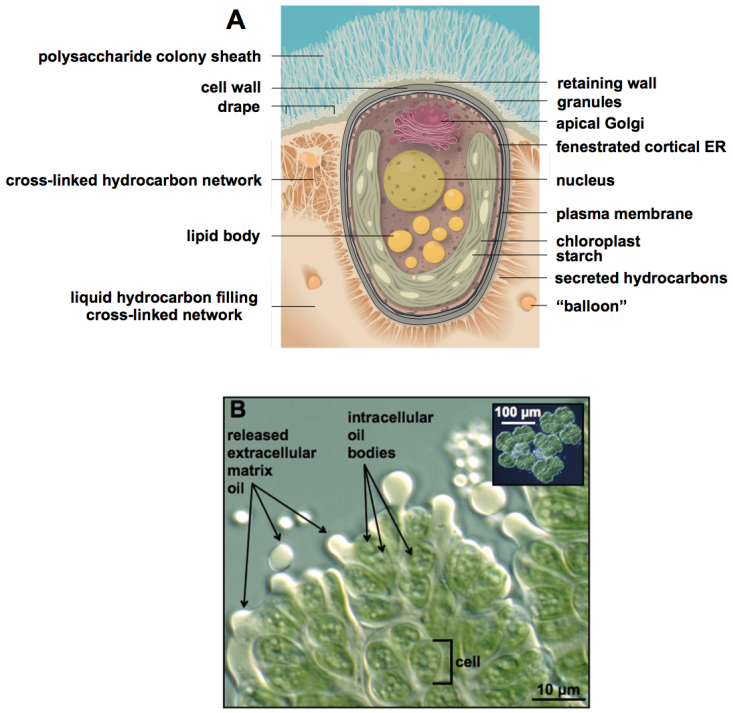


Figure 1. Model and image of *Botryococcus braunii*. A. Model of *B. braunii* colony focused on a single cell (Figure adapted from Weiss TL et al., 2012 *Eukaryotic cell*. **11**:1424-1440). **B.** Transmitted microscopic image *B. braunii* colony when pressure was applied to release oils from extracellular matrix (Figure adapted from Weiss TL et al., 2010 *J. Biol. Chem*. **285**:32458-32466).

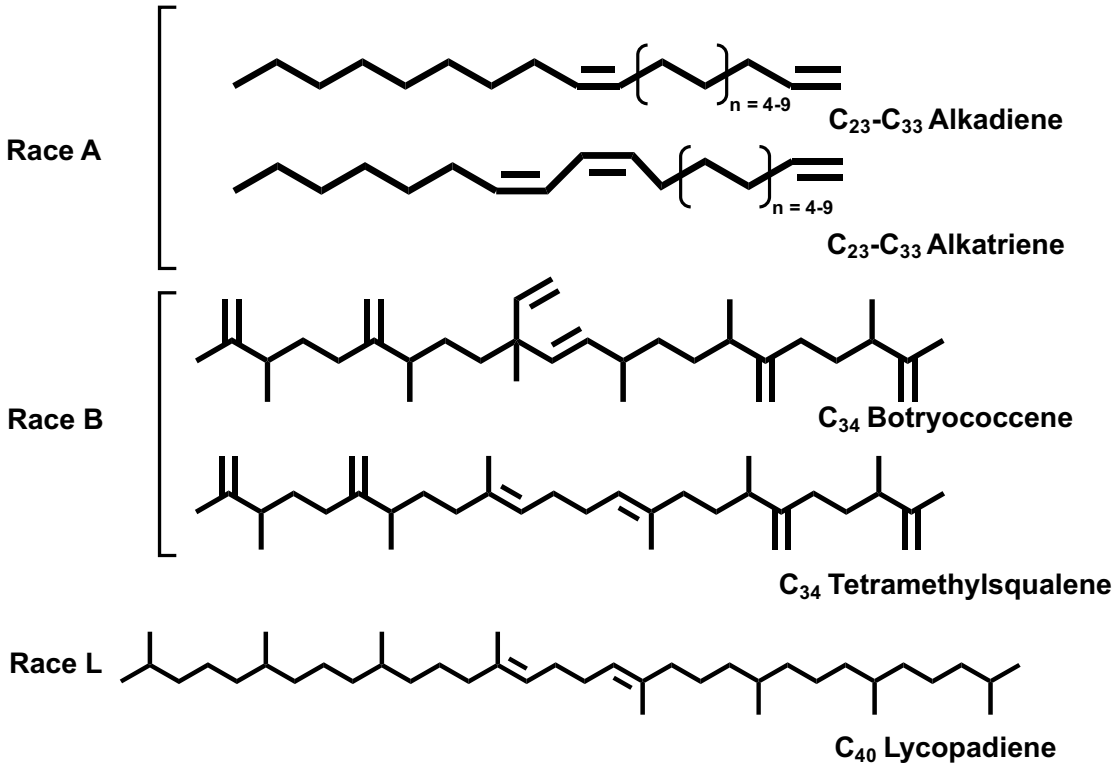


Figure 2. Hydrocarbons produced by three races of *Botryococcus braunii*.

in its natural habitat by allowing the colonies to float and increase exposure to light for a higher rate of photosynthesis[37, 48, 49]. The hydrocarbons stored in the extracellular matrix are sequestered by a retaining wall that surrounds the entire colony (Figure 1A)[37]. The retaining wall can also for a drape that is hypothesized to penetrate into the interior of biopolymer matrix to divide the mother colony into two daughter colonies (Figure 1A)[37]. A polysaccharide colony sheath extends from the retaining wall out toward the growth medium (Figure 1A)[37]. This alga also possesses a unique fenestrated endoplasmic reticulum (ER) that is contiguous to the cell membrane (Figure 1A)[37]. The fenestrae in the ER may act as a secretory pore for the delivery of hydrocarbons to the cell membrane, cell wall and finally to the extracellular matrix[37].

B. braunii has been considered a potential host for the production of renewable fuel for several reasons. First, geochemical evidences show the contribution of all three races of *B. braunii* in the existing coal deposits and oil shales[50-54]. Second, this alga produces prodigious amount of hydrocarbons as liquid oils, the majority of which is secreted into the extracellular matrix[44]. Third, catalytic hydrocracking of hydrocarbons from *B. braunii* can result in fuels that are highly compatible with the existing petroleum infrastructure[4]. For instance, under standard hydrocracking and distillation, C_{34} botryococcene results in 67% gasoline, 15% aviation fuel, 15% diesel fuel and 3% residual oils[4].

Despite the aforementioned advantages of *B. braunii*, its use for the production of biofuel feedstocks faces several challenges. *B. braunii* is a photosynthetic microalgae and requires light, water, CO_2 and inorganic nutrients for its growth. Several studies on

culture conditions show that the hydrocarbon production in this alga is directly related to its growth rather than the requirements for special nutrients[5]. It is a very slow growing alga with a doubling time of 6-7 days, probably due to the commitment of most of its cellular energy to the synthesis of energetically expensive hydrocarbon molecules[55]. To overcome the slow growing nature of this alga, genetic manipulation and use of molecular biology techniques are necessary. Biolistic transformation and preparation of protoplasts by cell wall digestion has not been successful due to the colony organization and presence of the extensive extracellular hydrocarbon matrix[56]. Recently, isolation of single cells from this alga was accomplished by the use of chemical treatments such as glycerol[56]. However, the survival of single cells remains a real challenge as these cells can remain in vital condition for a week only if maintained at a higher concentration (>2 $\times 10^7$cells/mL)[56]. Another aspect of *B. braunii* research is to determine genes involved in the biosynthesis of the above-mentioned hydrocarbons, and to introduce these genes into faster growing heterologous hosts such as the green microalgae *Chlamydomnas reinhardtii* or tobacco plants for the large-scale production of hydrocarbons for biofuel purposes[48, 57-60].

1.6 The isoprenoid pathway

With over more than 55,000 members, isoprenoids (a.k.a. terpenoids) are one of the largest and most structurally diverse family of natural products found in all kingdoms of life, with the majority of isopreonids produced in plant species[61]. Depending on the function, isoprenoids and their derivatives are classified as primary and secondary metabolites[62]. These metabolites play important roles in photosynthesis, respiration,

membrane fluidity, regulation of growth and development, plant-pathogen interactions, and many other cellular functions[62]. They are also used as essential oils, drugs, flavors, colorants, and agrochemicals[62]. A common precursor for the biosynthesis of these isoprenoids is C_5 isopentyl diphosphate (IPP), which contains one isoprene unit[61]. The five-carbon isoprene unit is the building block of isoprenoid compounds, which are made by repetitive joining of several isoprene units[63]. Based on the number of isoprene units present, isoprenoids are classified as monoterpenes (C_{10}), sesquiterpenes (C_{15}), diterpenes (C_{20}), triterpenes (C_{30}), sesterterpenes (C_{35}) or tetraterpenes (C_{40})[64]. Isopentyl diphosphate (IPP) biosynthesis in living organisms occurs via two different pathways: the mevalonate pathway (MVA) or the 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway[61-63, 65]. The cytosolic localized MVA pathway is present in plants, animals, fungi, archaea and some gram-positive bacteria, whereas the MEP pathway is present in green algae and plants, and is compartmentalized into the plastid[63]. The MEP pathway is also utilized by eubacteria for isoprenoid production, and is localized into their cytosol[66].

Like other green algae, *B. braunii* also utilizes the MEP pathway for biosynthesis of isoprenoids and derivatives[67]. The first step of the MEP pathway begins with condensation of pyruvate and glyceraldehyde-3-phosphate by the 1-deoxy-D-xylulose 5-phosphate synthase (DXS) enzyme to produce 1-deoxy-D-xylulose-5-phosphate (DXP), and is considered a rate limiting step of the pathway (Figure 3)[61]. The DXP molecule is rearranged and reduced to 2-*C*-methyl-D-erythritol 4-phosphate (MEP) by the DXP reductoisomerase (DXR) enzyme (Figure 3)[61]. MEP then undergoes

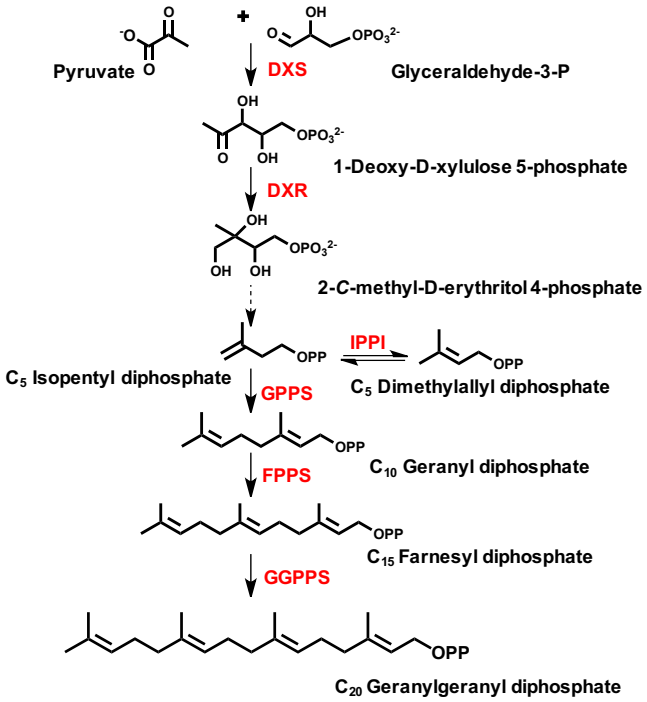


Figure 3. MEP pathway of isoprenoid biosynthesis. DXS = 1-deoxy-D-xylulose 5-phosphate (DXP) synthase, DXR = DXP reductoisomerase, IPPI = Isopentyl diphosphate Δ-isomerase, GPPS = Geranyl diphosphate synthase, FPPS = Farnesyl diphosphate synthase and GGPPS = Geranylgeranyl diphosphate synthase.

conversion in several enzymatic steps to produce a mixture of IPP and dimethylallyl diphosphate (DMAPP)[61]. Equilibrium between IPP and DMAPP is controlled by the IPP Δ-isomerase (IPPI) enzyme (Figure 3)[61]. DMAPP is primarily used as a chemically active substrate for addition to IPP to synthesize higher carbon numbered prenyl diphosphate molecules such as C_{10} geranyl diphosphate (GPP), C_{15} farnesyl diphosphate (FPP) and C_{20} geranylgeranyl diphosphate (GGPP)[63]. DMAPP is also used as a substrate for cytokinin synthesis, and isoprene biosynthesis in plants[63]. Although the reactions of MEP pathway yielding IPP and DMAPP are localized in the plastid of green algae, these molecules or GPP, FPP and GGPP must be exported from the plastid for the biosynthesis of cytosolic isoprenoids such as sterols from FPP[63]. To date, there is a limited information available on the export of prenyl diphosphate precursors from the plastid[63]. GPP is synthesized by GPP synthase (GPPS) as a result of the condensation between DMAPP and IPP molecules (Figure 3). GPP is used as a precursor for monoterpenes such as menthol, camphor and limonene[63]. FPP synthase (FPPS) produces FPP by either utilizing IPP and GPP or two IPP molecules and DMAPP as precursor molecules (Figure 3). FPP is at a major branch point of the isoprenoid pathway and is used as a precursor for the biosynthesis of sesquiterpenes, triterpenes, brassinosteroids, dolichols, oligoprenols and polyprenols[63]. In the case of GGPP synthase, it can utilize all three prenyl diphosphates, FPP, GPP and DMAPP, as substrates to produce the GGPP molecule (Figure 3). Organisms such as plants use GGPP for the production of a variety of isoprenoid molecules such as the tail for chlorophylls, carotenoids, phylloquinone,

plastoquinone, tocopheorols, gibberellins, diterpenes and abscisic acid[63]. Furthermore, FPP and GGPP are also used for protein prenylation[63].

1.7 Squalene synthase

1.7a Squalene

The two chapters (II & III) of this thesis will involve study of squalene synthase (SS)-like genes/enzymes. Thus background on SS is provided in section 1.7 of this chapter. Squalene ($C_{30}H_{50}$), a triterpenoid hydrocarbon, is widely distributed among several different organisms for use as an intermediate in the biosynthesis of cholesterol in mammals, hopanoids in bacteria, ergosterol in fungi, and phytosterols in plants[68]. Squalene is also used as a precursor for the biosynthesis of a wide varieties of triterpene molecules[49]. In addition, squalene has other critical roles such as regulation of membrane fluidity, formation of membrane rafts, and protection against pathogens in host organisms[49, 69]. Squalene was first isolated in 1916 from liver oils of sharks belonging to the family *Squalidae*, and the formal name squalene was given later in 1920[70, 71]. It was characterized as a highly unsaturated hydrocarbon which is colorless, odorless and an oily liquid with a strong ability to reflect light[71]. The absolute amount of squalene varies between organisms, and is found in large quantities in shark liver, oils of human skin, plant seeds, wheat germ, rice bran and plant oils such as olive oil[68, 69, 72]. Currently, squalene is utilized for wide range of applications in food, cosmetics and the pharmaceutical industry[73]. Squalene acts as a natural antioxidant, and is proven to have radioprotective and cardioprotective functions[68]. Squalene also has anti-carcinogenic properties as it successfully inhibits chemically induced tumorigenesis in the colon,

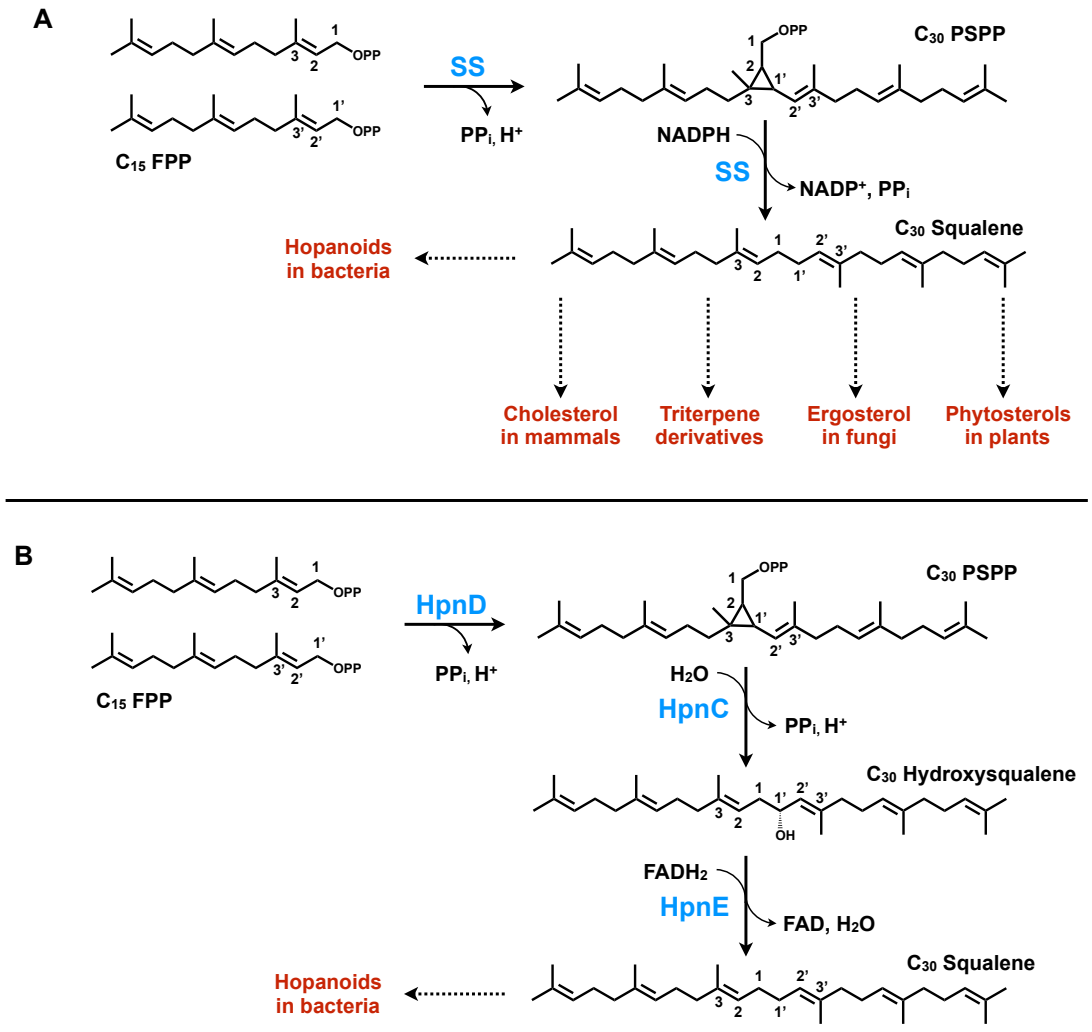


Figure 4. Two pathways for squalene biosynthesis. A. Conversion of two farnesyl diphosphate (FPP) molecules into squalene in a two-step reaction catalyzed by single squalene synthase (SS) enzyme. **B.** A novel three-step squalene pathway catalyzed by three separate enzymes in bacteria.

lungs and skin[68]. Due to its potential for increasing immune responses, squalene has been widely used as a component of flu vaccines[68]. It is also used as a moisturizer in cosmetics and a surfactant in dry cleaning[69]. Because of its highly branched structure, squalene is currently being explored as a potential biofuel feedstock molecule for production of high-octane gasoline[74, 75]. In fact, the catalytic cracking of the squalene molecule resulted in gasoline range molecules with an octane number of 98[75].

1.7b Squalene biosynthesis

Several studies involving squalene biosynthesis in plants, algae, animals, fungi and a few bacteria show a single squalene synthase (SS) enzyme catalyzing a two step reaction[57, 76-79]. The first step of the SS reaction involves condensation of two molecules of C_{15} FPP to form the stable cyclopropyl intermediate presqualene diphosphate (PSPP), and in the second step PSPP undergoes reductive rearrangement in the presence of the cofactor NADPH to form a 1-1' linkage between the two FPP to yield C_{30} squalene (Figure 4A)[48, 80]. The PSPP intermediate is not released from the active site of SS during the two-step reaction. Squalene synthases from several different organisms have been cloned and characterized[57, 76-79]. Consistent with the two-step squalene formation reaction, extensive sequence conservation exists among SS enzymes at the amino acid level, with five conserved domains proposed to play an important role in catalysis[57, 79, 81]. Except for bacteria, all eukaryotic SS enzymes contain a C-terminal membrane spanning domain[57, 78]. Experiments with rat hepatic SS using analytical subcellular fractionation, immunodetection with a monospecific antibody, and immunoelectron microscopy show that the SS protein and its enzyme activity is localized in the ER[82]. Furthermore, deletion

of the C-terminal transmembrane domain resulted in a soluble and active SS enzyme suggesting the C-terminal residues are dispensable for SS activity[83]. These studies suggested the C-terminal transmembrane domain of the SS protein is used for anchoring it to the ER membrane and the remaining large N-terminal region containing the catalytic site is on the cytoplasmic face of the ER[79]. This structure orientation of SS would allow it to take the water soluble substrate FPP and co-factor NADPH from the cytosol, and release the hydrophobic reaction product squalene iinto the membrane of the ER[84]. In another study investigating the localization of the SS protein, the C-terminal transmembrane domain of yeast SS was fused to a reporter gene encoding for the green fluorescent protein[85]. Fluorescence microscopy showed that the fusion construct was localized to the ER, a result that is consistent with the role of the C-terminal transmembrane domain of SS as a signal peptide for targeting the ER membrane[85]. In addition, the fluorescent signal from the fusion construct was also observed in the lipid droplets connecting to the ER[85].

Recently, a new and unanticipated pathway for squalene biosynthesis from farnesyl diphosphate was discovered in the bacterial species *Zymomonas mobilis* and *Rhodopseudomonas palustris* (Figure 4B)[49]. Three genes from the hopanoid biosynthetic gene cluster encoding for HpnC, HpnD and HpnE enzymes were found to constitute a new squalene biosynthetic pathway (Figure 4B)[49]. HpnC and HpnD are squalene synthase-like proteins whereas HpnE encodes for a putative flavoprotein in the amine oxidoreductase family[49]. HpnD is a PSPP synthase and catalyzes the first step of the pathway where two molecules of FPP are converted to PSPP (Figure 4B)[49]. HpnC takes

PSPP as substrate and catalyzes the hydrolytic rearrangement reaction to produce hydroxysqualene (HSQ) in the second step of the pathway (Figure 4B). The third enzyme, HpnE, catalyzes the unusual reduction reaction where the hydroxyl group of HSQ is displaced in an FAD-dependent manner to produce squalene (Figure 4B)[49]. Until now, the squalene biosynthetic enzyme(s) from few bacteria have been characterized[49]. The discovery of a three-step squalene biosynthetic pathway and the conservation of these three genes encoding for HpnC, HpnD and HpnE in the genomes of many Gram positive and Gram negative species suggests the novel squalene pathway might be widespread in bacteria[49].

1.7c Regulation of squalene synthase

Squalene formation is a critical branch point of the isoprenoid pathway where carbon flux from the central pathway is taken towards the biosynthesis of sterols in all eukaryotic organisms and hapanoids in prokaryotes[49, 86]. Sterols are essential to all eukaryotes as they play important roles in a wide range of cellular functions[87]. They are known to maintain the fluidity and permeability of cellular membranes, regulate the activity of membrane bound enzymes, serve as a precursor for steroid hormones such as testosterone and estrogen in mammals, dysteroids in insects, and brassinosteroids in plants[87, 88]. As the SS reaction is the first committed step in sterol biosynthesis, and the FPP substrate is used for production of non-sterol isoprenoid derivatives, SS is considered a potential target for regulation of steroid metabolism[84]. Studies investigating SS regulation in plants shows that SS enzyme activity was found to correlate with *in vivo* biosynthesis of sterols in tobacco[89]. Sterol accumulation in tobacco cell suspension

cultures increased parallel to the fresh weight gain, and was maximal during the rapid growth phase of the culture cycle[89]. When SS activity was measured, it was found to be highest (4- fold increase) during the exponential phase corroborating the carbon committed role of SS in sterol biosynthesis[89]. In another study, sterol accumulation as well as SS enzyme activity rapidly declined when plant cells were treated with fungal elicitors[90]. This rapid inhibition was due to channeling of FPP away from SS and towards the biosynthesis of sesquiterpenoid phytoalexins[90].

Follow-up experiments involving fungal elicitor treatment of tobacco cell suspension cultures showed that SS is regulated at transcriptional, translational and post-translational levels[88]. In this study, time course experiments over 72-hrs were conducted during the rapid growth phase (4-days after subculturing) of tobacco cell suspension cultures[88]. In a control culture, the SS mRNA level was transiently induced 3-fold and remained maximum at 12-hr after the start of the experiment[88]. Induction of mRNA levels corresponded to a 2-fold increase in the level of SS protein and SS enzyme activity by 48-hr[88]. In a parallel experiment where cell cultures were treated with fungal elicitor, SS enzyme activity declined by 70% and the SS mRNA level declined by 80% within the first 14-hr of elicitor treatment, and then remained constant during the rest of 72-hr time course[88]. On the contrary, the SS protein level did not change until 36-hr post-elicitor treatment but started to rapidly decline to undetectable levels by 48-hr[88]. These results suggest that decrease of SS activity after elicitor treatment is not due to the decrease in the level of protein but rather by some post-translational modification of the SS protein[88]. The rapid decline of SS protein after 48-hr could be due to selective

degradation of SS protein in response to elicitor[88]. In the case of mRNA, its regulation in elicitor treated cells could be due to decreased transcription rate of SS gene or post-transcriptional control such as destabilization of SS mRNA[88]. Expression of the tobacco SS promoter fused with the β-glucuronidase (GUS) reporter gene in transgenic plants showed a high-level of GUS reporter expression in shoot meristems and significantly lower expression in other tissues[88]. This suggests the SS gene is primarily expressed in the meristematic regions of the stems[88]. This data was further supported by a significantly higher level of SS protein and enzymatic activity present in the apical meristem and stem when compared to that of root and leaf tissues[88].

Mammalian SS is highly regulated by sterols and cytokines[84, 91]. The level of SS enzymatic activity, protein and mRNA increased rapidly in a cholesterol deficient condition but decreases to very low levels in the presence of excess cholesterol[84]. When hamsters were treated with lipopolysaccharides and cytokines to mimic infection and inflammation, mRNA, protein and enzymatic activity of SS declined rapidly whereas the hydroxymethylglutaryl CoA reductase (HMGR) enzyme activity increased rapidly[91]. The HMGR enzyme catalyzes the rate limiting step of the MVA pathway used for cholesterol biosynthesis[91]. Opposite and rapid changes in SS and HMGR expression support the flux diversion hypothesis where the limited FPP produced by the cells is taken away from squalene biosynthesis towards a non-sterol pathways such as protein prenylation required for signaling during the acute phase response of cells to infection and inflammatory stimuli[84, 91]. Cholesterol homeostasis genes are regulated by binding of the sterol regulatory element binding protein (SREBP) family of transcription factors,

SREBP-1a, SREBP-1c/ADD-1 and SREBP-2, and one or more accessory transcription factor to the promoter of the gene of interest[84]. Analysis of the human SS promoter showed the presence of three SRE-like sequences, HSS-SRE-1, Inv-SRE-3 and SRE-1(8/10), and the binding sites for three of the accessory factors that co-operate with SREBPs in other genes[84]. Characterization of SS promoter elements showed a 50-fold induction of SS promoter activity when cells were deprived of sterols[84]. Rapid promoter activation was a result of the additive effect of HSS-SRE-1 and Inv-SRE-3 whereas SRE-1(8/10) contributed very little to SS promoter activity[84]. When the role of transcription factors was investigated, SREBP-1a activated SS gene transcription by additive interactions with HSS-SRE-1 and Inv-SRE-3, while SREBP-2 activated transcription by binding to all three SRE sequences present in the HSS promoter[84]. Additional experiments found a critical role for SREBP-1a than SREBP-2 in mediating SS gene transcription during sterol deprivation conditions[84].

1.7d Structure-function of squalene synthase

Squalene synthase enzyme catalyzes Mg^{2+} dependent dimerization of two FPP molecules to produce PSPP as an intermediate and squalene as the final product of the reaction (Figure 4A)[79]. Because of its critical role in the biosynthesis of sterol and sterol-like molecules in living organisms, SS is considered a potential therapeutic target to reduce cholesterol related diseases in humans and control infections caused by fungi and parasites[92-94]. Structure-function studies to elucidate the mechanism of the SS reaction have primarily utilized squalene synthase enzymes from yeast, rat and human[81, 92, 95]. Amino acid sequence alignments of eukaryotic SS proteins from mammals (human and

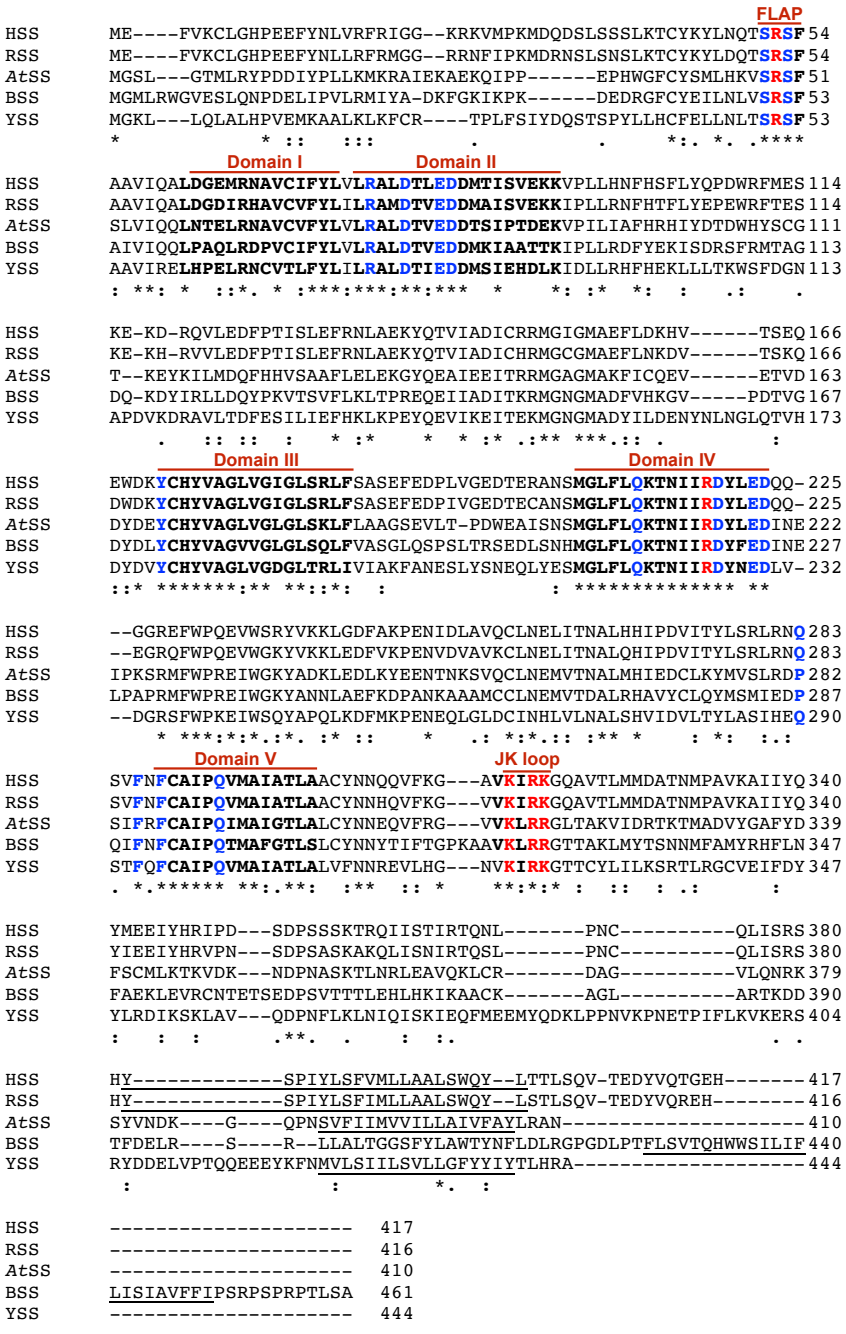


Figure 5. Amino acid sequence alignment of eukaryotic squalene synthase. Conserved domains and regions found in a typical squalene synthase are highlighted in bold and labeled. Transmembrane domain predicted by TMpred for each protein are underlined. HSS = Human SS, RSS = Rat SS, *At*SS = SS from *Arabidopsis thaliana*, BSS = SS from *Botryococcus braunii* race B, and YSS = SS from *Saccharomyces cerevisae.*

rat), plants (*Arabidopsis thaliana*), green algae (*Botryococcus braunii*), and fungi (*Saccharomyces cerevisae*) show conservation of five functional domains (domain I to domain V), a FLAP domain, a JK loop region and the C-terminal transmembrane domain (Figure 5)[76-78, 81, 94]. Site-directed mutagenesis (SDM) experiments, biochemical assays and crystal structures of SS enzymes have shown the importance of all these conserved regions in the two-step SS reaction[81, 92, 94]. Crystal structures of human squalene synthase (HSS) bound to substrate analogs of FPP and the PSPP intermediate, as well as several SS inhibitors have been successfully determined[92-94].

HSS is an α-helical and monomeric protein with a centrally located active site surrounded by the five conserved domains, the FLAP domain and the JK loop region[94]. The crystal structure of HSS with farnesyl thiodiphosphate (FSPP), an analog of FPP, shows two binding sites (S1 and S2) for each FPP substrate molecule (Figure 6A)[94]. The S2 FSPP binds to serine residues (Ser^{51} and Ser^{53}) of the FLAP domain via hydrogen bonding whereas the S1 FSPP binds to two conserved aspartic acid (Asp) rich motifs, **^{80}DxxED^{84}** of domain II and **^{219}DxxED^{223}** of domain IV, and to the conserved Tyr^{171} residue found in domain III (Figure 6A)[94]. Kinetic experiments have suggested a sequential binding mechanism for the two FPP substrates; the acceptor FPP first binds to the S2 site (nonactivating binding site) followed by binding of the donor FPP to the S1 site (activating binding site) (Figure 6A)[95]. Three Mg^{2+} divalent metal ions coordinate the binding of S1 FSPP to the aspartate-rich motifs (Figure 6A)[94]. SDM experiments to mutate Asp^{219} and Asp^{223} to shorter residues, neutral residues and positive residues resulted in total loss of SS activity (Figure 5)[81]. Furthermore, change of these two

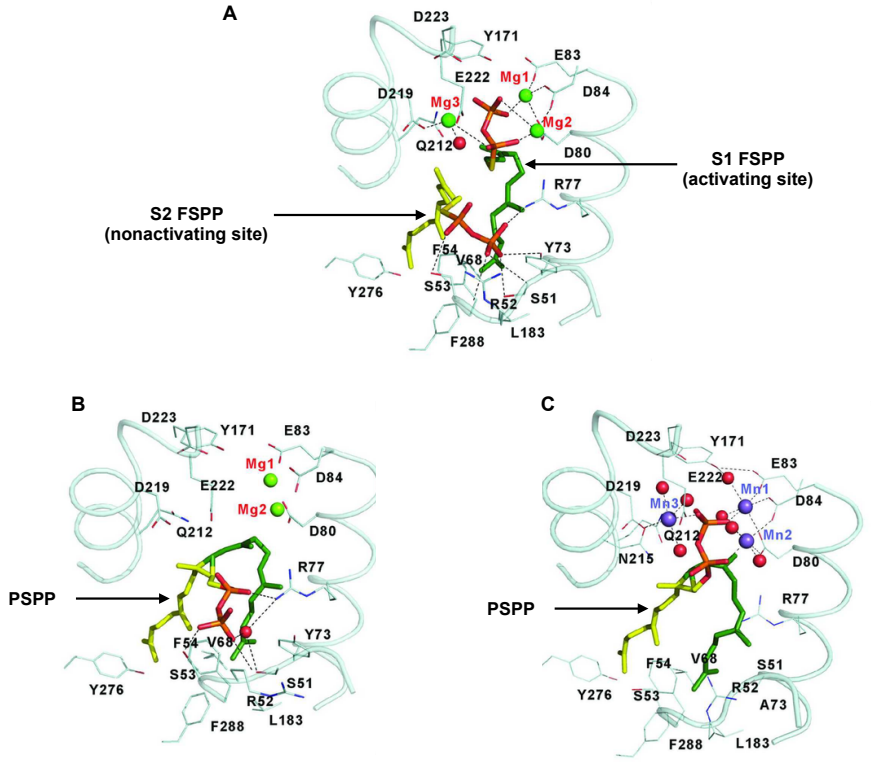


Figure 6. X-ray crystal structures of human squalene synthase (HSS). A. Structure of HSS bound to substrate analog farnesyl thioldiphosphate (FSPP) in the presence of Mg^{2+} metal ion. **B.** Structure of mutant HSS (Y73A) bound to presqualene diphosphate (PSPP) intermediate in the presence of Mg^{2+}. **C.** Structure of HSS bound to PSPP intermediate in the presence of Mn^{2+}. Diphosphate group of FSPP or PSPP is colored in red. Figure A, B and C are adapted from Liu et al., 2014 *Acta Crys*. **D70**:231-241.

aspartic acid residues to either glutamic acid and asparagine did not rescue the enzyme activity suggesting that both the size and acidic nature of Asp^{219} and Asp^{223} is essential for SS reaction[81]. When the SS assay was conducted without Mg^{2+}in the reaction buffer, no enzyme activity was observed[94]. This result is consistent with the critical role of Mg^{2+} in initiating the condensation reaction by ionizing the diphosphate moiety from the S1 FPP substrate to produce an allylic carbocation and diphosphate ion[94]. Tyrosine (Tyr^{171}) found in domain III also interacts with S1 FPP and is found to be highly conserved in all SS proteins (Figures 5 and 6A)[94]. When Tyr^{171} was mutated to phenylalanine, serine, and tryptophan, complete loss of PSPP formation, and squalene formation from PSPP was observed[81]. This suggested an essential role for the phenol group of Tyr^{171} in both steps of the SS reaction[81]. It is hypothesized that the hydroxyl group of Tyr^{171} promotes ionization of diphosphate from S1 FPP and the resulting carbocation is stabilized by the π electrons present in the aromatic ring of tyrosine (Tyr^{171})[81]. In addition, the allylic carbocation could also be stabilized by electrostatic interactions with a magnesium bound diphosphate ion and via the side chain of a glutamine residue (Gln^{212}) present in domain IV (Figures 5 and 6A)[94]. When Gln^{212} was mutated to leucine and asparagine residues, both mutants completely lost SS activity suggesting the critical role of Gln^{212} in SS catalysis[94]. The highly unstable allylic carbocation then undergoes nucleophilic attack by the 2,3-π bond of S2 FPP to produce the cyclopropyl ring containing PSPP intermediate (Figure 4A)[94].

PSPP is converted to squalene without dissociating from the SS enzyme suggesting that the synthesis of PSPP from FPP, and squalene from PSPP takes place in

the same active site of the enzyme[94]. The structure of HSS with PSPP and Mg^{2+} showed the binding of the diphosphate to the serine residues (Ser^{51} and Ser^{53}) of the FLAP domain, and the position of the two hydrophobic tails of PSPP is the same as previously seen for S1 FSPP and S2 FSPP (Figure 6B)[94]. Furthermore, the salt-bridge interaction observed between the α-phosphate group of PSPP and arginine (Arg^{77}) may provide electrostatic interaction to initiate the cleavage of the diphosphate moiety from PSPP (Figure 6B)[94]. However, a structure-function study of dehydrosqualene synthase (CrtM) with PSPP has suggested switching of the diphosphate from the nonactivating site (S2 site) to the activating site (S1 site) in order to initiate the cleavage of diphosphate from PSPP, and subsequent reactions to form the final product (Figures 6A and 6B)[96]. CrtM is an SS-like enzyme that utilizes FPP in a two-step reaction to produce dehydrosqualene, a molecule similar to squalene except for the presence of *cis* double bond at 1-1' linkage between the two FPP molecules, and is used as an intermediate for carotenoid biosynthesis in bacteria[96]. In the CrtM reaction, two FPP molecules are condensed to PSPP in the first step[96]. PSPP then undergoes an NADPH independent rearrangement in the second step to produce dehydrosqualene[96]. As mentioned earlier, opening of the cyclopropyl ring of PSPP in the second step to produce squalene requires NADPH as a reducing cofactor. It was suggested that the SS enzyme undergoes a conformational change in the presence of NADPH to position the diphosphate of PSPP towards the activating site (Figures 6B and 6C)[93, 94]. Mutation of tyrosine (Tyr^{73}) to alanine was made in HSS to allow it tolerate the movement of ligands such as NADPH[94]. The crystal structure of mutant HSS with PSPP and Mn^{2+}, an analogue of biological cofactor Mg^{2+},

showed the binding of the diphosphate moiety of PSPP to the two-conserved Asp rich motifs, **^{80}DxxED^{84}** of domain II and **^{219}DxxED^{223}** of domain IV, via a trio of Mn^{2+} bridges (Figure 6C)[94]. The diphosphate moiety of PSPP is indeed bound to the same binding site as that of S1 FSPP, thus allowing its ionization to initiate the second half reaction (Figures 6A and 6C)[94]. The highly unstable cyclopropylcarbinyl cation formed after cleavage of diphosphate group from PSPP could be stabilized by the π electrons present in the aromatic ring of Phe^{288} found in domain V (Figures 5 and 6C)[94]. Squalene synthase activity was completely lost when Phe^{288} was mutated to either tyrosine, tryptophan, leucine, aspartic acid or arginine residues, suggesting its critical role in catalysis[81]. The crystal structure of HSS showed that Phe^{288} acts as a floor for the S1 site (Figures 6A, 6B and 6C). Furthermore, the crystal structure of HSS (F288A)/PSPP complex showed the movement of the S1 hydrocarbon tail downwards by 4.5Å, and results in the misorientation of the diphosphate group suggesting a stabilization as well structural role for Phe^{288} in the second half reaction[94]. Besides Phe^{288}, residues Phe^{286}, Gln^{283} and Gln^{293} found in or near domain V are also involved in the conversion of PSPP to squalene[81]. Substitution of these three residues with amino acids of different properties retained only partial activity in the second step of the SS reaction (Figure 5)[81].

Reduction of PSPP to squalene in the second step involves hydride transfer from NADPH to the cationic intermediate of PSPP (Figure 4A)[97]. Although not required in the first half reaction, NADPH was shown to stimulate PSPP synthesis by 40-fold[95]. The SS enzyme is unique when compared to other NADPH-dependent enzymes as it does not contain any dinucleotide binding motifs and its structure does not show the presence of a

Rossmann-fold (Figure 6A)[94]. This suggested the SS enzyme might have a unique mechanism for binding NADPH[94]. Positive residues Arg^{52}, Arg^{218}, Lys^{315}, Arg^{317} and Lys^{318} found in FLAP domain, domain II, domain IV and the J-K loop have been identified as critical residues for NADPH binding (Figure 5)[94]. The binding of NADPH by these residues most likely occurs through interaction of the positive charge present on their side chains to the negatively charged phosphate groups of NADPH[94]. Mutation of Arg^{52} to Gln almost completely lost SS activity, and this defect was not rescued in the presence of increasing concentrations of NADPH (Figure 5)[94]. Mutation of Arg^{218} also drastically decreased SS activity but the inhibitory effect was overcome by high concentrations of NADPH (Figure 5)[94]. The double mutants K315E/R317E and K315E/K318E showed a synergistic effect resulting in complete loss of SS activity (Figure 5)[94]. Although the crystal structure of HSS with PSPP and NADPH has not been determined, it seems likely that NADPH binds to the active site via positive residues such that it orients itself to facilitate the hydride transfer to the cationic intermediate of the PSPP molecule[94].

1.8 Hydrocarbon biosynthesis in race A of *B. braunii*

Race A of *B. braunii* produces odd-carbon-numbered *n*-alkadiene and *n*-alkatriene hydrocarbons ranging from C_{23}-C_{33} as its predominant hydrocarbons, and are the products of the fatty acid biosynthetic pathway (Figure 2)[36, 41]. Depending on the strain, total hydrocarbons in race A range from 0.4% to 61% of their dry weight[36]. The structural backbone of *n*-alkadiene and *n*-alkatriene shows similarity to the backbone of the fatty acid C_{18} oleic acid (18:1, *cis*-ω^9) (Figure 7)[98]. Radioactive feeding experiments

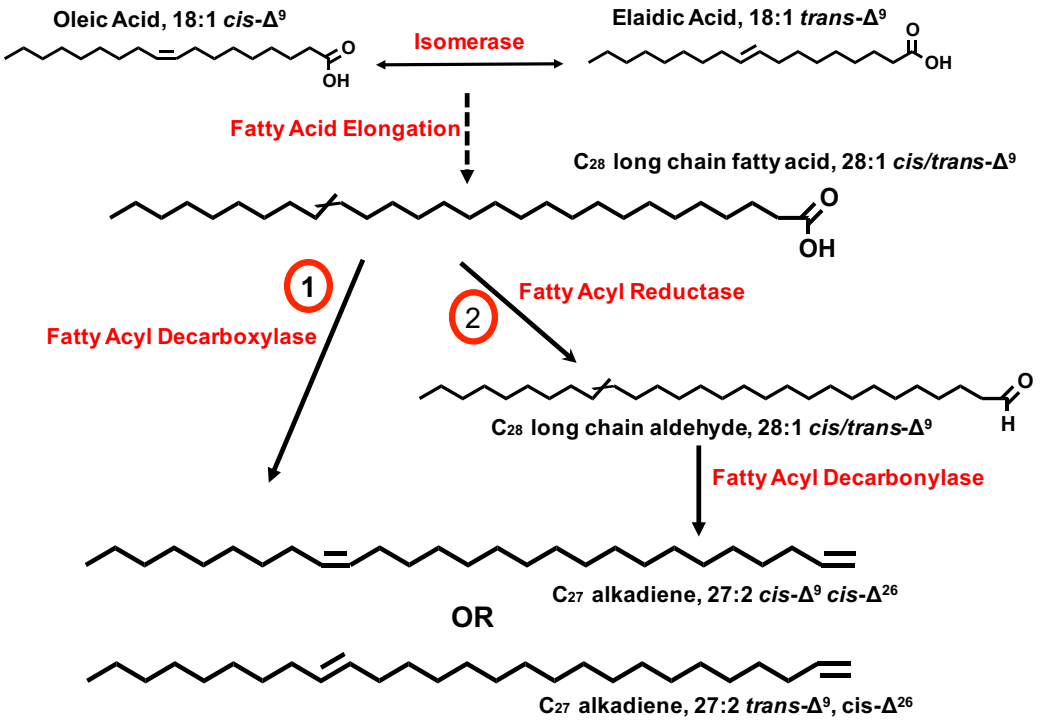


Figure 7. Possible biosynthetic pathways for C_{27} alkadiene in race A. Oleic acid or elaidic acid is elongated to produce C_{28} fatty acid. In possibility one, the elongated C_{28} fatty acid is decarboxylated by fatty acyl decarboxylase enzyme to produce C_{27} alkadiene. In possibility two, the elongated C_{28} fatty acid is first reduced by fatty acyl reductase enzyme to produce aldehyde which is further decarbonylated by fatty acid decarbonylase enzyme to produce C_{27} alkadiene.

indeed show oleic acid as the direct precursor for biosynthesis of alkadienes/alkatrienes[98]. Furthermore, the intracellular concentration of oleic acid decreases as the alkadienes/alkatrienes production increases and vice-versa[98]. A decarboxylation-elongation mechanism for alkadiene/alkatriene production has been proposed as one of the possible biosynthetic routes and needs to be experimentally verified (Figure 7; possibility 1)[98-101]. The C_{18} oleic acid could be first elongated via the fatty acid elongation pathway to produce an elongated fatty acid of desired length which is then decarboxylated by a putative fatty acyl decarboxylase enzyme to produce *cis* double containing alkadienes/alkatrienes (Figure 7)[98-100]. Isomerase activity for conversion of C_{18} oleic acid (18:1, *cis*-ω^9) to C_{18} elaidic acid (18:1, *trans*-ω^9) exists in some strains of race A, and race A can utilize elaidic acid as a precursor to produce trans-alkadienes (Figure7)[100]. For simplification, possible biosynthetic pathways for only C_{27} alkadiene is shown in figure 7. In the case of alkatrienes, the additional double bond could be due to direct desaturation of precursors oleic acid and elaidic acid or from the desaturation of intermediates during the fatty acid elongation process (Figure 2 and 7)[5, 100].

An aternative mechanism for alkadiene/alkatriene production is also possible based on studies conducted with microsomes of race A (Figure 7; possibility 2)[101-104]. A fatty acyl-CoA reductase enzyme was purified from the microsomes, and the enzyme assays showed its ability to produce fatty acid aldehyde when fatty acyl-CoA (palmitoyl-CoA) substrate and NADH cofactor was provided in the reaction[104]. Similarly, a decarbonylase enzyme was also isolated from the race A microsomes[102]. Incubation of

the purified enzyme with aldehyde (octadecanal) substrate showed decarbonylase activity and resulted in the production of an alkane (heptadecane)[102]. However, the fatty acyl-CoA and octadecanal substrates used for reductase and decarbonylase activity, respectively, are not considered natural substrates or intermediates of the alkadiene/alkatriene biosynthetic pathway[36, 102, 104]. According to the alternative mechanism, an elongated fatty acid obtained from oleic acid or elaidic acid could be reduced by the fatty acyl reductase enzyme to produce the corresponding aldehyde intermediate (Figure 7; possibility 2)[101-104]. The aldehyde then undergoes a decarbonylation reaction by fatty acyl decarbonylase to produce *cis* or *trans* double bond containing alkadiene/alkatrine (Figure 7; possibility 2)[101-104].

1.9 Hydrocarbon biosynthesis in race B of *B. braunii*

Among the three races of *B. braunii*, hydrocarbon biosynthesis in race B has been extensively studied and the biosynthetic pathways have been mostly elucidated[48, 58]. Race B produces large amounts of liquid hydrocarbons ranging from 30 to 86% of its cellular dry weight[36]. These hydrocarbons are the product of the MEP pathway, and are the triterpenoids squalene, C_{31}-C_{34} methylated squalene, and C_{30}-C_{37} botryococcene (Figure 8)[5, 36, 67]. Hydrocarbon composition in race B varies depending on the strain characterized and the culture conditions used for growth[36]. The majority of botryococcenes and squalenes produced in race B are methylated homologues C_{32} botryococcene, C_{34} botryococcene and C_{34} tetramethylsqualene whereas C_{31} botryococcene, C_{33} botryococcenes, and C_{31}-C_{33} methylsqualenes account for minor components of the total hydrocarbon pool[48]. Due to the structural similarity of C_{30}

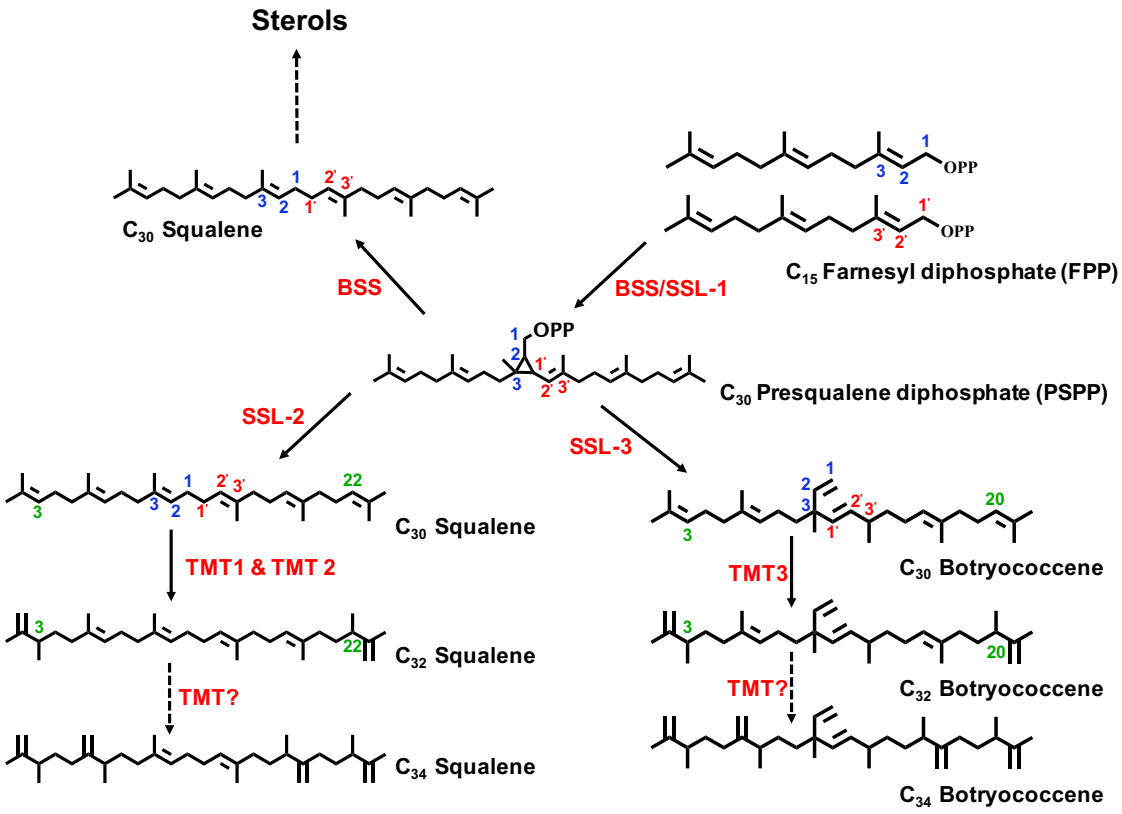


Figure 8. Triterpenoid hydrocarbon biosynthetic pathway in race B. Carbon numbers shown in green indicates the methylation position for the hydrocarbon molecule and the numbering system is different from farnesyl diphosphate molecule. BSS = race B squalene synthase, SSL = squalene synthase-like and TMT = triterpene methyl transferase. Dashed arrow indicated multiple reactions.

botryococcene to C_{30} squalene, its biosynthesis has long been hypothesized to occur via a single SS enzyme or SS-like enzyme(s) (Figure 8)[105]. Extensive *in vitro* and *in vivo* studies have now identified one SS (SS from race B; BSS) and three SS-like (SSL-1, SSL-2 and SSL-3) enzymes involved in the biosynthesis of C_{30} squalene and C_{30} botryococcene, and show a unique mechanism of triterpene production in this alga (Figure 8)[57, 58]. BSS is a typical SS enzyme and catalyzes the two-step reaction for squalene formation discussed in the sections above (Figure 8)[57, 58]. An additional mechanism for squalene formation is observed in race B and the reaction is catalyzed by two SS-like enzymes SSL-1 and SSL-2 (Figure 8)[58]. The SSL-1 enzyme condenses two molecules of FPP into PSPP, which is then utilized by SSL-2 as a substrate and rearranged in the presence of NADPH to produce C_{30} squalene (Figure 8)[58]. C_{30} botryococcene biosynthesis also utilizes two enzymes, SSL-1 and SSL-3 (Figure 8)[58]. The PSPP product formed by SSL-1 is utilized as a substrate by SSL-3 to carry out an alternative rearrangement reaction in the presence of NADPH to form a 1-3' linkage between the two FPP molecules to produce C_{30} botryococcene (Figure 8)[58]. Amino acid sequence analysis of the four triterpene synthases (BSS, SSL-1, SSL-2 & SSL-3) from race B showed conservation of all five domains, the FLAP domain and the J-K loop region found in a typical squalene synthase[58]. Except for BSS and SSL-2, SSL-1 and SSL-3 do not contain a transmembrane domain[57, 58]. However, the localization of botryococcene synthase activity in the membrane fraction of race B suggests that SSL-1 and SSL-3 could be loosely associated to the membrane via interaction with other membrane bound proteins[58, 105]. As shown in Figure 8, there are two mechanisms of

squalene production in race B and each may serve a specific purpose [48, 105]. It has been proposed that squalene produced by conventional squalene synthase (BSS) is involved in housekeeping functions such as sterol biogenesis, whereas the squalene produced using SSL-1 and SSL-2 may be used to synthesize polymethylenesqualene diols that create a cross-linked hydrocarbon network of the extracellular matrix[48, 58, 105]. Three S-adenosyl methionine (SAM) dependent triterpene methyl transferases (TMTs) were identified that are involved in the methylation of C_{30} squalene and C_{30} botryococcene (Figure 8)[48]. TMT-1 and TMT-2 catalyze the methylation of C_{30} squalene at carbon positions 3 and 22 to yield C_{32} dimethylsqualene, whereas TMT3 is responsible for two successive methylations of C_{30} botryococcene at carbon positions 3 and 20 to produce C_{32} botryococcene (Figure 8)[48]. Additional TMTs responsible for further methylation of C_{32} squalene and C_{32} botryococcene into their tetramethylated derivatives remain to be identified (Figure 8).

1.10 Hydrocarbon biosynthesis in race L of *B. braunii*

Race L of *B. braunii* produces tetraterpenoid hydrocarbons derived from the MEP pathway, and total hydrocarbon composition can range from 0.1% to 10% of the cellular dry weight[36, 59, 101]. The C_{40} lycopadiene is the predominant hydrocarbon produced by race L[42, 59]. The biosynthetic pathway for lycopadiene was not known until the recent studies detailed in this thesis[59], and extensive characterization of the lycopadiene pathway is described in chapter II of this thesis.

CHAPTER II

A SQUALENE SYNTHASE-LIKE ENZYME INITIATES PRODUCTION OF TETRATERPENOID HYDROCARBONS IN *BOTRYOCOCCUS BRAUNII* RACE L

2.1 Background and rationale

Microalgae are a promising next-generation source of feedstocks for biofuel production with the potential to serve as a practical alternative to petroleum-based transportation fuels[1]. Depending on the microalgal species, the oils produced vary from triacylglycerols to hydrocarbons[2, 5]. Hydrocarbon-based fuels are preferred over other biofuels as they are highly compatible with existing petroleum infrastructures and possess superior fuel properties[32, 34]. The colony-forming green microalga *Botryococcus braunii* is an exciting candidate for biofuel feedstock production as it produces up to 61% of its dry weight as liquid hydrocarbon oil[36]. These hydrocarbons are produced inside the cells of the colony, seen as intracellular oil bodies, and secreted into the colony extracellular matrix (ECM) where the majority of the hydrocarbons are stored[36] (Figure 9A). Most importantly, catalytic hydrocracking of hydrocarbons from this alga results in petroleum-equivalent fuels of gasoline, kerosene and diesel[4]. Intriguingly,

---

Portions of this chapter have been reprinted as per Creative Commons Attribution 4.0 International License of[59]: (1) **Hem R. Thapa, Mandar T. Naik, Shigeru Okada, Kentaro Takada, István Molnár, Yuquan Xu, Timothy P. Devarenne** (2016). A squalene synthase-like enzyme initiates production of tetraterpenoid hydrocarbons in *Botryococcus braunii* Race L. *Nature Commun*. **7**:11198.

geologic evidence also shows a direct contribution of this alga to the formation of currently utilized fossil fuel deposits around the globe[50-54]. Despite the aforementioned advantages of *B. braunii*, its use for biofuel feedstock production is hindered by a slow growth rate, and the lack of transformation systems to achieve targeted genetic modification[5]. Thus, the identification of *B. braunii* hydrocarbon biosynthetic pathways and associated genes/enzymes can provide options for metabolically engineering these pathways into heterologous hosts with better growth characteristics and the ability to be genetically manipulated. This would then allow the development of improved versions of hydrocarbon biosynthetic enzymes in order to direct production towards the most commercially desirable products[106].

There are three different races of *B. braunii* based on the hydrocarbons synthesized. Race A produces fatty acid-derived C_{23}-C_{33} alkadienes and alkatrienes. Races B and L produce isoprenoid-derived hydrocarbons: methylsqualenes and C_{30}-C_{37} botryococcene triterpenoids in race B, and the C_{40} tetraterpenoid lycopadiene, the focus of this study, in race L (Figure 2)[36]. Green algae have been shown to possess only the plastid-localized methyl erythritol phosphate (MEP) pathway to supply isoprenoid precursors for terpene production[67, 107], and thus lycopadiene is predicted to be generated from C_5 precursors *via* this route (Figure 3). However, the exact mechanism of lycopadiene biosynthesis from C_{20} prenyl diphosphate intermediates has been a mystery and two possible biosynthetic routes have been suggested (Figure 9B)[42]. The first entails C_{20} geranylgeranyl diphosphate (GGPP) reduction by GGPP reductase (GGR) to produce C_{20} phytyl diphosphate (PPP; Figure 9B). Two molecules of PPP would then

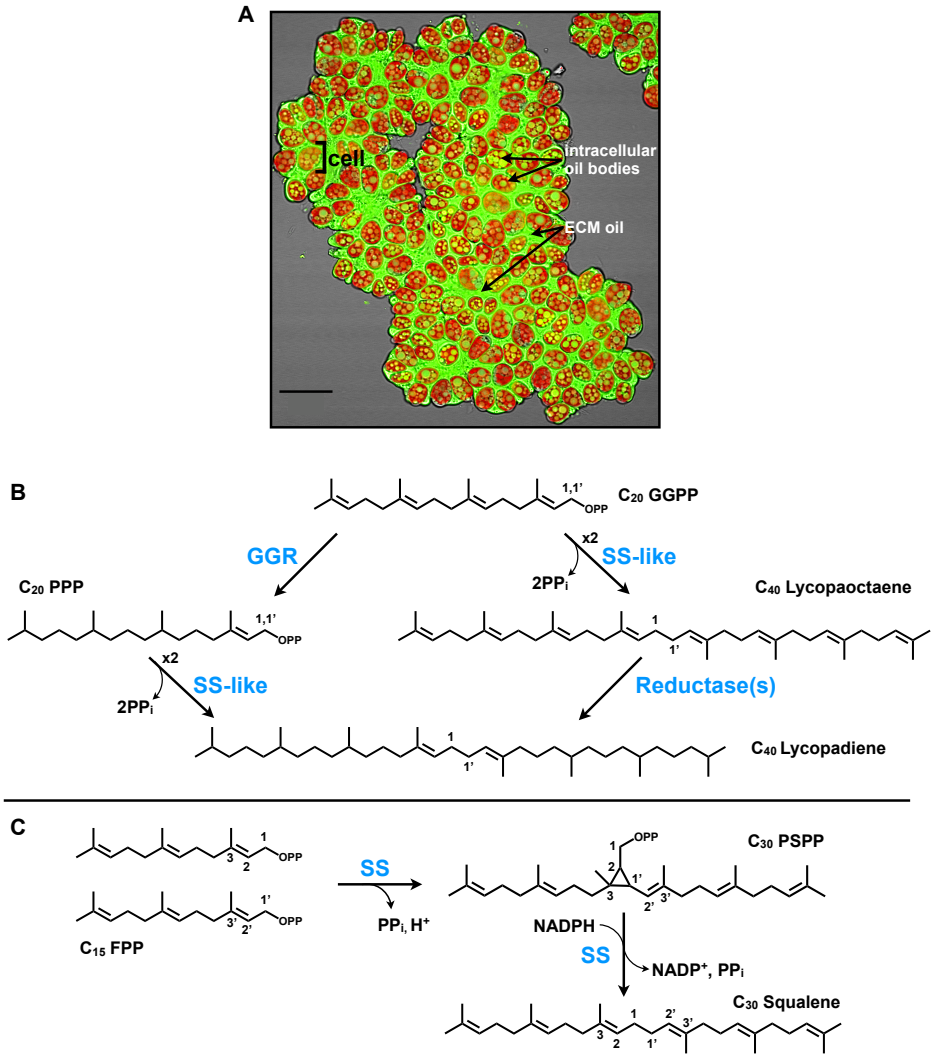


Figure 9. Background information on lycopadiene biosynthesis in race L. A. Confocal microscopy image of an L race colony of *B. braunii*. Shown are hydrocarbon oils stained using the neutral lipid-binding stain Nile red (false colored green), and chlorophyll autofluorescence (red). Indicated are intracellular hydrocarbon oil bodies, hydrocarbon oil in the ECM, and an individual cell as defined by chlorophyll autofluorescence from the single chloroplast in each cell. Image is an overlay of Nile red signal, chlorophyll autofluorescence signal, and a bright-field image. Scale bar = 20 μm. **B**. Two possible pathways for lycopadiene biosynthesis. Possibility 1: GGPP reduction to PPP followed by condensation of two PPP molecules to directly produce lycopadiene. Possibility 2: condensation of two molecules of GGPP to form lycopaoctaene, which would then be reduced to lycopadiene. **C**. Two step reaction catalyzed by single SS enzyme: two molecules of FPP are condensed to squalene through the PSPP intermediate. In (B), and (C) enzyme names are colored blue.

undergo head-to-head condensation (1-1' linkage) to produce lycopadiene (Figure 9B). The second possibility is the head-to-head condensation of two GGPP molecules to produce lycopaoctaene, followed by stepwise enzymatic reduction to produce lycopadiene (Figure 9B).

Using either proposed route, the condensation of PPP or GGPP is predicted to proceed in a reaction mechanism similar to that carried out by the enzyme squalene synthase (SS), which forms C_{30} squalene, a precursor required for sterol biosynthesis in eukaryotes[79] and hopanoids in some prokaryotes[108]. SS enzymes catalyze a two-step reaction. First, the condensation of two C_{15} farnesyl diphosphate (FPP) molecules yields the cyclopropyl intermediate presqualene diphosphate (PSPP; Figure 9C). Second, PSPP undergoes NADPH-dependent reductive rearrangement to form squalene with a 1-1' linkage between the two FPP molecules (Figure 9C)[80]. Herein we report the elucidation of the first committed step in the lycopadiene hydrocarbon biosynthetic pathway in *B. braunii* race L, which is catalyzed by a new SS-like enzyme.

2.2 Results

2.2a L race hydrocarbon content and related enzyme activity

Previous studies reported *trans,trans*-lycopadiene as the predominant hydrocarbon (98% of total hydrocarbons) produced by race L, with a small amount of lycopatriene also documented[42, 43, 109]. However, the lycopatriene structure was not determined. Thus, we first set out to ascertain and refine the hydrocarbon profile of *B. braunii* race L to provide a baseline for our biosynthetic models. Analysis of purified hydrocarbons by GC-MS showed that lycopadiene accounts for 95% of the total

hydrocarbon pool (Figure 10, panel I), with six other minor compounds constituting the remaining 5%: lycopatriene, lycopatetraene, lycopapentaene, lycopapentaene isomer, lycopahexaene, and a $C_{35}H_{64}$ molecule (Figure 10, panels II-VII; Figures 11-14). Ozonolysis experiments suggested lycopatriene and lycopapentaene share an identical reduced C_{20} moiety with lycopadiene (Figures 15-17). NMR spectroscopy was used to confirm identity and structure, including double bond positions, of each molecule (Figure 10, Table 1, Figures 18-25). This is the first report of lycopatetraene, lycopapentaene, and lycopahexaene hydrocarbons from *B. braunii*, race L. The lack of detectable amounts of lycopaoctaene and lycopaheptaene in the total hydrocarbon fraction could be due to rapid conversion of these metabolites into more highly saturated homologues. Interestingly, the unique $C_{35}H_{64}$ isoprenoid hydrocarbon detected contains seven isoprene units, and thus is assigned to the recently named C_{35} terpene class "sesquarterpenes"[110].

In order to gain more insight into the mechanism of lycopadiene biosynthesis and assess the validity of the two possible lycopadiene biosynthetic routes (Figure 9B), a cell-free *in vitro* enzyme assay was developed using total, soluble, and microsomal protein fractions with ^{3}H-labeled PPP and GGPP substrates for potential incorporation into lycopadiene and lycopaoctaene, respectively. Squalene production by SS enzyme activity with ^{3}H-labeled FPP as substrate was used as a control[105]. Lycopadiene production from PPP was not detected in any protein fraction tested, while lycopaoctaene biosynthesis from GGPP was easily observed, mainly in the microsomal fraction (Figure 26A). This indicates a lycopaoctaene synthase (LOS) activity is

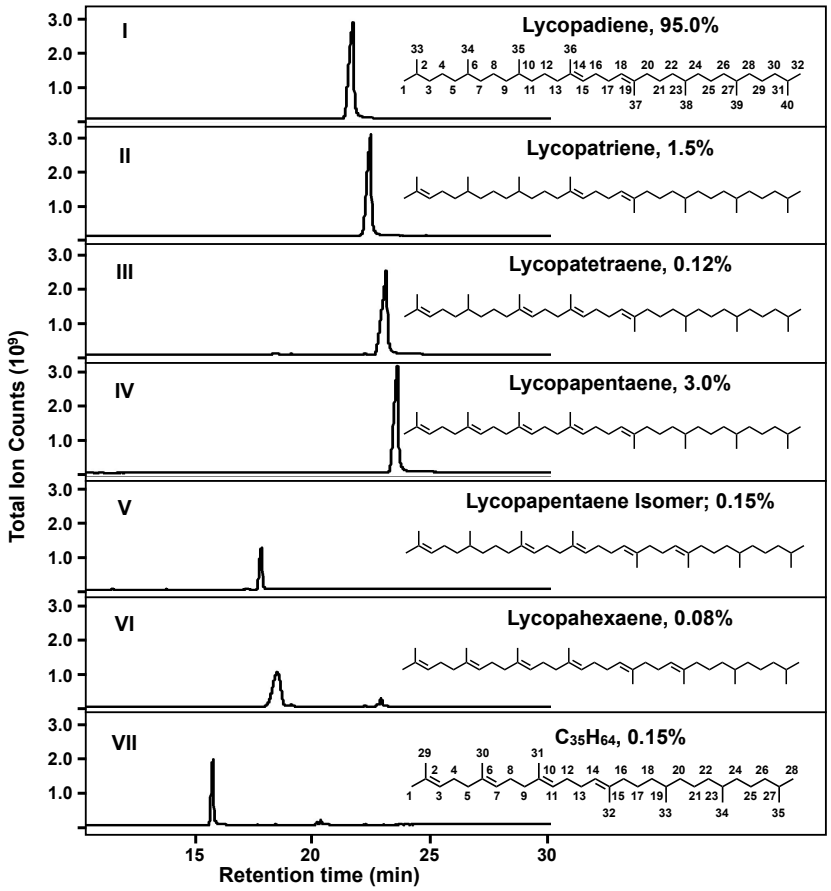


Figure 10. Hydrocarbon composition in *B. braunii* race L. Representative GC-MS profiles of purified hydrocarbons and their corresponding names and structures (panels I - VII). Structures of hydrocarbons were determined by ^{13}C-NMR. Lycopadiene carbon numbering (panel I) is used for numbering all C_{40} hydrocarbons of race L. Percentage next to the name of each hydrocarbon corresponds to its average amount present in total hydrocarbons pool from three independent culture cycles (n = 3).

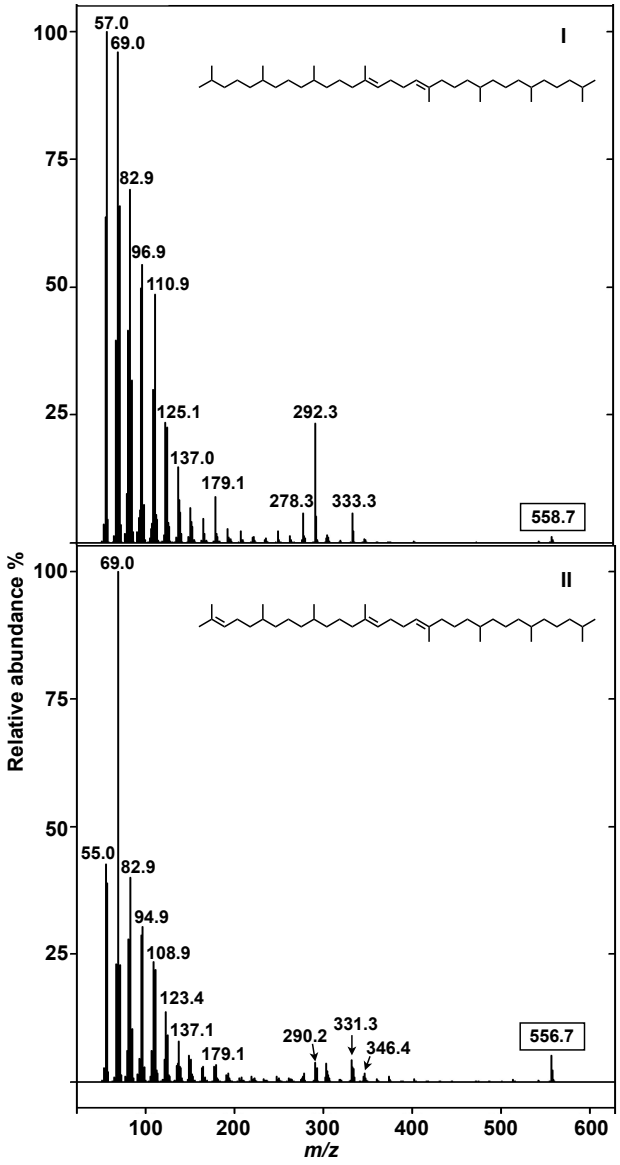


Figure 11. Analysis of lycopadiene and lycopatriene by GC-MS. The mass spectra from GC-MS and structures of lycopadiene and lycopatriene are shown in panel I and panel II, respectively. Molecular ion in box corresponds to the parent ion of each hydrocarbon molecule.

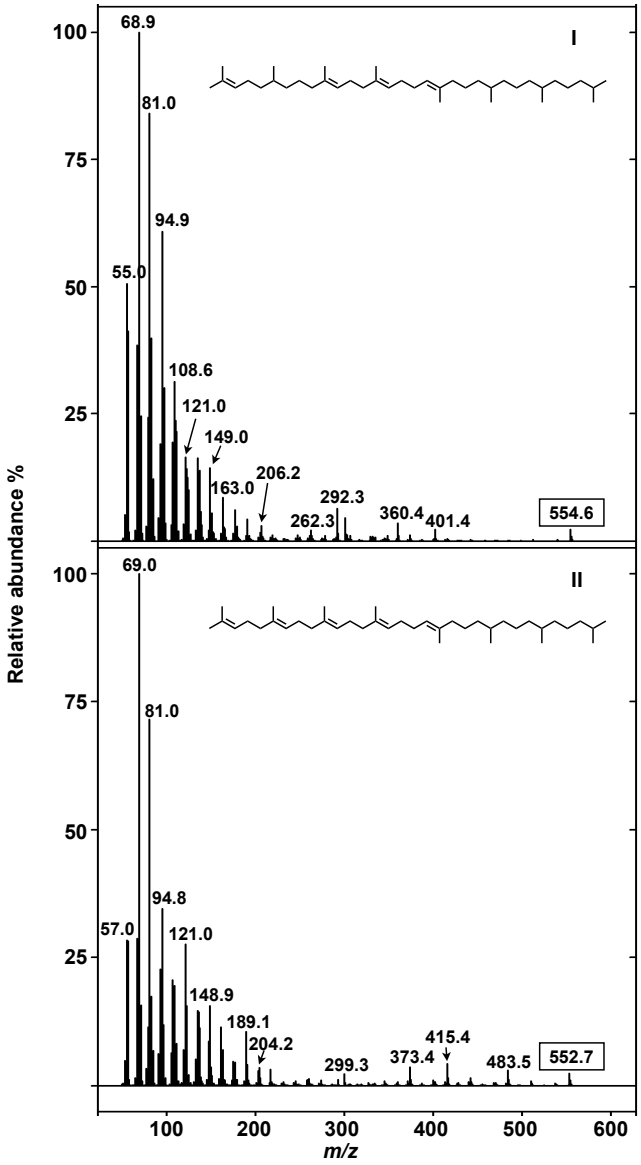


Figure 12. Analysis of lycopatetraene and lycopapentaene by GC-MS. The mass spectra from GC-MS and structures of lycopatetraene and lycopapentaene are shown in panel I and panel II, respectively. Molecular ion in box corresponds to the parent ion of each hydrocarbon molecule.

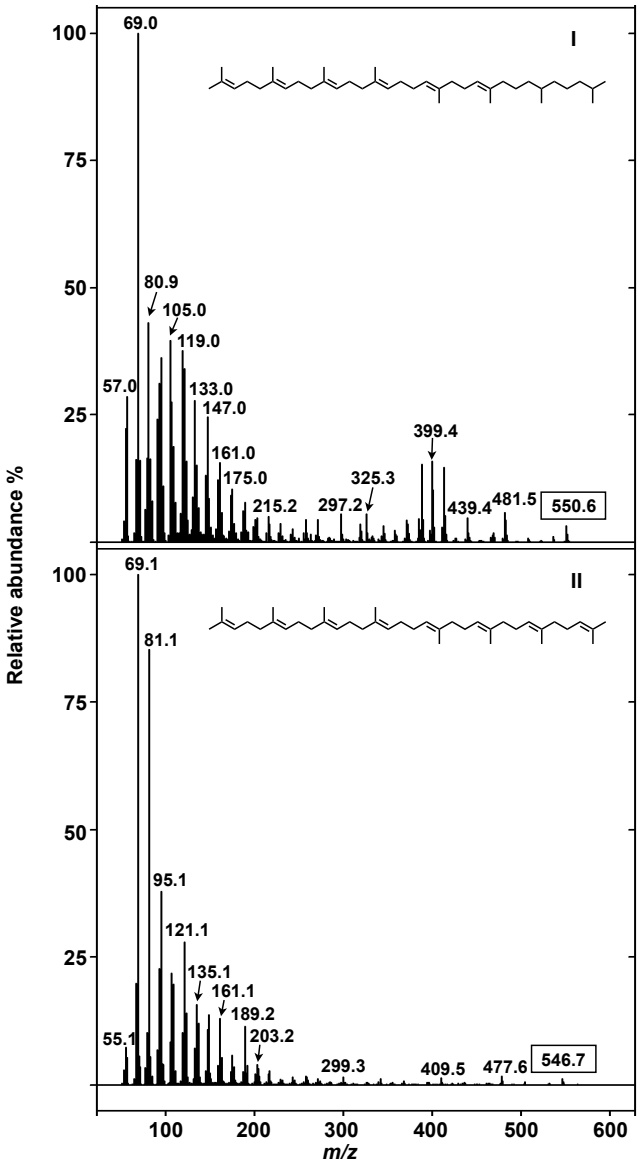


Figure 13. Analysis of lycopahexaene and lycopaoctaene by GC-MS. The mass spectra from GC-MS and structures of lycopahexaene and lycopaoctaene are shown in panel I and panel II, respectively. Molecular ion in box corresponds to the parent ion of each hydrocarbon molecule.

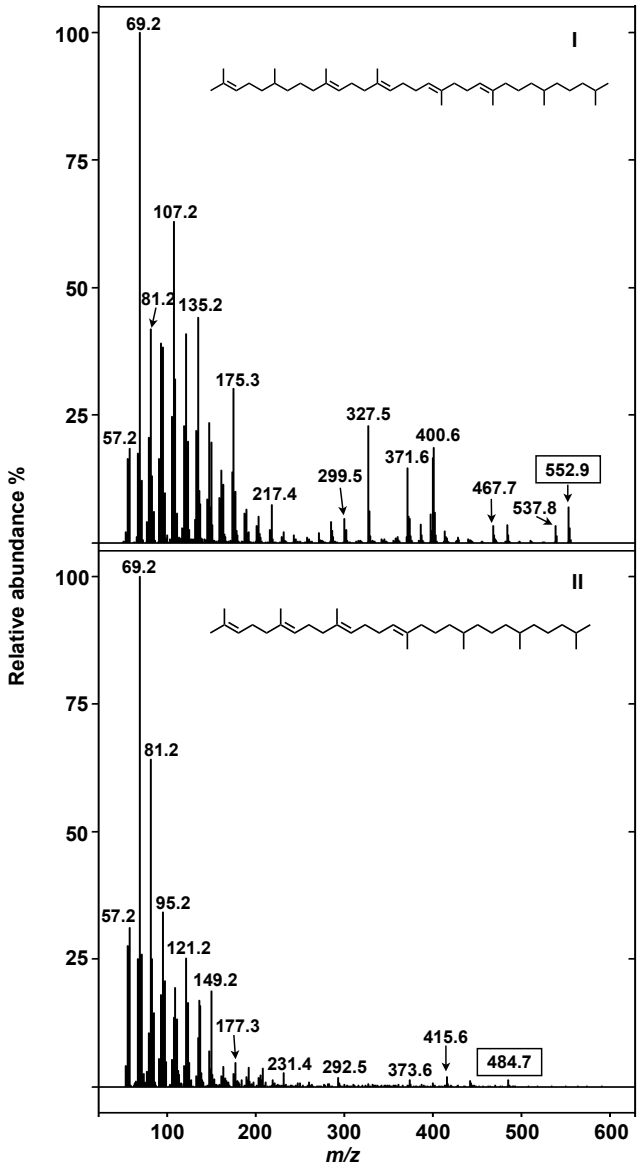


Figure 14. Analysis of lycopapentaene isomer and $C_{35}H_{64}$ by GC-MS. The mass spectra from GC-MS and structures of lycopapentaene isomer and $C_{35}H_{64}$ are shown in panel I and panel II, respectively. Molecular ion in box corresponds to the parent ion of each hydrocarbon molecule.

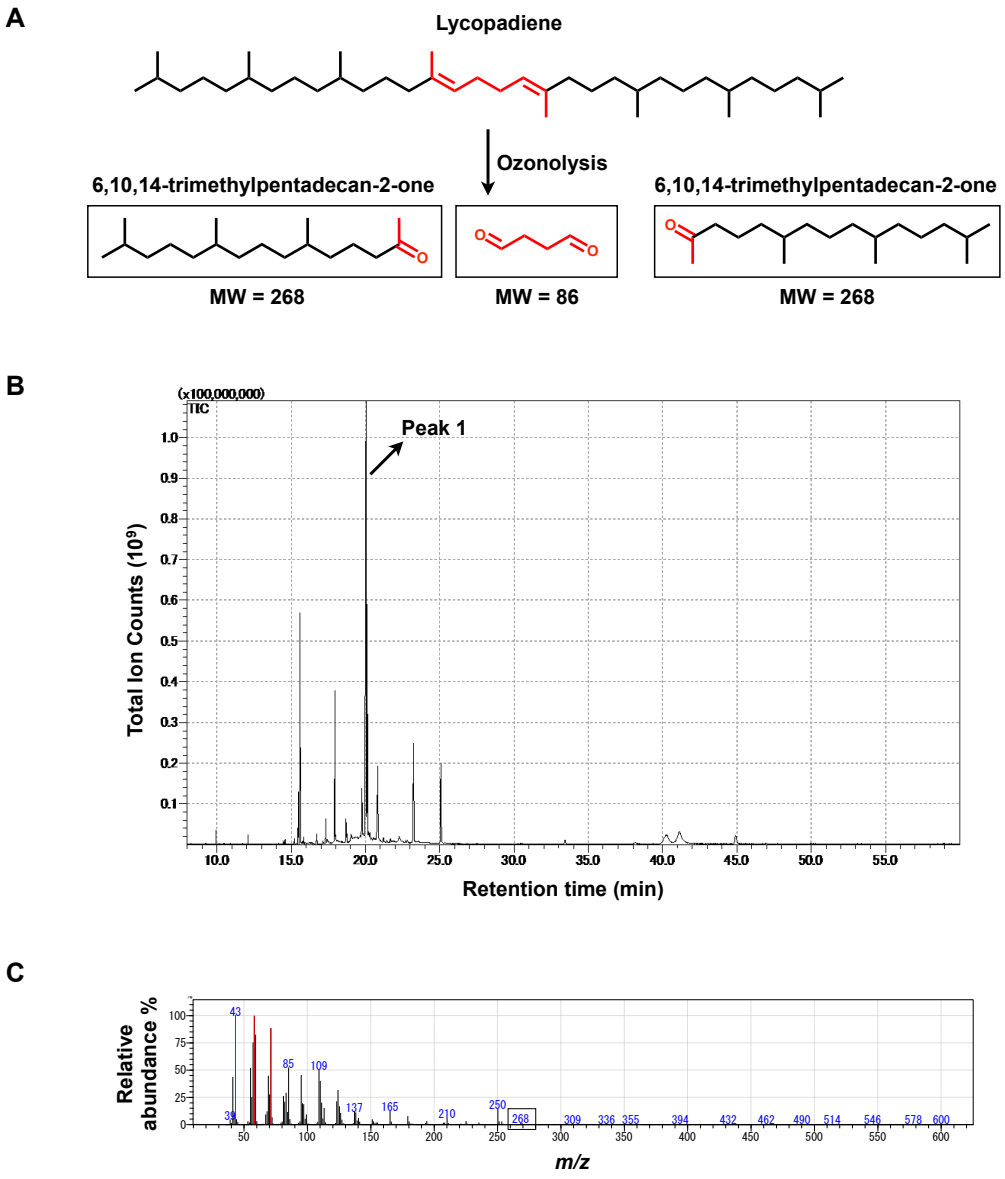


Figure 15. Ozonolysis analysis of lycopadiene. A. Expected ozonolysis products from lycopadiene. **B**. GC-MS analysis of crude ozonolysis products from lycopadiene showed a predominant peak (Peak 1) at 20.04 min. **C**. The mass spectra of Peak 1 showed a parent ion of *m/z* = 268 (in box) and was identified as the expected molecule 6,10,14-trimethylpentadecan-2-one by the Shimadzu mass spectrum database.

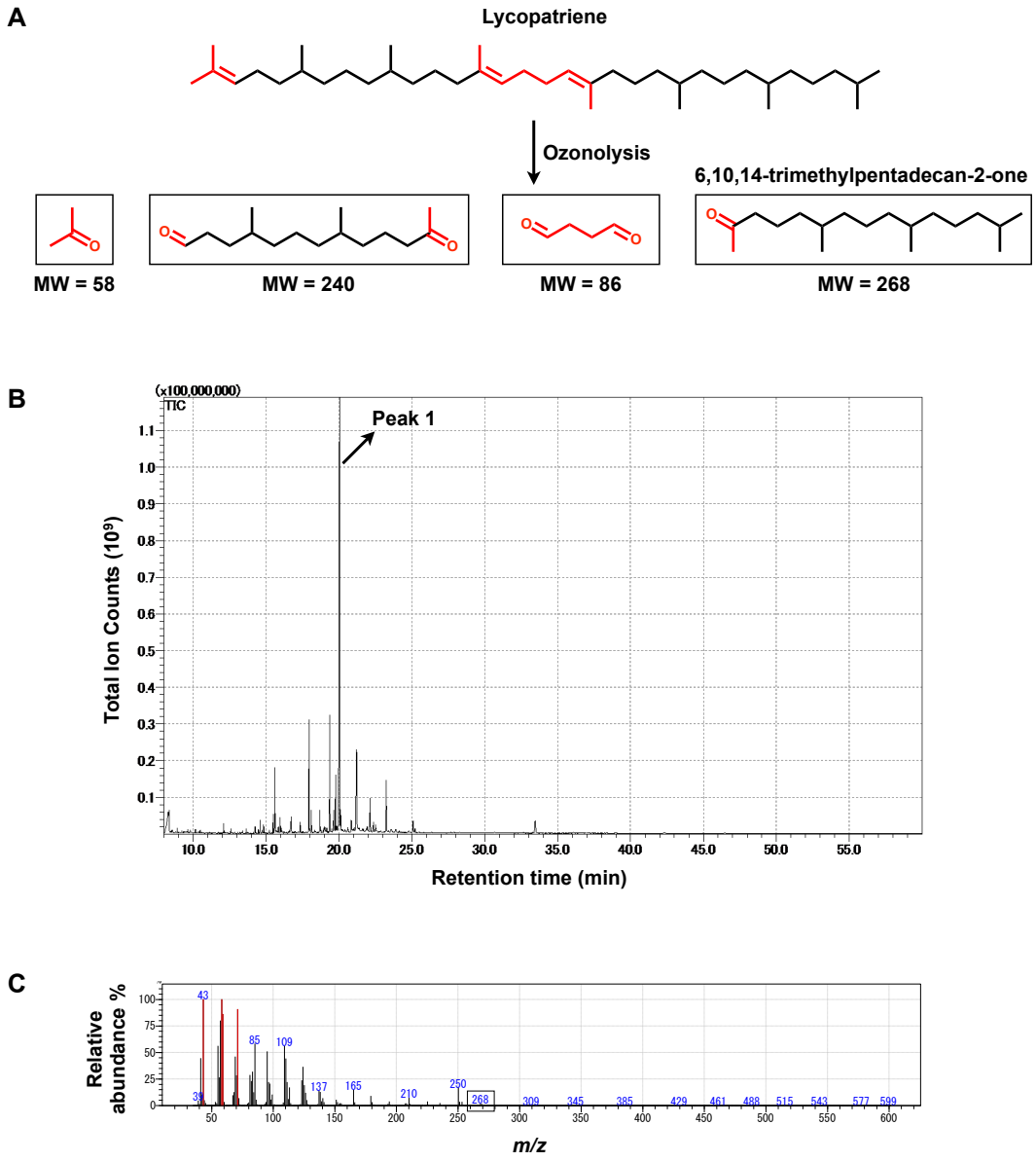


Figure 16. Ozonolysis analysis of lycopatriene. A. Expected ozonolysis products from lycopatriene. **B.** GC-MS analysis of crude ozonolysis products from lycopatriene showed a predominant peak (Peak 1) at 20.04 min. **C.** The mass spectra of Peak 1 showed a parent ion of *m/z* = 268 (in box) and was identified as the expected molecule 6,10,14-trimethylpentadecan-2-one by the Shimadzu mass spectrum database.

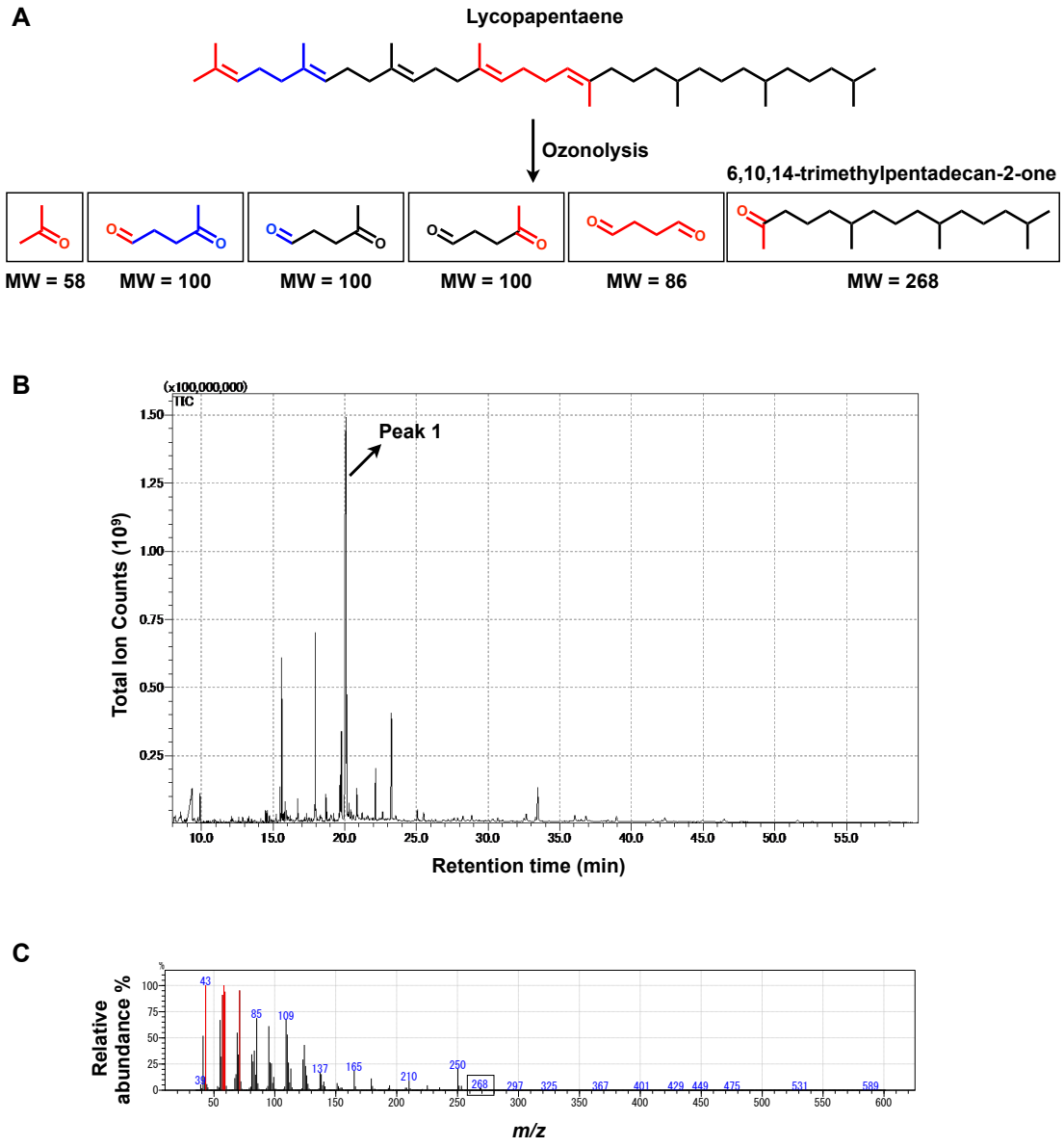


Figure 17. Ozonolysis analysis of lycopapentaene. A. Expected ozonolysis products from lycopapentaene. **B.** GC-MS analysis of crude ozonolysis products from lycopapentaene showed a predominant peak (Peak 1) at 20.04 min. **C.** The mass spectra of Peak 1 showed a parent ion of *m/z* = 268 (ion in box) and was identified as the expected molecule 6,10,14-trimethylpentadecan-2-one by the Shimadzu mass spectrum database.

Carbon	δ ^{13}C (δ ^{1}H) chemical shifts							
	Lycopadiene	Lycopatriene	Lycopatetraene	Lycopapentaene	Lycopapentaene Isomer	Lycopahexaene	Lycopaoctaene	$C_{35}H_{64}$
1	22.7 (0.89)	25.7 (1.71)	25.7 (1.71)	25.7 (1.71)	25.7 (1.71)	25.7 (1.71)	25.7 (1.71)	26.0 (1.68)
2	28.0 (1.54)	131.1	131.1	131.1	131.1	131.1	131.1	131.4
3	39.4 (1.16)	125.1 (5.12)	125.2 (5.12)	124.2 (5.15)	124.8 (5.10)	124.3 (5.15)	124.3 (5.15)	124.8 (5.13)
4	24.7 (1.29)	25.6 (1.98)	25.5 (1.99)	26.7 (2.09)	25.5 (1.99)	26.8 (2.06)	27.2 (2.03)	26.3 (2.01)
5	37.4(1.08, 1.28)	37.1 (1.13, 1.34)	37.1 (1.13, 1.34)	39.8 (2.00)	37.1 (1.13, 1.34)	39.8 (1.99)	39.8 (2.00)	40.0 (1.98)
6	32.8 (1.40)	32.7 (1.40)	32.7 (1.40)	135.9	32.7 (1.40)	135.9	135.9	135.1
7	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	124.2 (5.15)	36.6 (1.08, 1.28)	124.3 (5.15)	124.3 (5.15)	124.8 (5.13)
8	24.5 (1.20, 1.34)	24.5 (1.20, 1.34)	24.5 (1.20, 1.34)	26.7 (2.09)	24.5 (1.20, 1.34)	26.8 (2.06)	27.2 (2.03)	26.3 (2.01)
9	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	39.3 (1.95)	39.8 (2.00)	39.8 (2.00)	39.8 (1.99)	39.8 (2.00)	40.0 (1.98)
10	31.2 (2.17)	30.8 (2.16)	135.9	135.9	135.9	135.9	135.9	135.1
11	36.8 (1.09, 1.27)	36.8 (1.09, 1.27)	124.1 (5.15)	124.2 (5.15)	124.2 (5.15)	124.3 (5.15)	124.3 (5.15)	124.8 (5.13)
12	25.4 (1.39)	25.4 (1.39)	25.5 (1.99)	26.7 (2.09)	26.7 (2.09)	26.8 (2.06)	27.2 (2.03)	27.2 (2.06)
13	40.0 (1.96)	40.0 (1.96)	40.0 (1.96)	39.8 (2.00)	39.8 (2.00)	39.8 (1.99)	39.8 (2.00)	27.2 (2.06)
14	135.9	135.9	135.9	135.9	135.9	135.9	135.9	124.8 (5.13)
15	124.0 (5.15)	124.1 (5.15)	124.1 (5.15)	124.2 (5.15)	124.2 (5.15)	124.3 (5.15)	124.3 (5.15)	135.1
16	28.3 (2.04)	28.3 (2.04)	28.3 (2.04)	28.3 (2.04)	28.3 (2.04)	28.3 (2.04)	28.3 (2.04)	40.0 (1.98)
17	28.3 (2.04)	28.3 (2.04)	28.3 (2.04)	28.3 (2.04)	28.3 (2.04)	28.3 (2.04)	28.3 (2.04)	24.4 (1.62)
18	124.0 (5.15)	124.1 (5.15)	124.1 (5.15)	124.2 (5.15)	124.2 (5.15)	124.3 (5.15)	124.3 (5.15)	37.7 (1.06, 1.24)
19	135.9	135.9	135.9	135.9	135.9	135.9	135.9	33.0 (1.36)
20	40.0 (1.96)	40.0 (1.96)	40.0 (1.96)	39.8 (2.00)	39.8 (2.00)	39.8 (1.99)	39.8 (2.00)	37.7 (1.06, 1.24)
21	25.4 (1.39)	25.4 (1.39)	25.4 (1.39)	25.4 (1.39)	26.7 (2.09)	26.8 (2.06)	27.2 (2.03)	25.1 (1.21)
22	36.8 (1.09, 1.27)	36.8 (1.09, 1.27)	36.6 (1.08, 1.28)	36.6 (1.08, 1.28)	124.2 (5.15)	124.3 (5.15)	124.3 (5.15)	37.7 (1.06, 1.24)
23	31.2 (2.17)	30.8 (2.16)	30.8 (2.16)	31.0 (2.23)	135.9	135.9	135.9	33.0 (1.36)
24	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	39.8 (2.00)	39.4 (1.96)	39.8 (2.00)	37.7 (1.06, 1.24)
25	24.5 (1.20, 1.34)	24.5 (1.20, 1.34)	24.5 (1.20, 1.34)	24.5 (1.20, 1.34)	24.5 (1.20, 1.34)	24.5 (1.20, 1.34)	27.2 (2.03)	24.8 (1.19, 1.32)
26	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	36.6 (1.08, 1.28)	124.3 (5.15)	39.6 (1.13)
27	32.8 (1.40)	32.7 (1.40)	32.7 (1.40)	32.8 (1.40)	32.7 (1.40)	32.8 (1.40)	135.9	28.2 (1.52)
28	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	37.4 (1.08, 1.28)	39.8 (2.00)	22.9 (0.86)
29	24.7 (1.29)	24.7 (1.29)	24.7 (1.29)	24.7 (1.29)	24.7 (1.29)	24.7 (1.29)	27.2 (2.03)	17.9 (1.61)
30	39.4 (1.16)	39.4 (1.16)	39.4 (1.16)	39.4 (1.16)	39.4 (1.16)	39.4 (1.16)	124.3 (5.15)	16.3 (1.61)
31	28.0 (1.54)	28.0 (1.54)	28.0 (1.54)	28.0 (1.55)	28.0 (1.55)	28.0 (1.55)	135.9	16.3 (1.61)
32	22.7 (0.89)	22.7 (0.89)	22.7 (0.89)	22.7 (0.89)	22.7 (0.89)	22.7 (0.89)	25.7 (1.71)	16.3 (1.61)
33	22.7 (0.89)	17.6 (1.63)	17.6 (1.63)	17.7 (1.63)	17.7 (1.63)	17.7 (1.63)	17.7 (1.63)	20.0 (0.83)
34	19.7 (0.87)	19.7 (0.87)	19.7 (0.87)	16.0 (1.62)	19.8 (0.87)	16.0 (1.61)	16.0 (1.62)	20.0 (0.83)
35	19.7 (0.87)	19.7 (0.87)	15.9 (1.61)	16.0 (1.62)	16.0 (1.62)	16.0 (1.61)	16.0 (1.62)	22.9 (0.86)
36	15.9 (1.61)	15.9 (1.61)	15.9 (1.61)	16.0 (1.62)	16.0 (1.62)	16.0 (1.61)	16.0 (1.62)	
37	15.9 (1.61)	15.9 (1.61)	15.9 (1.61)	16.0 (1.62)	16.0 (1.62)	16.0 (1.61)	16.0 (1.62)	
38	19.7 (0.87)	19.7 (0.87)	19.7 (0.87)	19.8 (0.87)	16.0 (1.62)	16.0 (1.61)	16.0 (1.62)	
39	19.7 (0.87)	19.7 (0.87)	19.7 (0.87)	19.8 (0.87)	19.8 (0.87)	19.8 (0.87)	16.0 (1.62)	
40	22.7 (0.89)	22.7 (0.89)	22.7 (0.89)	22.7 (0.89)	22.7 (0.89)	22.7 (0.89)	17.7 (1.63)	

Table 1. NMR assignments for hydrocarbons from race L. ^{13}C and ^{1}H (in parentheses) chemical shifts determined for hydrocarbons from race L. Carbon numbering from 1 to 40 and 1 to 35 is the same as shown for lycopadiene and $C_{35}H_{64}$ in Figure 10, panel I and VII, respectively.

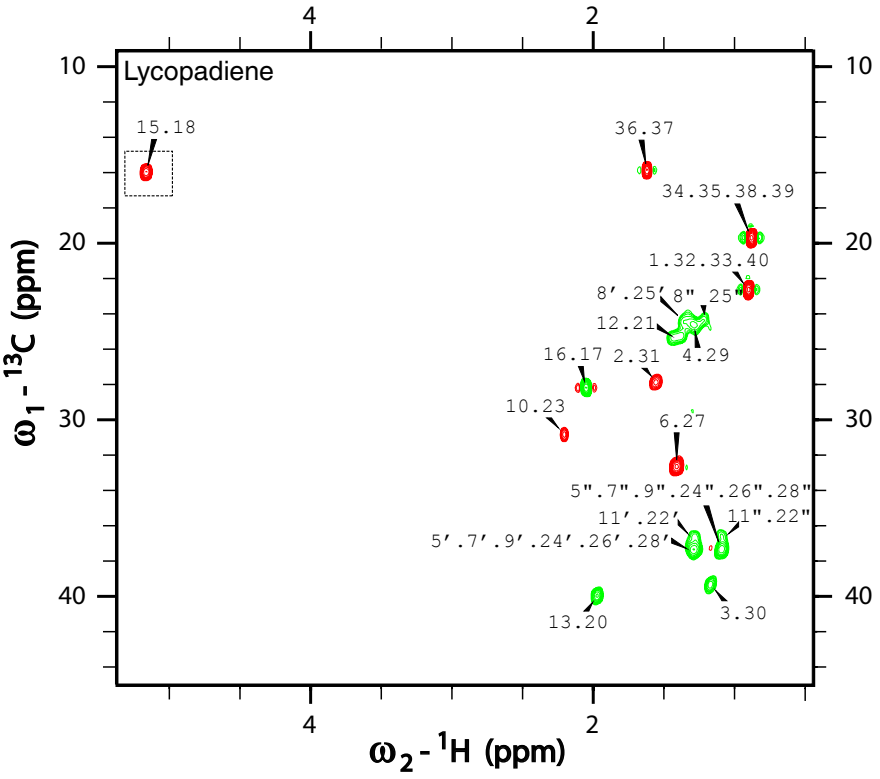


Figure 18. NMR analysis of lycopadiene. 2D-^{13}C-HSQC spectrum of lycopadiene. Positive (red) contours represent -CH or -CH_3 correlations, while negative (green) contours indicate -CH_2. Peaks originating from the molecule of interest are marked with resonance assignment, while those from impurities are left unmarked. Peaks aliased on ^{13}C axis are shown in a dashed line box.

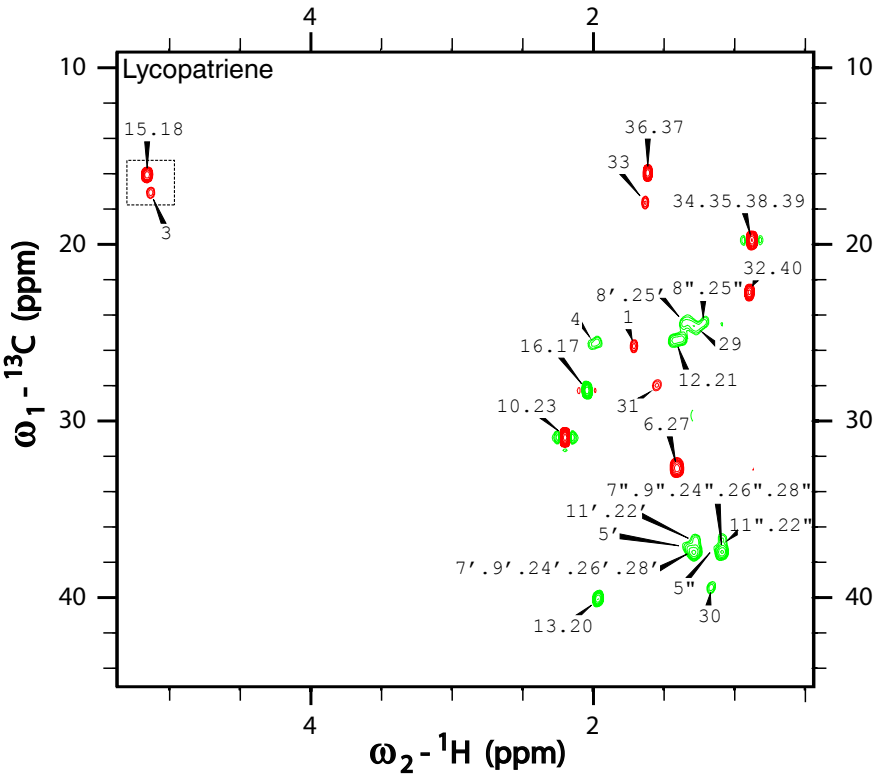


Figure 19. NMR analysis of lycopatriene. Positive (red) contours represent -CH or -CH_3 correlations, while negative (green) contours indicate -CH_2. Peaks originating from the molecule of interest are marked with resonance assignment, while those from impurities are left unmarked. Peaks aliased on ^{13}C axis are shown in a dashed line box.

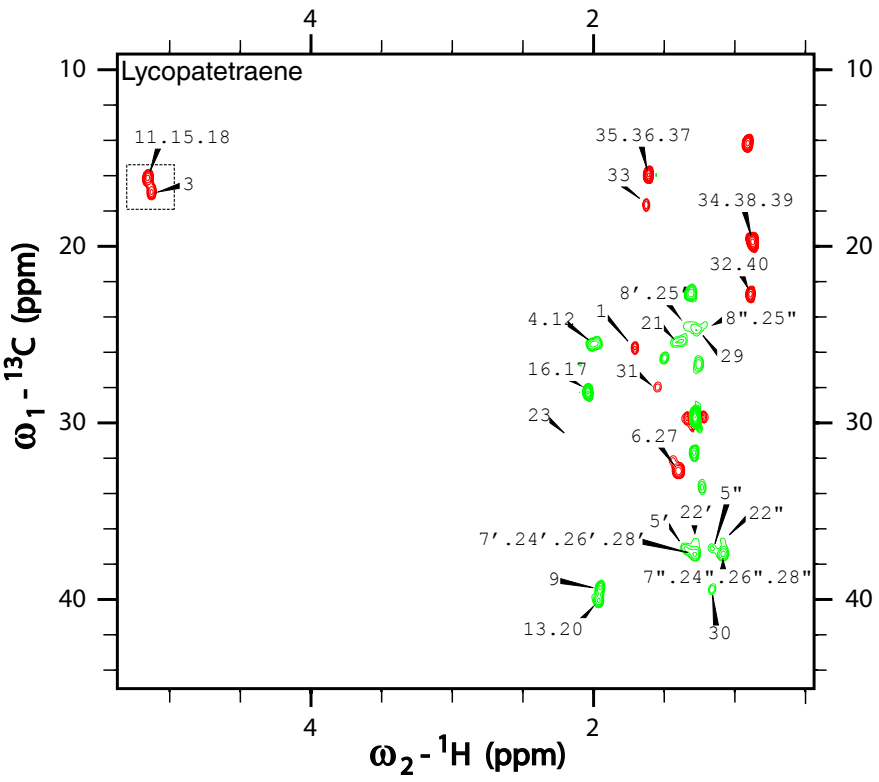


Figure 20. NMR analysis of lycopatetraene. 2D-^{13}C-HSQC spectrum of lycopatetraene. Positive (red) contours represent -CH or -CH_3 correlations, while negative (green) contours indicate -CH_2. Peaks originating from the molecule of interest are marked with resonance assignment, while those from impurities are left unmarked. Peaks aliased on ^{13}C axis are shown in a dashed line box.

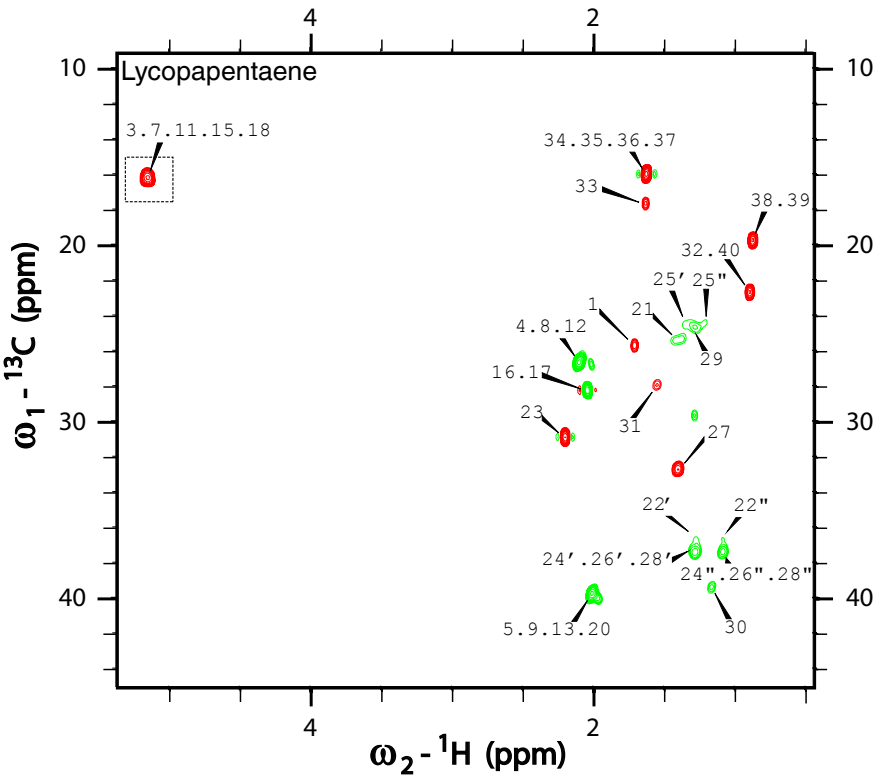


Figure 21. NMR analysis of lycopapentaene. 2D-^{13}C-HSQC spectrum of lycopapentaene. Positive (red) contours represent -CH or -CH_3 correlations, while negative (green) contours indicate -CH_2. Peaks originating from the molecule of interest are marked with resonance assignment, while those from impurities are left unmarked. Peaks aliased on ^{13}C axis are shown in a dashed line box.

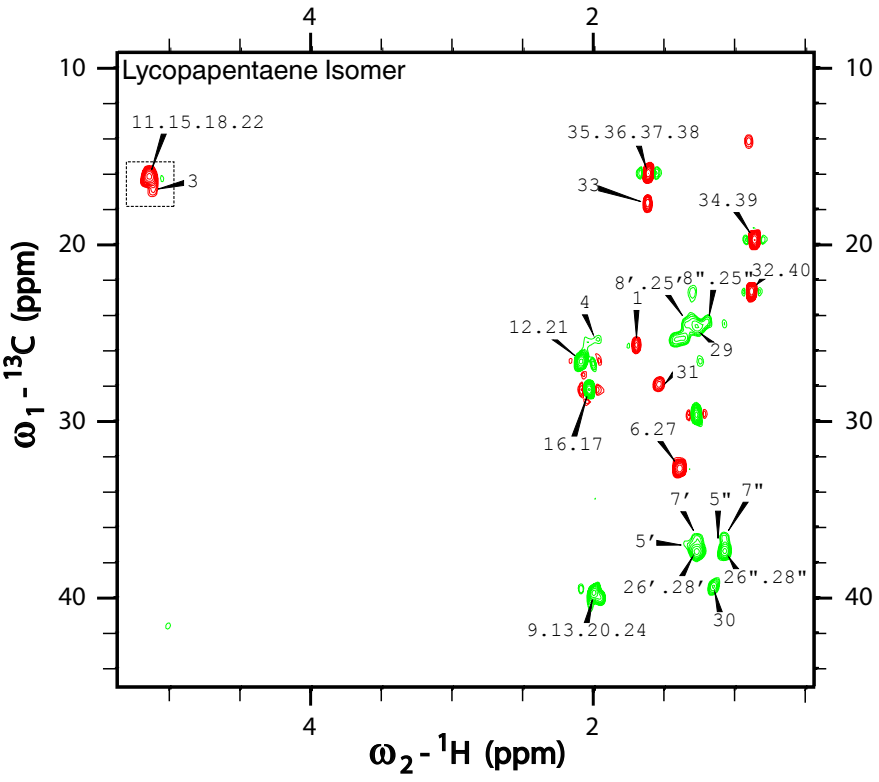


Figure 22. NMR analysis of lycopapentaene isomer. Positive (red) contours represent -CH or -CH_3 correlations, while negative (green) contours indicate -CH_2. Peaks originating from the molecule of interest are marked with resonance assignment, while those from impurities are left unmarked. Peaks aliased on ^{13}C axis are shown in a dashed line box.

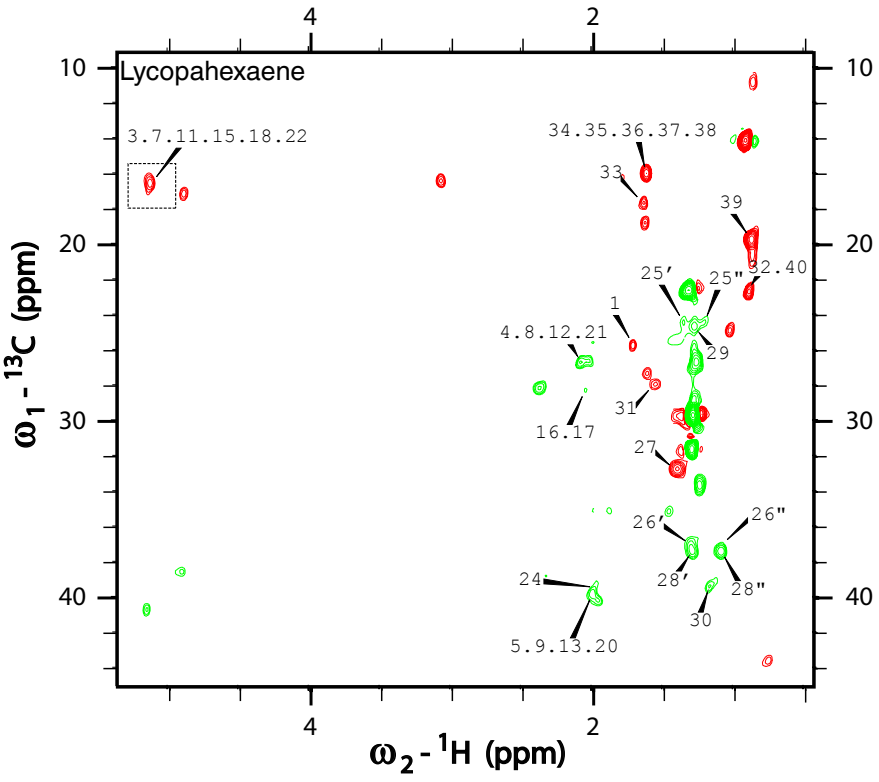


Figure 23. NMR analysis of lycopahexaene. 2D-^{13}C-HSQC spectrum of lycopahexaene. Positive (red) contours represent -CH or -CH_3 correlations, while negative (green) contours indicate -CH_2. Peaks originating from the molecule of interest are marked with resonance assignment, while those from impurities are left unmarked. Peaks aliased on ^{13}C axis are shown in a dashed line box.

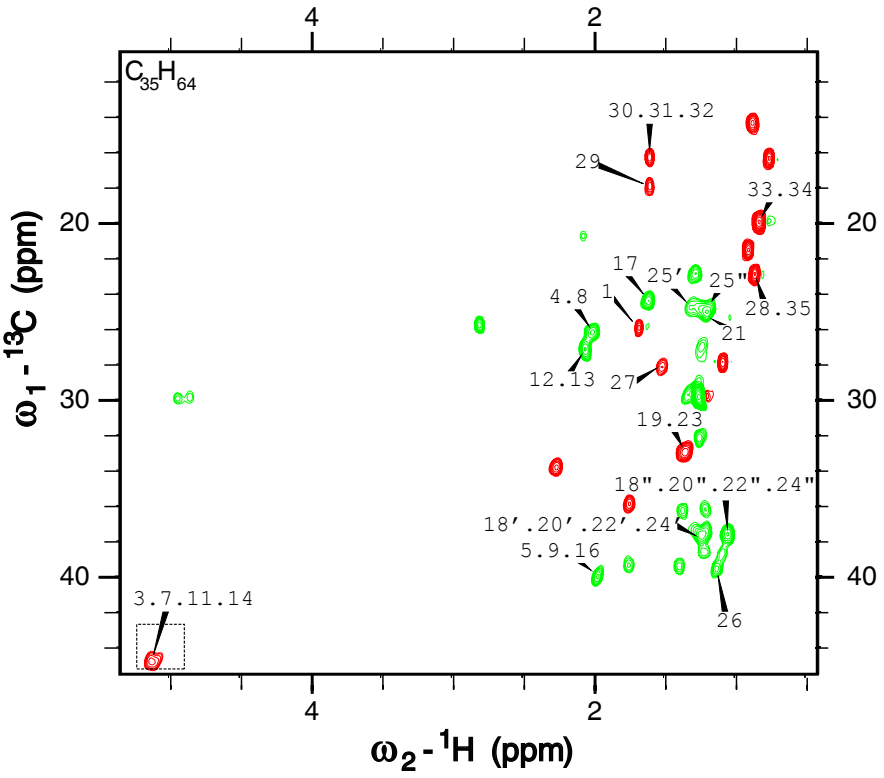


Figure 24. NMR analysis of $C_{35}H_{64}$. 2D-^{13}C-HSQC spectrum of $C_{35}H_{64}$. Positive (red) contours represent -CH or -CH_3 correlations, while negative (green) contours indicate -CH_2. Peaks originating from the molecule of interest are marked with resonance assignment, while those from impurities are left unmarked. Peaks aliased on ^{13}C axis are shown in a dashed line box.

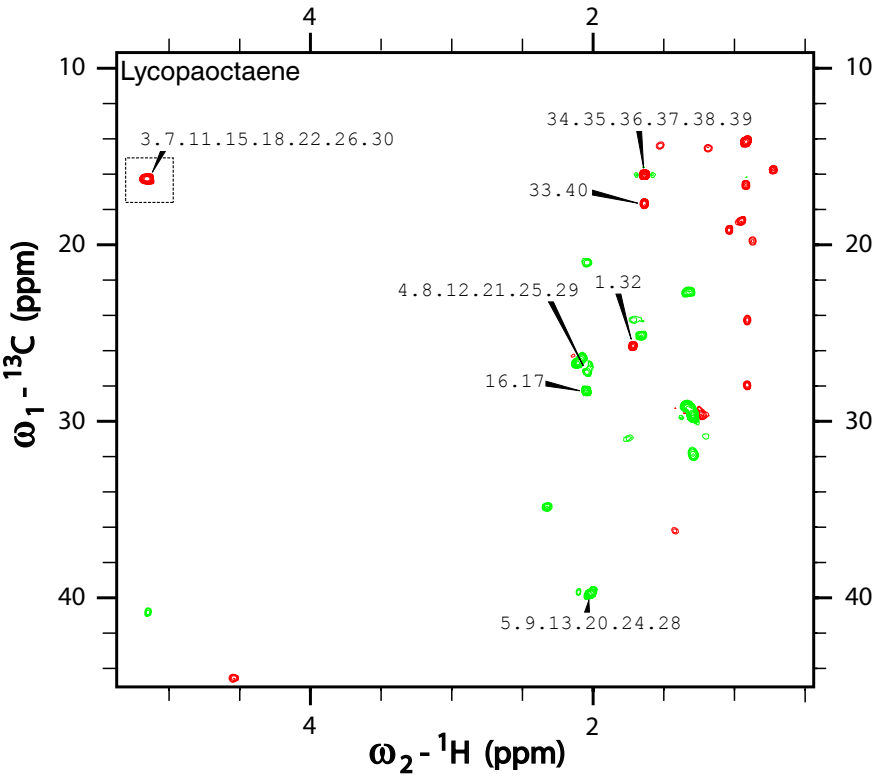


Figure 25. NMR analysis of lycopaoctaene. 2D-^{13}C-HSQC spectrum of lycopaoctaene. Positive (red) contours represent -CH or -CH_3 correlations, while negative (green) contours indicate -CH_2. Peaks originating from the molecule of interest are marked with resonance assignment, while those from impurities are left unmarked. Peaks aliased on ^{13}C axis are shown in a dashed line box.

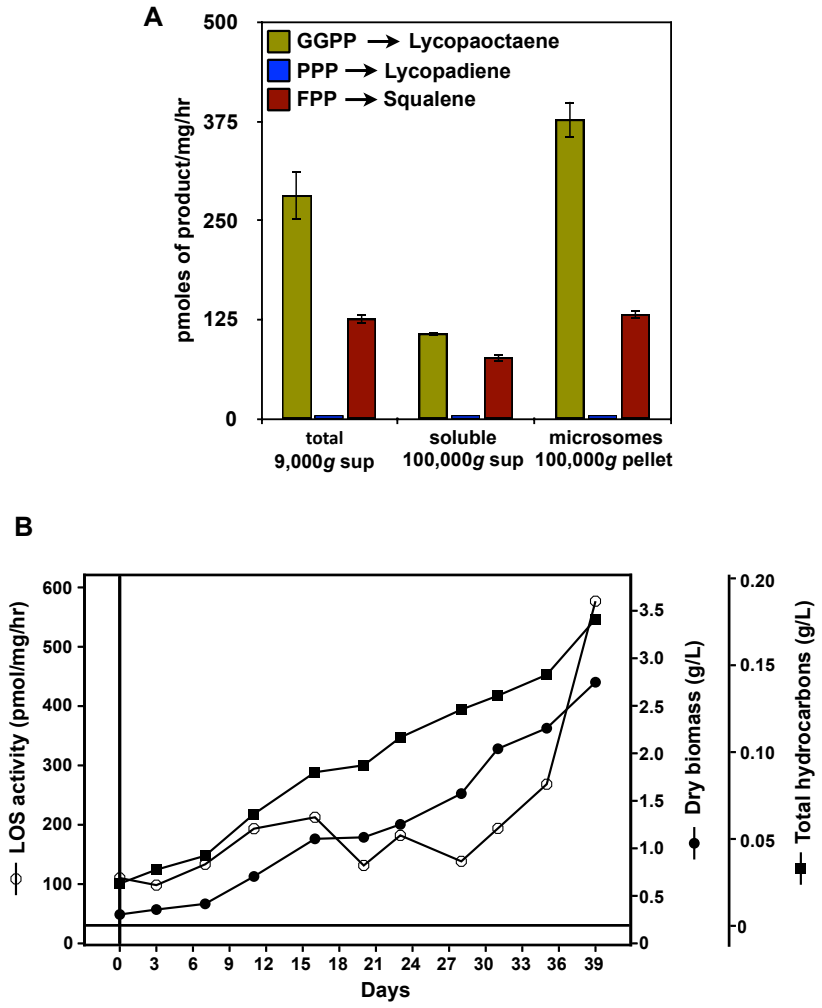


Figure 26. Hydrocarbon-related enzyme activities in *B. braunii* race L.
A. Radioactive enzyme assays were conducted using different protein fractions of race L cell lysates to test the two possible lycopadiene biosynthetic pathways. Enzyme activities were calculated based on the incorporation of ^{3}H-GGPP into lycopaoctaene, ^{3}H-PPP into lycopadiene, and ^{3}H-FPP into squalene. Squalene synthase enzyme activity is used as a positive control. Values shown are the mean ± standard error obtained from three independent experiments (n = 3). **B.** Lycopaoctaene synthase (LOS) enzyme activity, growth rate (dry biomass accumulation), and total hydrocarbon production in race L over a 39-day culture period. Race L shows a rapid increase in growth rate after 12 days of inoculation into new medium, and a direct correlation between LOS enzyme activity and hydrocarbon accumulation was observed, with LOS activity increasing rapidly after day 30 of the culture cycle.

localized to a membrane system, possibly the ER as is seen for SS[82]. Tellingly, LOS activity in these assays was detected when using the same cofactor and divalent metal cation as those used for SS activity measurement. These results suggest direct lycopadiene biosynthesis from PPP is not a significant contributor to C_{40} hydrocarbon production in race L, while conversion of GGPP into lycopaoctaene occurs readily, as proposed in the second hypothetical route (Figure 9B). Correspondingly, LOS enzyme activity directly correlated with lycopadiene accumulation over the growth cycle, suggesting this activity is related to hydrocarbon biosynthesis (Figure 26B). Detection of LOS activity from an algal homogenate in an assay similar to that of SS supports the notion that an LOS enzyme may be similar to a typical SS enzyme.

2.2b GGPP is a precursor for tetraterpenoid hydrocarbons in race L

To investigate whether GGPP is incorporated into lycopaoctaene, which would act as the precursor for lycopadiene and its unsaturated homologues, we conducted a pulse-chase experiment using ^{14}C-geranylgeraniol (GGOH) instead of ^{14}C-GGPP. GGOH was used for two reasons; first, GGPP will not be taken up by cells due to the highly charged diphosphate moiety, and second prenyl alcohols such as GGOH and FOH have been shown to be converted to their mono and diphosphate versions once taken up by plant cells[111] including *B. braunii*, race B[112]. We confirmed that enzyme activities exist in *B. braunii* race L for the conversion of GGOH first to GGP and subsequently to GGPP (Figure 27A) as has been seen in tobacco[111]. After a 48-hour ^{14}C-GGOH pulse, algae cells were transferred to fresh media containing non-labelled GGOH, samples

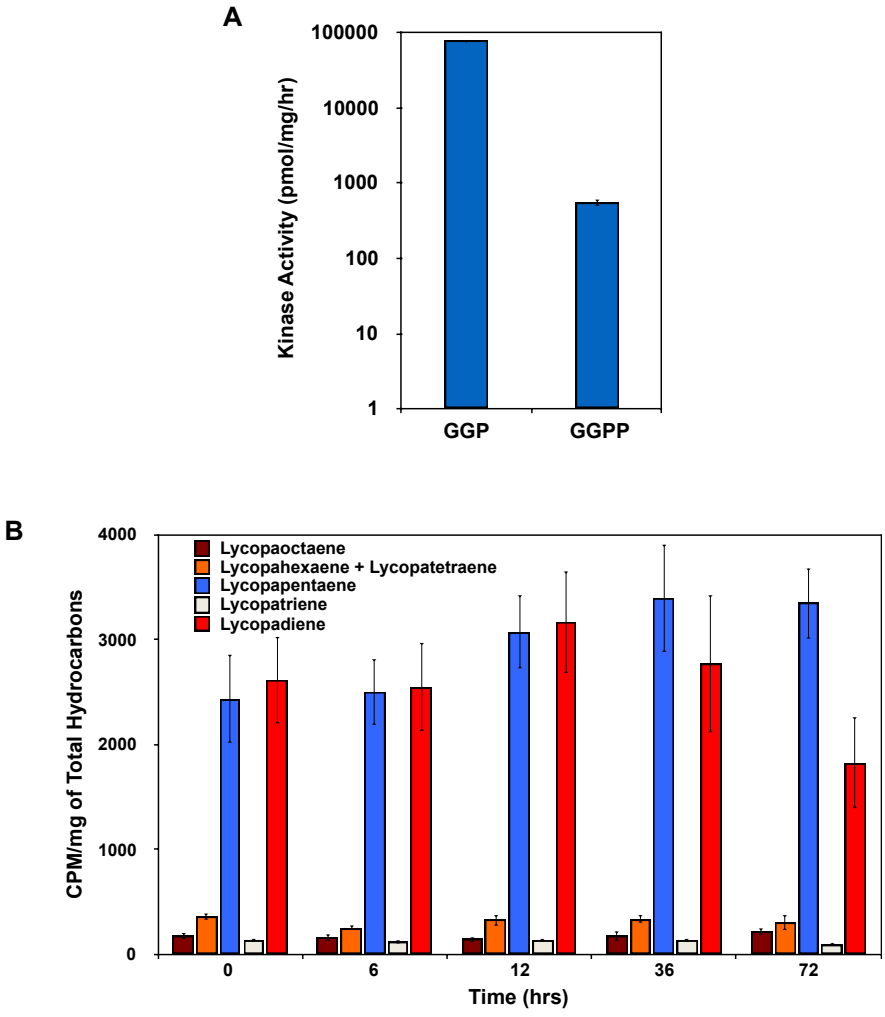


Figure 27. Geranylgeraniol kinase assay and radioactive feeding experiment. GGOH = Geranylgeraniol. **A.** The kinase assay shows incorporation of ^{14}C-GGOH into ^{14}C-geranylgeranyl monophosphate (GGP) and ^{14}C-geranylgeranyldiphosphate (GGPP) by a microsomal protein fraction of race L. **B.** Pulse-Chase experiments were conducted to analyze relationships between precursor and intermediate levels related to lycopadiene biosynthesis. Algae cells were pulsed with ^{14}C-GGOH and chased with nonlabelled GGOH for 72-hours followed by determination of ^{14}C incorporation into individual hydrocarbons. Data shown in (A) and (B) are an representatives from three independent experiments (n = 3).

collected over a 72 hour chase, hydrocarbons extracted and separated by thin layer chromatography (TLC), and radioactive incorporation into each hydrocarbon molecule determined (Figure 27B). Incorporation of ^{14}C into lycopaoctaene was determined based on the TLC spot with an identical R_f value to a lycopaoctaene standard (generated in yeast as described below). ^{14}C incorporation into lycopahexaene and lycopatetraene could not be separated due to TLC comigration. Except for lycopapentaene and lycopadiene, ^{14}C incorporation into lycopaoctaene, lycopahexaene/lycopatetraene, and lycopatriene was low and remained constant over the 72-hour chase (Figure 27B) suggesting rapid reduction of these molecules to their respective higher saturated homologues. In contrast, ^{14}C incorporation into lycopapentaene increased over the chase and eventually surpassed the radioactivity incorporated into lycopadiene (Figure 27B). This was surprising since lycopapentaene accounts for only 3% of the total hydrocarbon fraction (Figure 10) suggesting reduction of lycopapentaene as a rate-limiting step in lycopadiene biosynthesis. Furthermore, ^{14}C incorporation into lycopadiene increased until 12-hours and then declined until 72-hours after the chase (Figure 27B). This is consistent with previous studies showing lycopadiene is further utilized to make the cross-linked hydrocarbon network forming the race L extracellular matrix[113, 114]. Altogether, the feeding experiments show GGPP is the direct precursor for lycopaoctaene production, with lycopahexaene, lycopapentaene, lycopatetraene and lycopatriene as intermediates in lycopadiene biosynthesis.

2.2c Identification of a lycopaoctaene synthase enzyme

To identify the gene(s) responsible for lycopaoctaene biosynthesis, we generated and computationally screened an L race transcriptomic database for SS-like sequences. Two SS-like cDNAs were identified and named based on the function of their encoded proteins as detailed below; SS from race L (LSS) and lycopaoctaene synthase (LOS). Both the LSS and LOS proteins contain all five conserved activity domains, the transmembrane domain, and the NADPH binding residues found in typical SS enzymes (Figures 28 and 29)[79]. Comparison of the encoded amino acid sequences showed 57.5% and 49.2% identity to SS from race B (BSS)[57] for LSS and LOS, respectively (Figure 29). Additionally, comparison of LOS with other SS and SS-like enzymes showed 52.5% identity to LSS, 47.3% identity to the *Solanum lycopersicum* (tomato) SS (*Sl*SS), 44.6%, 61.7%, and 44.1% identity to the squalene synthase-like enzymes (SSL) 1, 2, and 3 from the B race of *B. braunii*, and 33.9% identity to the human SS (HSS; Figure 29). Several studies[79, 81, 92, 93], including the recent determination of the crystal structure of HSS[94], have elucidated the catalytic mechanism of the SS reaction. These studies have shown the FLAP domain, domains I – V, and the JK loop of HSS form the SS active site[94]. Furthermore, several HSS residues (Ser51, Ser53, Tyr171, Gln212), two DXXED motifs in domains II and IV, and the NADPH-binding residues were determined to be key for catalytic activity (Figure 29)[94]. All of these catalytic residues and motifs are also conserved in both the LSS and LOS enzymes (Figure 29), making it difficult at this time to determine which domains/residues are specific for LOS activity in a comparison of LSS and LOS.

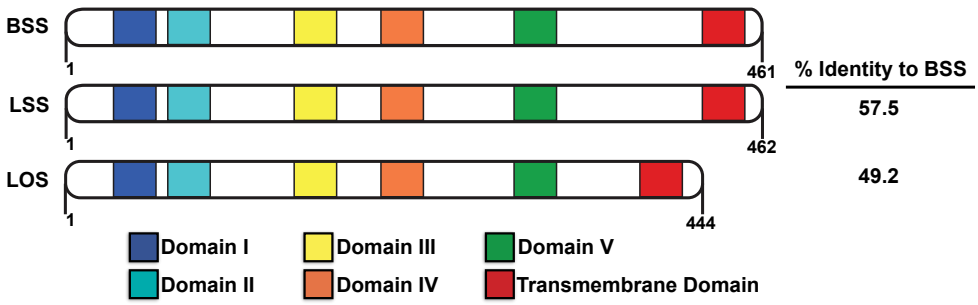
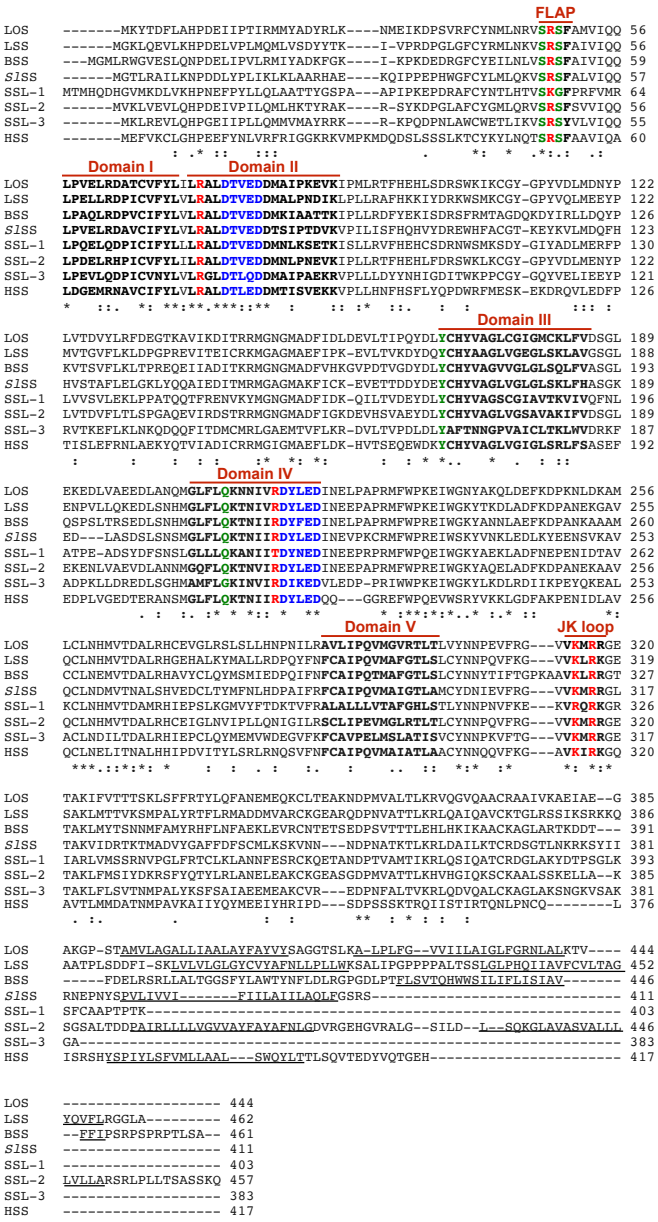


Figure 28. Protein domain alignment of SS-like cDNAs from *B. braunii*. SS= squalene synthase, BSS = race B SS, LSS = race L SS, LOS = lycopaoctaene synthase from race L.



Protein	% identity to LOS
LSS	52.5
BSS	49.2
*Sl*SS	47.3
SSL-1	44.6
SSL-2	61.7
SSL-3	44.1
HSS	33.9

Figure 29. Amino acid sequence alignment of SS and SS-like proteins. FLAP region, five conserved domains and putative NADPH binding residues (indicated in red) found in typical squalene synthase are labeled and shown in bold. Possible transmembrane region(s) predicted by TMpred for each protein are underlined. HSS = human squalene synthase, LOS = lycopaoctaene synthase, LSS = race L SS, BSS = race B SS, *Sl*SS = SS from *Solanum lycopersicum*, and SSL-1, SSL-2, SSL-3, SS-like enzymes from race B.

For initial characterization of the enzymatic activity of LSS and LOS, both proteins were expressed in *E. coli* and the purified recombinant proteins tested for enzyme activity. LSS was shown to yield squalene as its sole product in an *in vitro* assay with FPP as the substrate, as expected (Figures 30A and 30B, panel I). However, purified recombinant LOS incubated with GGPP in an *in vitro* assay showed lycopaoctaene as the sole reaction product (Figures 30A and 30B, panel II). Since LOS may have arisen from an SS paralogue that evolved to accept GGPP as substrate for lycopaoctaene production, we considered that LOS may have retained the ability to utilize FPP to produce squalene. Indeed, squalene production was detected when LOS was incubated with FPP in an *in vitro* reaction (Figures 30A and 30B, panel III). More surprising, LOS incubation with PPP yielded lycopadiene, albeit at levels much lower than lycopaoctaene or squalene (Figures 30A and 30B, panel IV). Similar *in vitro* incubations of LSS with GGPP or PPP did not result in lycopaoctaene or lycopadiene production (see below). Because LOS can utilize three different substrates, steady state kinetic experiments were performed with FPP, GGPP, and PPP to analyze LOS substrate specificity. The LOS enzyme shows higher substrate affinity (K_m) for GGPP compared to that of FPP and PPP, and the turnover number (k_{cat}) and catalytic efficiency (K_{cat}/K_m) for PPP are an order of magnitude less than for FPP and GGPP (Figure 31).

To further characterize LOS, it was coexpressed in yeast with *Arabidopsis thaliana* GGPP synthase-11 (*At*GGPPS11)[115] since unlike FPP, yeast GGPP biosynthesis is limited and considered a bottleneck for GGPP-derived isoprenoid production[116]. Coexpression of LOS and *At*GGPPS11 in yeast resulted in lycopaoctaene production

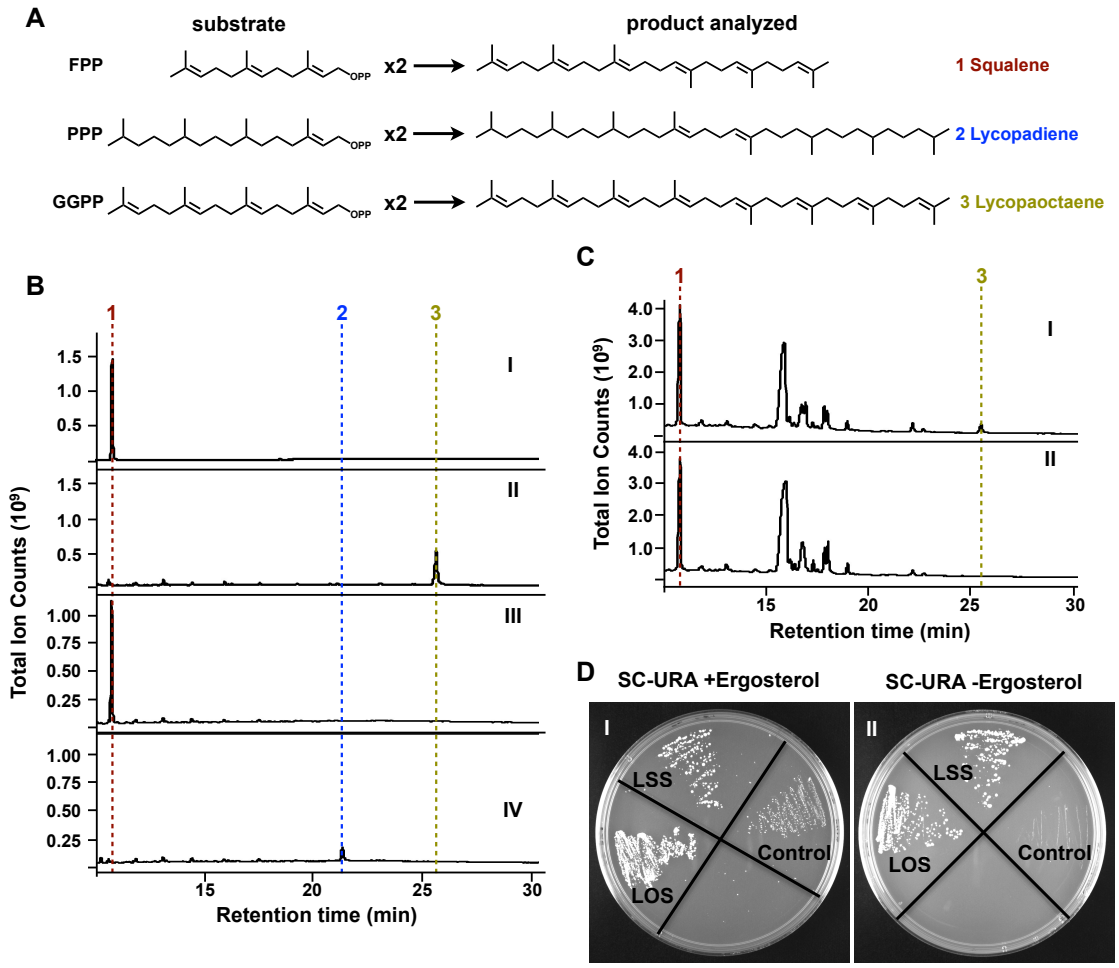


Figure 30. Functional characterization of SS-like enzymes from race L. A. Diagram of reactions showing substrates and products for the analysis in B and D. **B.** GC-MS profiles of LSS and LOS *in vitro* reaction products: I, LSS products using FPP as substrate to produce squalene (1); II, LOS products using GGPP as substrate to produce lycopaoctaene (3); III, LOS products using FPP as substrate to produce squalene (1); and IV, LOS products using PPP as substrate to produce lycopadiene (2). **C.** GC-MS profiles of LOS yeast *in vivo* reaction products: I, *n*-hexane extractable metabolites of yeast expressing *At*GGPPS11 and LOS to make lycopaoctaene (3); or II, yeast expressing *At*GGPPS11 alone. **D.** Expression of LSS and LOS in the *SS* (*ERG9*) yeast knockout strain: I, in the presence of ergosterol; or II, absence of ergosterol to test restoration of ergosterol prototrophy. The control carries the expression vector without a transgene. Data shown in (B), (C) and (D) are an representatives from three independent experiments (n = 3).

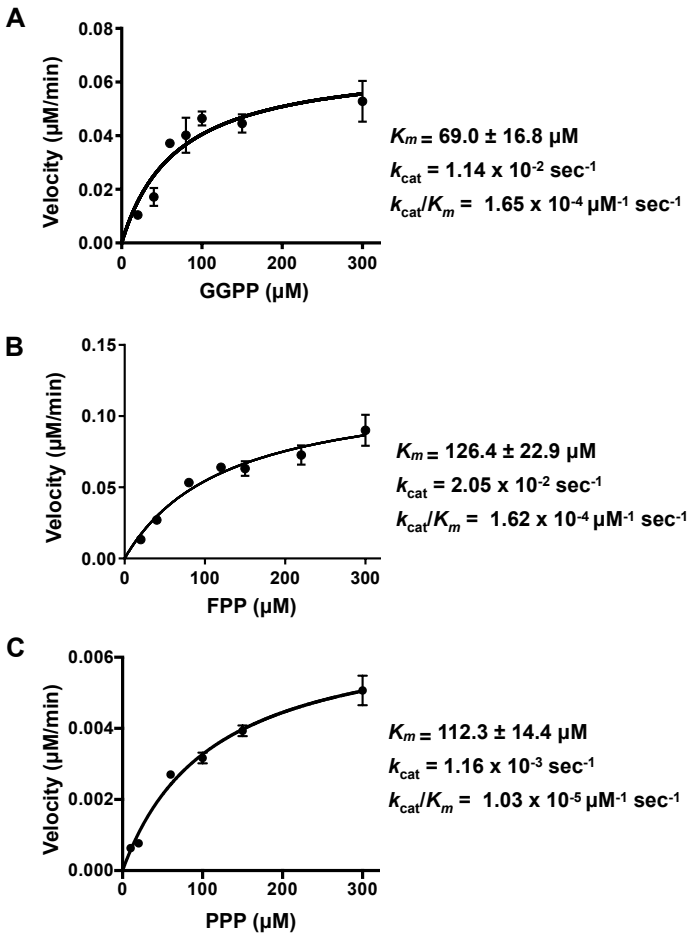


Figure 31. Kinetics of the LOS enzyme reaction with different substrates. A. LOS kinetics using geranylgeranyl diphosphate (GGPP) as substrate. **B**. LOS kinetics using farnesyl diphosphate (FPP) as substrate. **C**. LOS kinetics using phytyl diphosphate (PPP) as substrate. Values shown are the mean ± standard error obtained from three independent measurements.

(Figure 30C, panel I), which was undetectable when *At*GGPPS11 was expressed without LOS (Figure 30C, panel II). Furthermore, when expressed in a yeast SS knockout strain, LOS restored ergosterol prototrophy indicating its ability to produce squalene *in vivo* (Figure 30D, panel I and II). LSS was also able to restore ergosterol prototrophy as expected (Figure 30D, panel I and II). Taken together, the results in Figure 30 suggest LSS is a true SS enzyme, whereas LOS appears to be a promiscuous SS-like enzyme with broader substrate chain length and saturation specificity.

Previous studies have reported the production of lycopaoctaene (a.k.a. lycopersene) *in vitro* from protein extracts during studies conducted to decipher the carotenoid biosynthetic pathway[32-34], but subsequent studies determined lycopaoctaene could not be an intermediate in carotenoid biosynthesis[35,36] raising the possibility that lycopaoctaene production in these studies was an *in vitro* artifact. A more relevant study reported the ability of purified yeast SS to utilize GGPP for lycopaoctaene production *in vitro*, at levels much lower than native squalene production[37]. Importantly, we did not detect any lycopaoctaene production when *At*GGPPS11 was expressed alone in wild-type yeast (Figure 30C, panel II), suggesting that yeast SS does not utilize GGPP *in vivo* under the conditions employed. Similarly, the ability of LOS to catalyze the conversion of PPP to lycopadiene may not have biological significance as we did not detect lycopadiene production using PPP as the substrate *ex vivo* in race L cell lysates (Figure 26A), and the efficiency of LOS to use PPP as a substrate is quite low (Figure 31). Together, these results suggest that the relevant *in vivo* route to lycopadiene production is unlikely to involve the condensation of two PPP molecules by LOS, but rather, as we

contend, the condensation of two GGPP molecules to form lycopaoctaene, for eventual conversion to lycopadiene.

2.2d LOS is promiscuous towards prenyl substrates

The LOS enzyme was further characterized using combinations of FPP, GGPP and PPP as substrates in *in vitro* reactions. Interestingly, in the presence of FPP and GGPP LOS produced significant amounts of squalene and a $C_{35}H_{58}$ molecule, and lesser amounts of lycopaoctaene (Figure 32, panel I). The $C_{35}H_{58}$ molecule is a chimera produced from head-to-head condensation of one FPP molecule and one GGPP molecule, and its identity was confirmed by GC-MS (Figure 33). Next, when LOS was supplied with FPP and PPP, squalene production predominated with small amounts of $C_{35}H_{64}$ and lycopadiene (Figure 32, panel II). $C_{35}H_{64}$, also a chimeric hydrocarbon, is produced as a result of head-to-head condensation of one molecule each of FPP and PPP. Finally, LOS incubation with GGPP and PPP produced lycopadiene and lycopapentaene as minor products and lycopaoctaene as the major product (Figure 32, panel III).

To the best of our knowledge, this is the first report where a wild type, non-mutated eukaryotic SS or SS-like enzyme has been shown to be able to utilize three naturally occurring prenyl diphosphate substrates to yield hydrocarbon products in all possible combinations. To support this notion, several SS and SS-like enzymes were tested for their ability to utilize FPP, GGPP, and PPP as substrates (Figure 34). The enzymes tested included LOS and three typical SS enzymes; LSS, BSS, and *Sl*SS. Additionally, SS-like enzymes SSL-1, SSL-2, and SSL-3 from race B were chosen because of their SS-like activities. SSL-1 utilizes FPP to produce PSPP, which is then

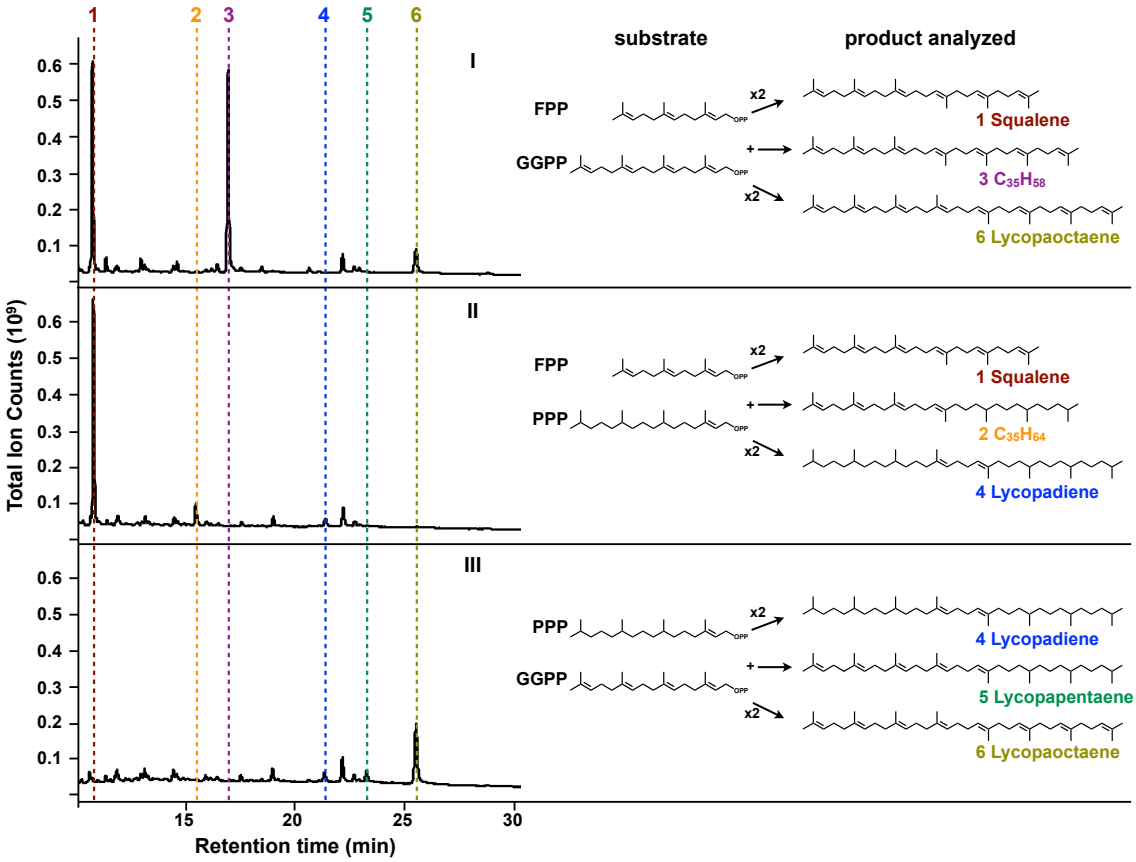


Figure 32. Characterization of substrate use of LOS enzyme. Diagram of reactions showing substrates and products and GC-MS profiles of LOS enzyme assay products with: **I,** Farnesyl diphosphate (FPP) and geranylgeranyl diphosphate (GGPP) substrates to produce squalene **(1)**, $C_{35}H_{58}$ **(3)** and lycopaoctaene **(6)**; **II,** FPP and phytyl diphosphate (PPP) substrates to make squalene **(1)**, $C_{35}H_{64}$ **(2)** and lycopadiene **(4)**; and **III,** GGPP and PPP substrates to produce lycopadiene (4), lycopapentaene **(5)** and lycopaoctaene **(6)**. Data shown are representatives from three independent experiments (n = 3).

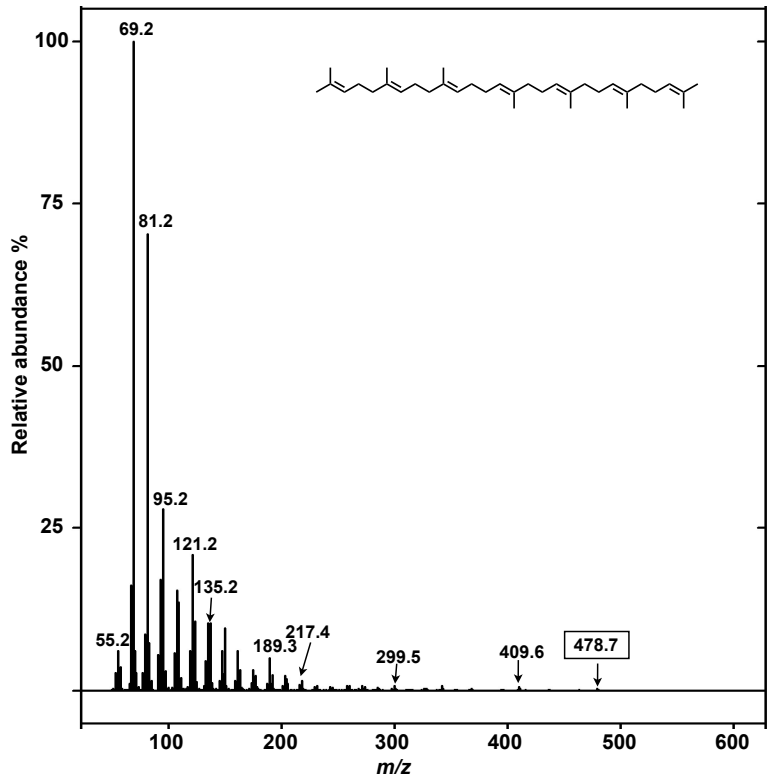


Figure 33. Analysis of $C_{35}H_{58}$ by GC-MS. The mass spectrum and structure of the $C_{35}H_{58}$ molecule produced by the LOS enzyme when FPP and GGPP are used in combination as substrate. Molecular ion in box corresponds to the parent ion of the molecule.

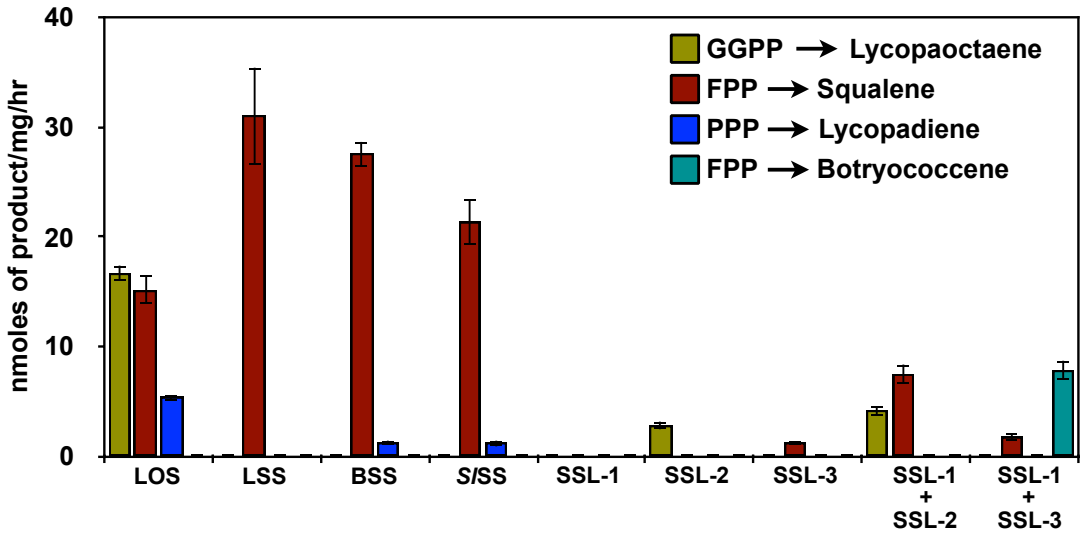


Figure 34. Characterization of substrate use for SS & SS-like enzymes. Radioactive enzyme assays using ^{3}H substrates geranylgeranyl diphosphate (GGPP), farnesyl diphosphate (FPP) and phytyl diphosphate (PPP) with selected SS or SS-like enzymes. SSL-1 produces presqualene diphosphate (PSPP) from FPP substrate, however, PSPP was not analyzed for in this assay. LOS, lycopaoctaene synthase; LSS, race L SS; BSS, race B SS; *Sl*SS, SS from *Solanum lycopersicum* (tomato); and SSL-1, SSL-2, SSL-3, SS-like enzymes from the B race of *B. braunii*. Values shown are the mean ± standard error obtained from three independent measurements (n = 3).

converted by SSL-2 to squalene, or by SSL-3 to C_{30}-botryococcene as the major product and squalene as a minor product[38]. As shown in figure 34, the typical SS enzymes display SS activity as their main catalytic function, with BSS and *Sl*SS also utilizing PPP to produce minute amounts of lycopadiene. In contrast, the B race SSL enzymes have limited substrate flexibility, allowing SSL-2 and SSL-1 plus SSL-2 to use GGPP to generate minor amounts of lycopaoctaene (Figure 34). However, LOS is the only enzyme tested that is able to utilize all three substrates and to produce significant amounts of lycopaoctaene, squalene, and lycopadiene (Figure 34).

2.2e The LOS reaction uses a cyclopropyl intermediate

We next conducted enzyme assays to determine if the LOS reaction mechanism with GGPP as substrate is similar to that of SS, i.e. utilizes a PSPP-like cyclopropyl diphosphate intermediate, which is termed prelycopaoctaene diphosphate (PLPP; Figure 35A). First, assays were conducted with or without a dinucleotide reducing agent, which would be required to convert PLPP to lycopaoctaene (Figure 35A). LOS successfully utilized both NADH and NADPH as reducing agents for lycopaoctaene production, with preference for NADPH (Figure 35B). In the absence of a dinucleotide reducing agent LOS activity was lost (Figure 35B), suggesting the presence of the PLPP reaction intermediate. This result is consistent with previous reports where yeast SS utilized GGPP *in vitro* to produce lycopaoctaene; however, without NADH or NADPH only the reaction intermediate PLLPP accumulated[117]. Next, we showed that LOS is strongly inhibited by squalestatin (Figure 35C), a potent SS inhibitor that mimics PSPP binding[105, 118-120]. Finally, we identified PLPP as a reaction intermediate by conducting a

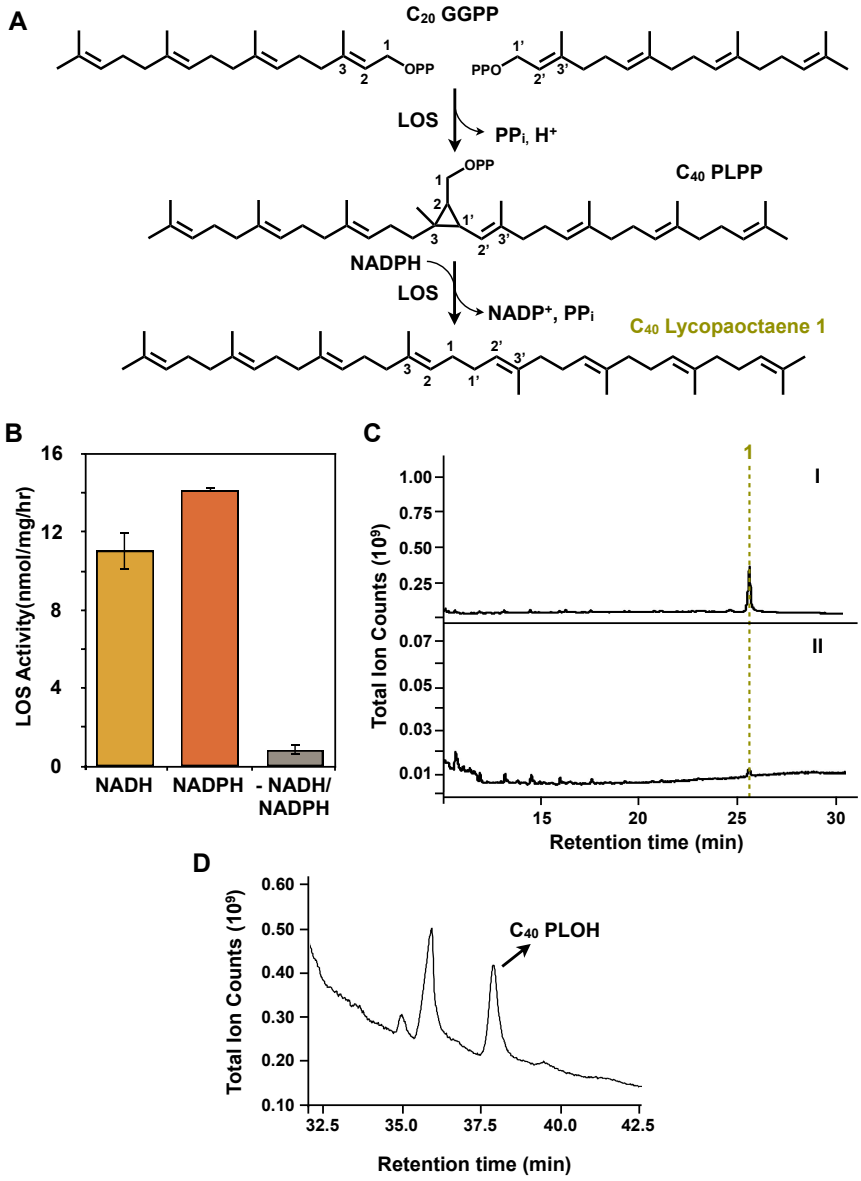


Figure 35. Characterization of the lycopaoctaene synthase reaction intermediate. A. Proposed LOS reaction mechanism to produce the prelycopaoctaene diphosphate (PLPP) intermediate and lycopaoctaene final product. **B.** Radioactive LOS enzyme assays conducted with ^{3}H-GGPP substrate with or without NADH or NADPH reducing agents. Values shown are the mean ± standard error obtained from three independent measurements (n = 3). **C.** GC-MS profiles of LOS enzyme assay with GGPP as substrate to produce lycopaoctaene **(1)**: **I,** in the absence; or **II,** presence of 20 nM squalestatin. **D.** GC-MS profile of products from LOS enzyme assay with GGPP substrate in the absence of NADPH to produce PLPP, which was hydrolyzed by acid phosphatase to yield prelycopaoctaene alcohol (PLOH) for GC-MS analysis. Data shown in (C) and (D) are representatives from three different experiments (n = 3).

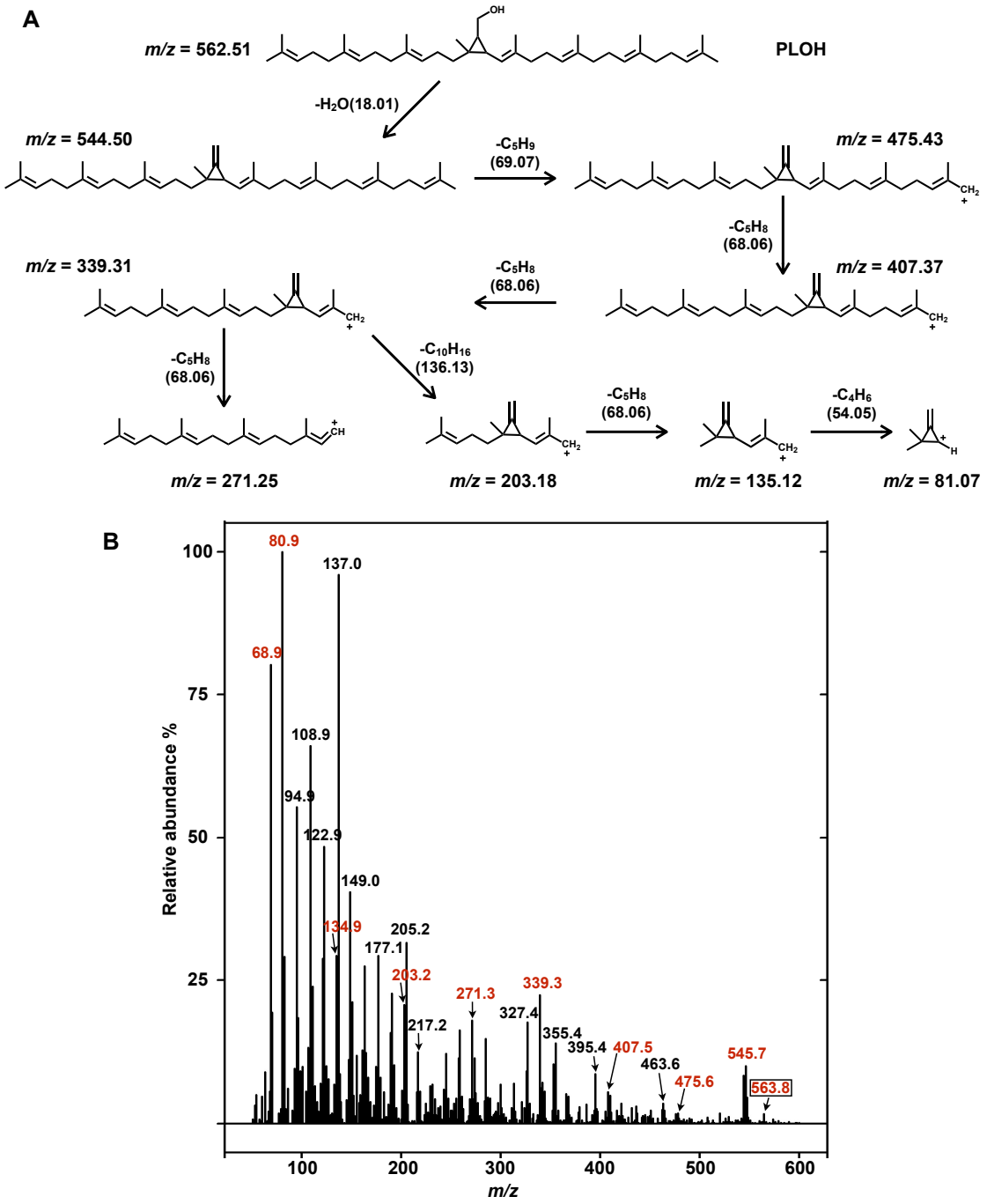


Figure 36. GC-MS analysis of prelycopaoctaene alcohol (PLOH). A. Expected mass spectral fragmentation pattern of PLOH. Mass fragmentation pattern of PLOH shown above is adapted from Quershi et al, *J. Biol. Chem*. **248**:2755-2767 (1973). **B.** The mass spectrum of PLOH molecule when analyzed in chemical ionization mode. Molecular ion in box corresponds to the parent ion of the PLOH molecule and other ionic species in red correspond to fragments of PLOH molecule as shown above in A.

GGPP-based LOS assay in the absence of NADPH, followed by acid phosphatase treatment to convert PLPP to prelycopaoctaene alcohol (PLOH). Analysis by GC-MS showed the presence of PLOH (Figure 35D) with a fragmentation pattern consistent with previous reports of PLOH (Figure 36)[37]. Taken together, these results are consistent with a two-step LOS reaction to produce lycopaoctaene. By analogy to SS, LOS catalyzes the condensation of two GGPP units in the first half reaction to form the cyclopropylcarbinyl diphosphate intermediate PLPP, with concomitant release of one molecule of inorganic pyrophosphate (PP_i; Figure 35A). In the second half reaction, the PLPP cyclopropyl ring is cleaved and rearranged to form a 1-1' linkage, and further reduction by NADPH forms lycopaoctaene (Figure 35A).

2.2f Characterization of geranylgeranyl reductase-like genes from race L

Altogether our data supports the second possibility of lycopadiene biosynthesis where the LOS enzyme catalyzes the dimerization of two GGPP molecules to produce the eight double bond containing lycopaoctaene in the first step, which would then be sequentially reduced to produce lycopahexaene, lycopapentaene, lycopatetraene, lycopatriene, and lycopadiene (Figure 37). These multiple reductions could be catalyzed by a multifunctional reductase or multiple reductases with each enzyme catalyzing a stepwise reduction to form lycopadiene as the final product. The pulse-chase experiment with ^{14}C-GGOH showed reduction of lycopapentaene to lycopatetraene as the rate limiting step of the lycopadiene biosynthetic pathway (Figure 27B). This data suggests that at least two reductase enzymes could be involved in the reduction of lycopaoctaene

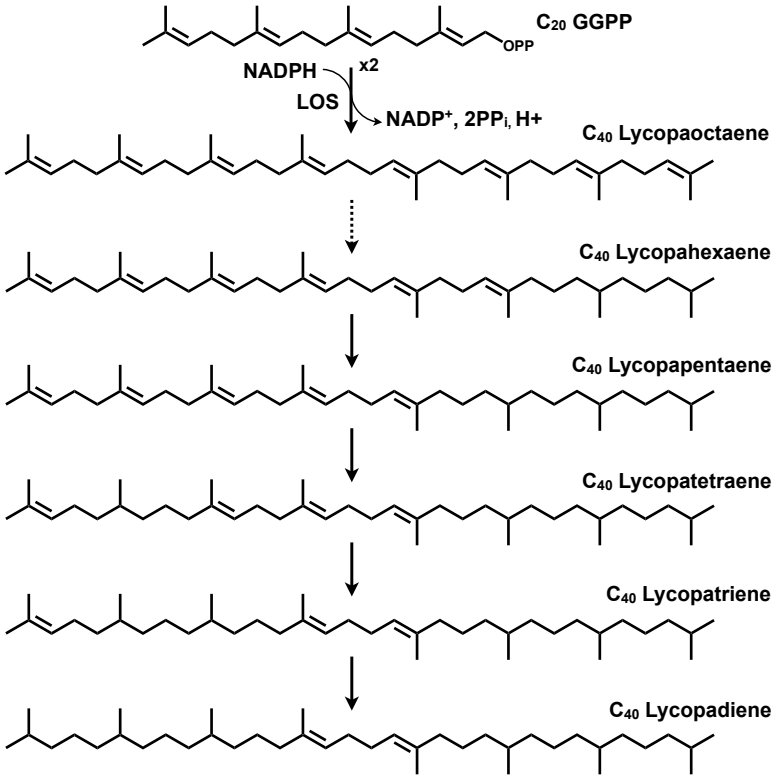


Figure 37. Proposed pathway for lycopadiene biosynthesis. The lycopaoctaene synthase (LOS) enzyme utilizes geranylgeranyl diphosphate (GGPP) in a two-step reaction to produce lycopaoctaene, which is further reduced by an unknown enzyme(s) to produce lycopahexaene, lycopapentaene, lycopatetraene, lycopatriene and finally lycopadiene.

to lycopadiene (Figure 37). The first putative reductase enzyme could reduce lycopaoctaene to lycopapentaene in a multi-step reduction reaction (Figure 37). Similarly, double bond reduction of lycopapentaene to lycopadiene could be catalyzed by second putative reductase enzyme in the pathway (Figure 37). This hypothesis is further supported by the structures of the lycopadiene intermediates. As shown in figure 37, reduction from lycopaoctaene to lycopapentaene shows reduction of the double bonds on one half of the molecule, and the reduction from lycopapentaene to lycopadiene involves the other half of the molecule.

Since each half of lycopaoctaene is made from a prenyl moiety of GGPP, the putative reductase(s) catalyzing the reduction of lycopaoctaene to lycopadiene could be similar to geranylgeranyl reductase (GGR). In plants such as *Arabidopsis thaliana*, a single multifunctional GGR (*At*GGR) enzyme catalyzes sequential reduction of three of the four double bonds of GGPP to produce dihydroGGPP, tetrahydroGGPP and phytyl diphosphate (PPP), in the chloroplast (Figure 38)[121]. PPP is used as the tail of the chlorophyll molecule and as a substrate for the biosynthesis of tocopherols and phylloquinone in photosynthetic organisms[121, 122]. Similarly, the *At*GGR enzyme also catalyzes sequential reduction of geranylgeranyl-chlorophyll to dihydrogeranylgeranyl-chlorophyll, tetrahydrogeranylgeranyl-chlorophyll and phytyl-chlorophyll, in etioplasts (Figure 38)[121]. Homologues of plant GGR found in archaea and bacteria also catalyze the sequential reduction of three of the four double bonds present in the geranylgeranyl side chain, and are involved in the biosynthesis of archaeal membrane lipids and bacteriochlorophyll[123, 124].

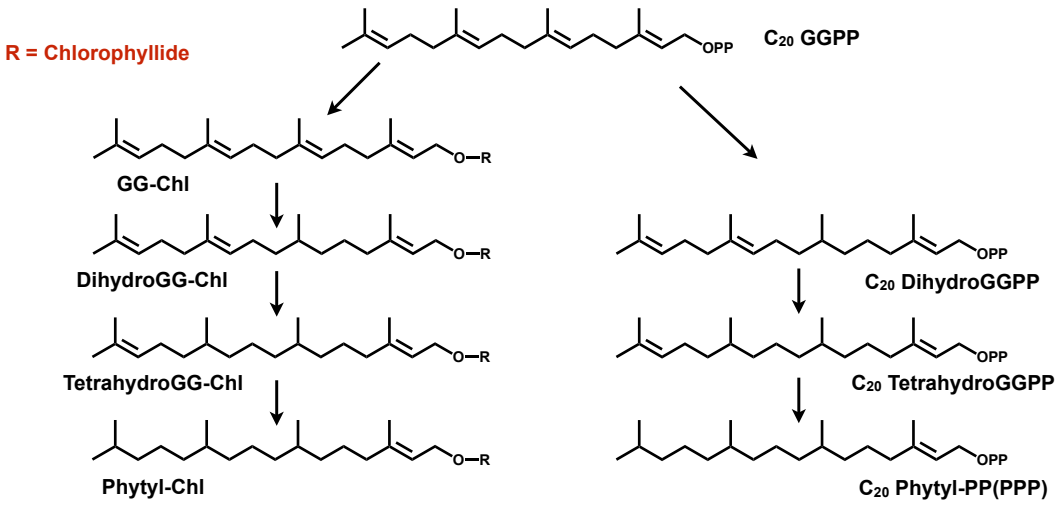


Figure 38. Reduction reaction catalyzed by *At*GGR enzyme. *At*GGR = geranylgeranyl reductase from *Arabidopsis thaliana*. GGPP = geranylgeranyl diphosphate, GG-Chl = geranylgeranyl-chlorophyll, DihydroGG-Chl = dihydrogeranylgeranyl-chlorophyll, TetrahydroGG-Chl = tetrahydrogeranylgeranyl-chlorophyll and Phytyl-Chl = phytyl-chlorophyll. The reactions shown above is drawn based on the study from Keller et al, *Eur. J. Biochem*. **251**:413-417 (1998).

To investigate the role of GGR-like enzymes in the reduction of lycopadiene biosynthetic intermediates, *GGR*-like cDNA sequences were searched for within the transcriptome of race L using BLAST and the *GGR* (*At*GGR) from *Arabidopsis thaliana* as a query. Two *GGR*-like cDNAs (*GGR-1* and *GGR-2*) were identified and cloned from race L. Alignment of the deduced amino acid sequence shows GGR-1 to be 66.8% and GGR-2 to be 21.6 % identical to *At*GGR (Figure 39). Both GGR-1 and GGR-2 contain a dinucleotide binding sequence motif found in typical GGR enzymes as they utilize the dinucleotide NADPH as an electron source for the reaction (Figure 39)[125]. Like *At*GGR, GGR-1 contains a chloroplast transit peptide (cTP) at its N-terminus suggesting plastid localization, and that GGR-1 could be a true GGR enzyme from race L (Figure 39). In contrast, GGR-2 does not contain a cTP but shows the presence of a predicted signal peptide and a transmembrane region at the N-terminus of the protein (Figure 39). For initial characterization of enzyme activity, we decided to express in *E. coli* cells engineered to produce lycopaoctaene molecule N-terminally truncated versions of GGR-1 and GGR-2 to delete the cTP for GGR-1 and the signal peptide and predicted transmembrane domain for GGR-2. An *in vivo* approach for characterization of GGR-1 and GGR-2 was chosen instead of an *in vitro* assay with purified enzymes for several reasons. First, lycopaoctaene, a potential substrate of the putative reductase is a highly hydrophobic molecule and is not soluble in an aqueous reaction environment. Second, the enzyme assay conditions such as cofactors and buffer conditions required for the reduction of lycopadiene biosynthetic intermediates is not known. Our initial strategy to

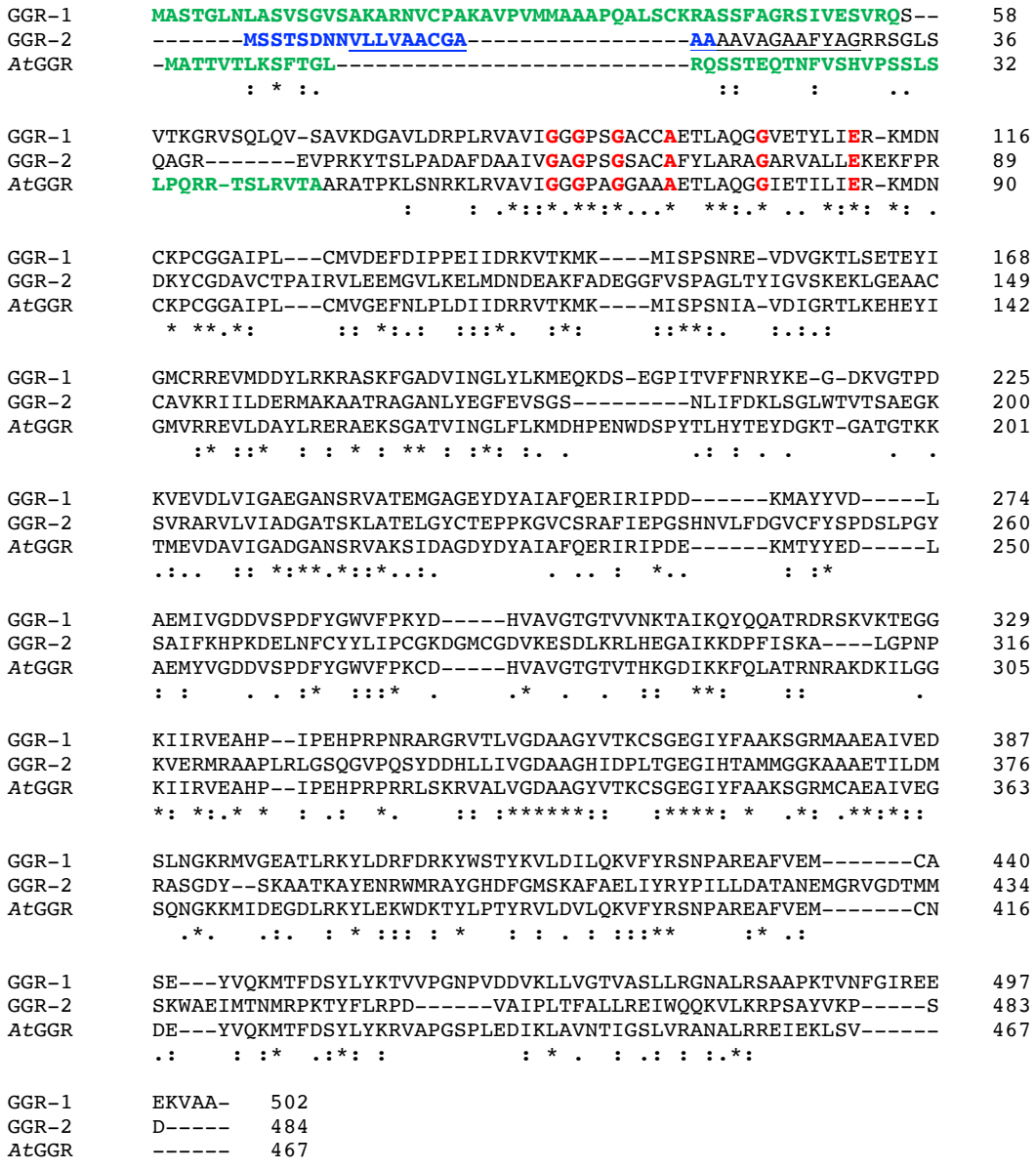


Figure 39. Amino acid sequence alignment of GGR-like sequences. GGR= geranylgeranyl reductase. GGR-1 and GGR-2 are from *B. braunii* race L and *At*GGR is from *Arabidopsis thaliana.* A chloroplast transit peptide predicted by the ChloroP 1.1 software for GGR-1 and *At*GGR is shown in green. A signal peptide predicted by the TargetP 1.1 software for GGR-2 is shown in blue. A possible transmembrane region predicted by TMpred for GGR-2 is underlined. Conserved dinucleotide binding sequences for each protein are highlighted in red.

produce lycopaoctaene in *E.coli* by overexpressing LOS did not yield any lycopaoctaene but resulted in the production of squalene and a small amount of $C_{35}H_{58}$ (Figure 40, panel I). This result is consistent with the previous report of low levels of intracellular GGPP precursors present in *E.coli*[126]. To overcome this limitation, GGPPsynthase-11 (*At*GGPPS11)[115] from *Arabidopsis thaliana* was overexpressed along with LOS in *E.coli*, and the analysis of *n*-hexane extracts show the production of high levels of lycopaoctaene but very low amounts of squalene and $C_{35}H_{58}$, suggesting robust turnover of FPP to GGPP by *At*GGPPS11 and GGPP to lycopaoctaene by LOS (Figure 40, panel II). Next, we overexpressed GGR-1 and GGR-2 separately in the *E.coli* line producing lycopaoctaene to test their ability to reduce lycopaoctaene into the higher saturated homologues, however, no lycopahexaene, lycopapentaene, lycopatetraene, lycopatriene, or lycopadiene were detected (Figure 41, panel IV and Figure 42, panel IV). Expression of both GGR-1 and GGR-2 slightly decreased the level of lycopaoctaene and increased the level of squalene and $C_{35}H_{58}$ (compare panel II and IV in Figures 41 and 42). It is possible that both GGR-1 and GGR-2 possess typical GGR enzyme activity and reduce GGPP to PPP, thus decreasing the level of the GGPP pool available for LOS to produce lycopaoctaene. Phytyl-PP production in the *E.coli* lines expressing GGR-1 and GGR-2 was not analyzed in these experiments and should be done in near future. It is difficult to speculate at this time on why the levels of squalene and $C_{35}H_{58}$ increases when GGR-1 and GGR-2 were coexpressed with *At*GGPPS11 and LOS. This preliminary data suggests that GGR-1 and GGR-2 may not be the reductase enzyme hypothesized to be involved in the reduction of the lycopaoctaene molecule. Further characterization of

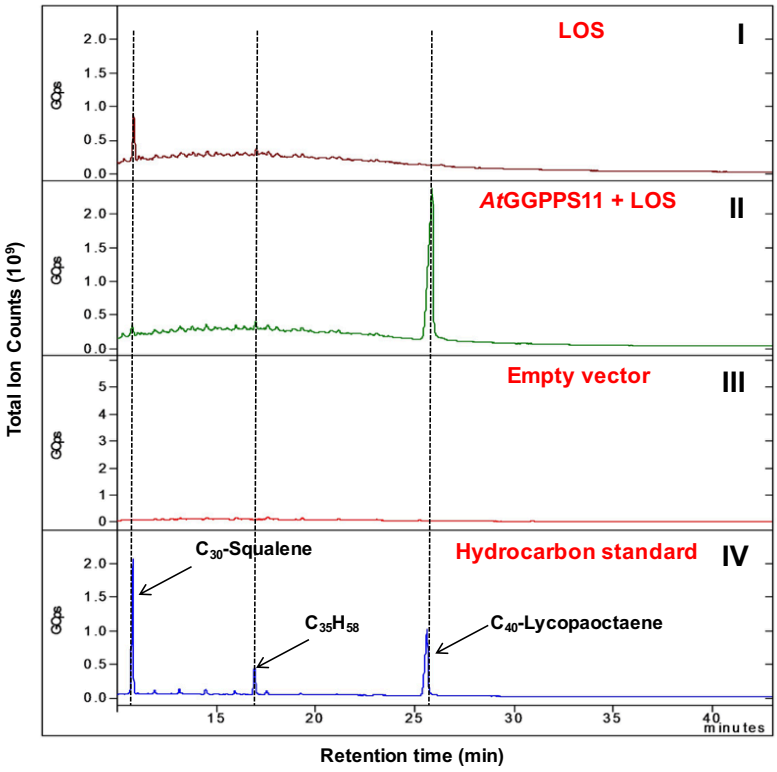


Figure 40. Generation of a lycopaoctaene producing *E.coli* line. Shown are GC-MS profiles of *n*-hexane extractable metabolites of *E.coli* cells expressing different gene constructs. I, LOS expression. **II,** *At*GGPPS11 and LOS expression. **III,** Empty vector expression. **IV,** Hydrocarbon standards. LOS = lycopaoctaene synthase; *At*GGPPS11 = geranylgeranyl diphosphate synthase 11 from *Arabidopsis thaliana*.

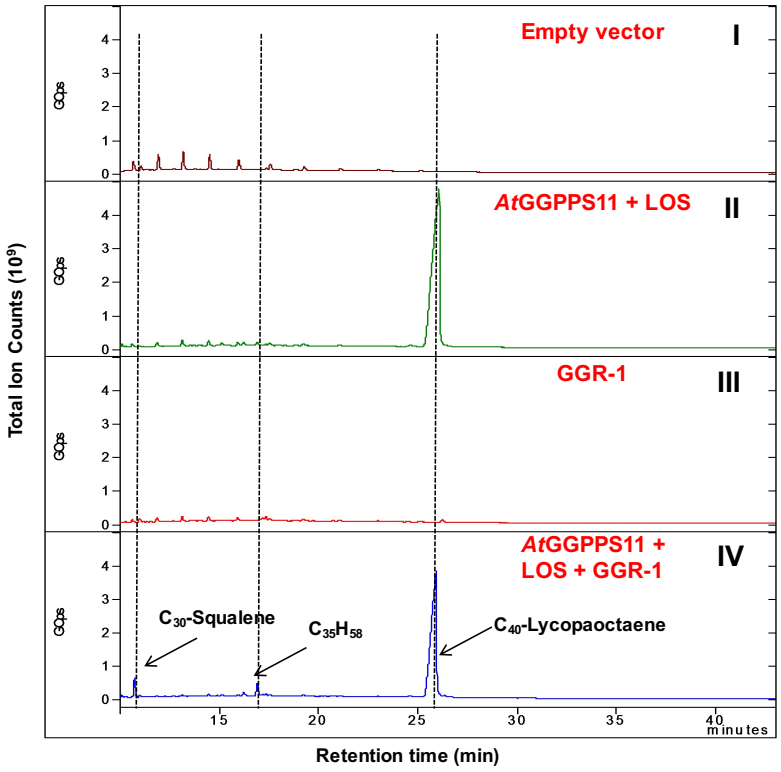


Figure 41. Characterization of GGR-1 in *E.coli*. Shown are GC-MS profiles of *n*-hexane extractable metabolites of *E.coli* cells expressing different gene constructs. I, Empty vector expression. **II,** *At*GGPPS11 and LOS expression. **III,** GGR-1 expression. **IV,** *At*GGPPS1, LOS and GGR-1 expression. *At*GGPPS11 = geranylgeranyl diphosphate synthase 11 from *Arabidopsis thaliana*, LOS = lycopaoctaene synthase and GGR-1 = geranylgeranyl reductase like-1.

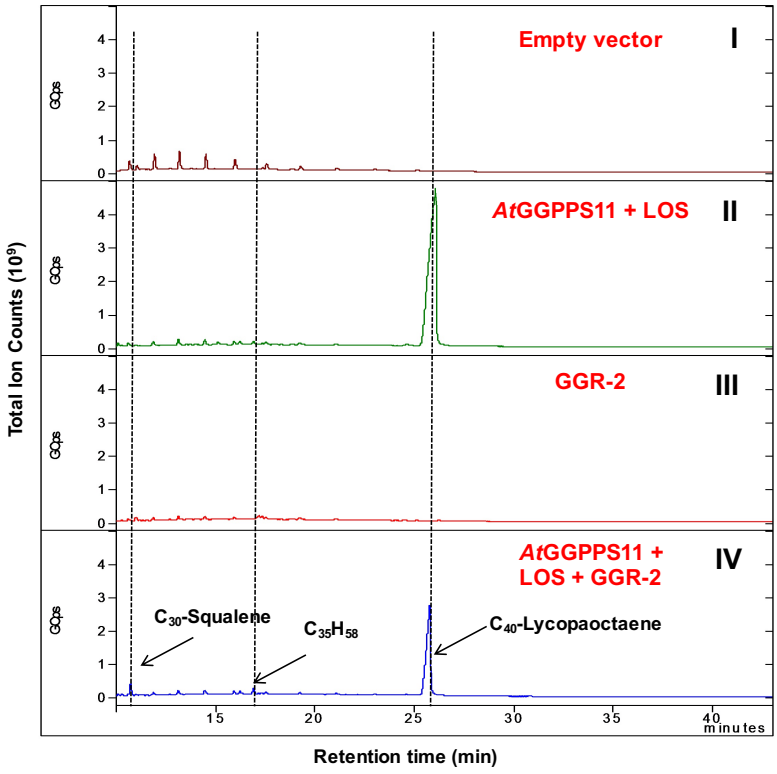


Figure 42. Characterization of GGR-2 in *E.coli*. Shown are GC-MS profiles of *n*-hexane extractable metabolites of *E.coli* cells expressing different gene constructs. I, Empty vector expression. **II,** *At*GGPPS11 and LOS expression. **III,** GGR-2 expression. **IV,** *At*GGPPS1, LOS and GGR-2 expression. *At*GGPPS11 = geranylgeranyl diphosphate synthase 11 from *Arabidopsis thaliana*, LOS = lycopaoctaene synthase and GGR-2 = geranylgeranyl reductase like-2.

GGR-1 and GGR-2 is necessary before ruling them out as a putative reductase enzyme of the lycopadiene biosynthetic pathway.

2.3 Discussion

Our studies identified the first committed step in lycopadiene biosynthesis in race L of *Botryococcus braunii,* and describe a new SS-like enzyme, LOS, that carries out this reaction. Importantly, this enzyme may be used to engineer the effective production of hydrocarbon biofuel feedstocks in other photosynthetic or heterotrophic organisms in the future. This novel enzyme produces the tetraterpenoid lycopaoctaene from GGPP in a reaction analogous to that of squalene synthase. Lycopaoctaene would then undergo sequential reduction by a yet-to-be identified reductase enzyme(s) to form lycopahexaene, lycopapentaene, lycopatetraene, lycopatriene, and finally lycopadiene as the major hydrocarbon product of the pathway (Figure 37). Furthermore, we show that the wild type LOS enzyme displays remarkable substrate flexibility, and can utilize at least three different, naturally occurring C_{15} and C_{20} prenyl diphosphate substrates *in vivo* and *in vitro* for the combinatorial biosynthesis of hydrocarbons belonging to three terpene classes; triterpenes, sesquarterpenes, and tetraterpenes.

When used under artificial reaction conditions, other SS-like enzymes have been described to show relaxed substrate flexibility and to yield products of varying lengths. One such enzyme is CrtM, a dehydrosqualene synthase from *Staphylococcus aureus* that is involved in C_{30} carotenoid biosynthesis[42-44]. In *S. aureus* and also when overexpressed in *E. coli,* CrtM uses two FPP molecules to produce dehydrosqualene, a C_{30} squalene-like molecule with a *cis* double bond at the 1-1' linkage of the two FPP molecules[45].

Expression of a GGPP synthase in *E. coli* also expressing CrtM led to the depletion of FPP and the overproduction of GGPP[42]. Under these conditions, CrtM was found to generate its native C_{30} product, a C_{35} homolog, and C_{40} phytoene, indicating the ability of CrtM to accept GGPP as a substrate when the natural FPP substrate is limiting[42]. CrtM was further engineered through mutations to accept C_{25} farnesylgeranyl diphosphate (FGPP), yielding C_{40}, C_{45} and C_{50} homologs of dehydrosqualene in the presence of FPP, FGPP, and GGPP[127, 128]. In another study, squalene synthases from yeast, pig, and rat have been shown to accept non-natural derivatives of FPP to generate several unnatural terpene products *in vitro*[129, 130]. Additionally, yeast SS has been shown to yield alternative products *in vitro* from FPP as a substrate under various non-physiological reaction conditions, such as in the absence of NADPH with extended incubation times, or in the presence of an unreactive NADPH analog[48-50]. Irrespective of these examples, the promiscuity of LOS is still remarkable, considering that the wild-type LOS enzyme is able to accept three naturally occurring prenyl diphosphate substrates and produce the chimeric hydrocarbons described here without the need for mutations or altered, non-physiological reaction conditions. Such remarkable intrinsic substrate promiscuity has not been documented for any other eukaryotic SS or SS-like enzyme up until now.

Ours and other studies[38,51] raise the question about how SS-like enzymes arise. LSS and LOS from *B. braunii* share >50% sequence identity and both contain the important catalytic residues and conserved domains of SS enzymes (Figure 29), but they still have different substrate preferences and product formation. Thus, we hypothesize

that an ancient SS gene may have undergone gene duplication in race L. The paralogue that gave rise to LSS maintained the ability to produce squalene as its sole product to support sterol biosynthesis for primary metabolism. However, the paralogue that yielded LOS developed the use of alternative substrates for hydrocarbon oil production, while also preserving SS activity. A similar scenario has been suggested as a possible mechanism for generating new secondary metabolites that may provide a fitness benefit to the host organism[127, 131]. For *B. braunii*, it has been proposed that hydrocarbon oils and their derivatives increase the buoyancy of colonies for increased exposure to sunlight[51,53,54]. Similar gene diversifications have also been proposed in the B race of *B. braunii* for botryococcene production, and in some bacteria for squalene generation. For botryococcene production, SS duplications gave rise to the SSL-1 and SSL-3 genes/enzymes responsible for botryococcene production while retaining a residual ability to produce squalene[38]. For bacteria, a three-enzyme system for squalene production arose from successive gene duplication events[51].

LOS produces a variety of terpenes *in vitro* when offered a mixture of substrates, and also do so *in vivo* in race L. In the *in vitro* mixed substrate assays with LOS (Figure 32), a $C_{35}H_{64}$ molecule was observed from FPP and PPP condensation. This product has the exact same GC retention time and mass spectrum as the $C_{35}H_{64}$ hydrocarbon detected in the total hydrocarbon pool (Figure 10, panel VII; Figure 32, panel II). This suggests that LOS is also promiscuous in *B. braunii* cells, but at best has limited access to the FPP and PPP pools since the $C_{35}H_{64}$ product only comprises 0.15% of the total hydrocarbons (Figure 10, panel VII). The FPP-GGPP condensation product $C_{35}H_{58}$ seen in these *in*

vitro assays (Figure 32, panel I) is not detected in the total hydrocarbon pool, possibly because this molecule is reduced to $C_{35}H_{64}$ *in vivo*. It should be noted that two previous studies described several C_{35} squalene-like isoprenoids from *Methanococcus jannaschii* and *Thermococcus barophilus*, including the $C_{35}H_{58}$ and $C_{35}H_{64}$ molecules identified here[55,56], although the enzymatic basis for the biosynthesis of these products remains uncharacterized. Additionally, the ability of LOS to produce lycopapentaene as an observed *in vitro* condensation product of GGPP and PPP (Figure 32, panel III) and accumulation high level of radioactivity during pulse-chase experiment (Figure 27B) may indicate the existence of an alternative *in vivo* biosynthetic route to this molecule, and may be partially responsible for the relative prevalence of lycopapentaene (3% of total hydrocarbons; Figure 10, panel IV) amongst the minor hydrocarbons of race L. This would suggest that LOS can also utilize PPP *in vivo*. However, LOS does not appear to utilize PPP as a single substrate for lycopadiene production *in vivo* as we were not able to detect lycopadiene synthesis activity in cell fractions using PPP as substrate (Figure 26A).

The promiscuous LOS described in this work is nevertheless utilized by *B. braunii* race L to produce lycopaoctaene preferentially over squalene, as shown by the large amounts of hydrocarbon oils produced by this alga, and the absence of detectable amounts of squalene or non-sterol squalene derivatives in the total hydrocarbon pool. In contrast, accumulation of the non-sterol squalene derivatives methylsqualenes in race B are thought to arise from squalene produced by the additional SS-like enzyme SSL-2[38].

Thus, detection of non-sterol squalene derivatives in race L would be expected if LOS was producing squalene *in vivo*.

We hypothesize that the observed preference of LOS for C_{40} lycopaoctaene biosynthesis over C_{30} squalene production *in vivo* may result from an increased flux from GGPP over FPP, as seen for *E.coli* cells producing lycopaoctaene (Figure 40), due to metabolic channeling to form a biosynthetic metabolon specific for lycopadiene production. Formation of such metabolons is common in plants, and is used for channeling substrates and intermediates required for the biosynthesis of different classes of plant natural products, including isoprenoids[57-59]. When expressed in yeast, LOS may produce squalene presumably due to a lack of this proposed metabolon, thus giving LOS access to the FPP pool. We therefore conclude that LOS is responsible for synthesizing lycopaoctaene as the first committed step towards lycopadiene hydrocarbon biosynthesis in race L of *B. braunii*.

2.4 Methods

2.4a Reagents

[1-^{3}H]-FPP (specific activity, 18.2 Ci/mmol) was purchased from PerkinElmer. [1-^{3}H]-GGPP (specific activity, 20.0 Ci/mmol), [1-^{3}H]-phytyl-PP (specific activity, 20.0 Ci/mmol), and nonradiolabeled phytyl-PP were purchased from American Radiolabeled Chemicals. When required, specific activities of radiolabeled chemicals were adjusted with non-labeled chemicals. Solvents and chemicals were purchased from VWR. All other reagents were purchased from Sigma unless otherwise noted.

2.4b Culturing of *Botryococcus braunii*

B. braunii, race L, Songkla Nakarin strain[16] was obtained from Algobank-Caen Microalgal Culture Collection, University of Caen Basse-Normandie, France and grown in a modified Chu 13 medium, pH 7.5[60] at 22°C under continuous aeration of filter-sterilized air enriched with 2.5% CO_2. The concentrations of chemicals in modified Chu 13 medium were as follows; KNO_3 (0.4 g/L), $MgSO_4 \cdot 7H_2O$ (0.1 g/L), K_2HPO_4 (0.052g/L), $CaCl_2 \cdot 2H_2O$ (0.054 g/L), FeNa EDTA (0.01 g/L), H_3BO_4 (2.86 mg/L), $MnSO_4 \cdot H_2O$ (1.54 mg/L), $ZnSO_4 \cdot 7H_2O$ (0.22 mg/L), $CuSO_4 \cdot 5H_2O$ (0.08 mg/L), $NaMoO_4 \cdot 2H_2O$ (0.06 mg/L) and $CoSO_4 \cdot 7H_2O$ (0.09 mg/L). The cultures were grown under a 12:12 hr light:dark cycle with a light intensity of 120 $\mu Em^{-2}s^{-1}$. Algae cells were subcultured by inoculation of 100 mL of a floating six-week-old culture to 750 mL of fresh medium.

2.4c Hydrocarbon staining and microscopy

Nile red (Sigma) staining of L race hydrocarbons was done by treating 500 µL of *B. braunii* L race colonies in medium at early phase (5 day) density with 2.5 µL of a stock solution of Nile red dissolved in acetone (0.15 mg/mL) for a final concentration of Nile red of 0.375 g/mL and acetone of 0.25%. Samples were kept in the dark and incubated at room temperature for 15 min prior to imaging. Confocal microscopy images were obtained using an Olympus IX83-FV1200 inverted confocal microscope with spectral detection unit using an Olympus UPLSAPO 60X oil immersion objective (NA=1.35). The confocal aperture diameter was set to 1 Airy disk unit, and Z-series images were acquired with voxel dimensions of 101x101x490 (nm; X-Y-Z,

respectively). Nile Red and chlorophyll were excited using a 488 nm laser and detected using 540-590 nm and 655-755 nm barrier filter settings, respectively. Bright-field images were acquired via the transmitted detector using the 488 nm laser. Maximum intensity projections and brightness/contrast adjustments were completed using the Olympus FV-ASW software version 4.

2.4d Hydrocarbon purification

Hydrocarbons were purified as described previously[38, 132] and as follows. Algal cells were harvested by vacuum filtration using a 10 μm nylon-mesh filter, freeze-dried, extracted twice with *n*-hexane for 2 hours to recover extracellular hydrocarbons, and then twice with $CHCl_3$:MeOH (2:1) for 12 hours stirring at room temperature to recover intracellular hydrocarbons. Both extracts were concentrated separately using a rotary evaporator and resuspended in a small amount of *n*-hexane before running through separate gravity-fed silica gel columns with *n*-hexane as solvent. The *n*-hexane eluent prior to the pigment front was collected as the hydrocarbon fraction. Both extracellular and intracellular hydrocarbon fractions were combined and evaporated to dryness to recover the total hydrocarbon pool of the algae. Individual hydrocarbon molecules were purified by injecting the total hydrocarbon sample onto an HPLC Develosil 60 silica column (20 mm × 250mm) using *n*-hexane as a mobile phase at a flow rate of 6 ml/min and detection at 210nm. An aliquot of the purified hydrocarbons were analyzed by GC-MS for purity analysis.

2.4e GC-MS conditions

Enzyme assays products, purified hydrocarbons, and yeast extracts were analyzed by GC-MS (Bruker 436-GC-SCION SQ Premium) using a 5% Phenyl BR-5ms capillary column (30 m × 0.25 mm, film thickness: 0.25 μm) in electron ionization (70 eV) mode. Initial oven temperature was 220°C, held for 1 min, then increased to 280°C at the rate of 5°C/min, and then ramped to 300°C at a rate of 2°C/min, and held for 20 min. Helium was used as a carrier gas at a flow rate of 2.58 mL/min. Temperatures of injection port, interface, and ion source were, 280ºC, 250ºC and 200ºC, respectively.

2.4f Hydrocarbon NMR analysis

1D NMR spectra of race L hydrocarbons molecules suffer from severe spectral overlap and the data interpretation is further complicated by degenerate chemical shifts as a result of symmetry planes in the lycopadiene structure. Consequently, hetero-nuclear 2D experiments were used for unambiguous structural characterization of each entity. NMR experiments were performed at 25°C on purified samples dissolved in deuterated chloroform using a 500 MHz Bruker Avance III HD spectrometer equipped with an inverse detection TXI probe. We have relied on 2D hetero-nuclear experiments acquired at natural abundance of ^{13}C nuclei to deduce structures of these molecules. Multiplicity-edited 2D HSQC spectrum was acquired for each compound and the resonances were assigned using a combination of 2D H2BC, 2D HMBC and 2D HSQCTOCSY experiments. The ^{13}C chemical shifts obtained for lycopadiene match those previously obtained[16].

2.4g Spectral data for hydrocarbons analyzed

Lycopadiene: ^{1}H NMR (500 MHz, $CDCl_3$): δ 0.87, 0.89, 1.08, 1.09, 1.16, 1.2, 1.29, 1.39, 1.4, 1.54, 1.61, 1.96, 2.04, 2.17, 5.15; ^{13}C NMR (125 MHz, $CDCl_3$): δ 15.9, 19.7, 22.7, 24.5, 24.7, 25.4, 28.0, 28.3, 31.2, 32.8, 36.8, 37.4, 39.4, 40.0, 124.0, 135.9; MS (*m/z*): [M^+] calcd. for $C_{40}H_{78}$, 558.61; found, 558.70. **Lycopatriene:** ^{1}H NMR (500 MHz, $CDCl_3$): δ 0.87, 0.89, 1.08, 1.09, 1.13, 1.16, 1.2, 1.29, 1.29, 1.39, 1.4, 1.54, 1.61, 1.63, 1.71, 1.96, 1.98, 2.04, 2.16, 5.12, 5.15; ^{13}C NMR (125 MHz, $CDCl_3$): δ 15.9, 17.6, 19.7, 22.7, 24.5, 24.7, 24.7, 25.4, 25.6, 25.7, 28.0, 28.3, 30.8, 32.7, 36.8, 37.1, 37.4, 39.4, 40.0, 124.1, 125.1, 131.1, 135.9; MS (*m/z*): [M^+] calcd. for $C_{40}H_{76}$, 556.59; found, 556.70. **Lycopatetraene:** ^{1}H NMR (500 MHz, $CDCl_3$): δ 0.87, 0.89, 1.08, 1.08, 1.13, 1.16, 1.2, 1.29, 1.39, 1.4, 1.54, 1.61, 1.63, 1.71, 1.95, 1.96, 1.99, 2.04, 2.16, 5.12, 5.15; ^{13}C NMR (125 MHz, $CDCl_3$): δ 15.9, 17.6, 19.7, 22.7, 24.5, 24.7, 25.4, 25.5, 25.7, 28.0, 28.3, 30.8, 32.7, 36.6, 37.1, 37.4, 39.3, 39.4, 40.0, 124.1, 125.2, 131.1, 135.9; MS (*m/z*): [M^+] calcd. for $C_{40}H_{74}$, 554.58; found, 554.60. **Lycopapentaene:** ^{1}H NMR (500 MHz, $CDCl_3$): δ 0.87, 0.89, 1.08, 1.08, 1.16, 1.2, 1.29, 1.39, 1.4, 1.55, 1.62, 1.63, 1.71, 2, 2.04, 2.09, 2.23, 5.15; ^{13}C NMR (125 MHz, $CDCl_3$): δ 16.0, 17.7, 19.8, 22.7, 24.5, 24.7, 25.4, 25.7, 26.7, 28.0, 28.3, 31.0, 32.8, 36.6, 37.4, 39.4, 39.8, 124.2, 131.1, 135.9; MS (*m/z*): [M^+] calcd. for $C_{40}H_{72}$, 552.56; found, 552.70. **Lycopapentaene isomer:** ^{1}H NMR (500 MHz, $CDCl_3$): δ 0.87, 0.89, 1.08, 1.08, 1.13, 1.16, 1.2, 1.29, 1.4, 1.55, 1.62, 1.63, 1.71, 1.99, 2, 2.04, 2.09, 5.1, 5.15; ^{13}C NMR (125 MHz, $CDCl_3$): δ 16.0, 17.7, 19.8, 22.7, 24.5, 24.7, 25.5, 25.7, 26.7, 28.0, 28.3, 32.7, 36.6, 37.1, 37.4, 39.4, 39.8, 124.2, 124.8, 131.1, 135.9; MS (*m/z*): [M^+] calcd. for $C_{40}H_{72}$, 552.56; found, 552.90. **Lycopahexaene:**

^{1}H NMR (500 MHz, $CDCl_3$): δ 0.87, 0.89, 1.08, 1.08, 1.16, 1.2, 1.29, 1.4, 1.55, 1.61, 1.63, 1.71, 1.96, 1.99, 2.04, 2.06, 5.15; ^{13}C NMR (125 MHz, $CDCl_3$): δ 16.0, 17.7, 19.8, 22.7, 24.5, 24.7, 25.7, 26.8, 28.0, 28.3, 32.8, 36.6, 37.4, 39.4, 39.4, 39.8, 124.3, 131.1, 135.9; MS (*m/z*): [M^+] calcd. for $C_{40}H_{70}$, 550.55; found, 550.60. **Lycopaoctaene:** ^{1}H NMR (500 MHz, $CDCl_3$): δ 1.62, 1.63, 1.71, 2, 2.03, 2.04, 5.15; ^{13}C NMR (125 MHz, $CDCl_3$): δ 16.0, 17.7, 25.7, 27.2, 28.3, 39.8, 124.3, 131.1, 135.9; MS (*m/z*): [M^+] calcd. for $C_{40}H_{66}$, 546.52; found, 546.70. **$C_{35}H_{64}$**: ^{1}H NMR (500 MHz, $CDCl_3$): δ 0.83, 0.86, 1.06, 1.13, 1.19, 1.21, 1.24, 1.32, 1.36, 1.52, 1.61, 1.68, 1.98, 2.01, 2.06, 5.13; ^{13}C NMR (125 MHz, $CDCl_3$): δ16.3, 17.9, 20.0, 22.9, 24.4, 24.8, 25.1, 26.0, 26.3, 27.2, 28.2, 33.0, 37.7, 39.6, 40.0, 124.8, 131.4, 135.1; MS (*m/z*): [M^+] calcd. for $C_{35}H_{64}$, 484.50; found, 484.70.

2.4h Ozonolysis experiments

Due to the limited amount of minor hydrocarbons in the total L race hydrocarbon pool, ozonolysis experiments were conducted only on those hydrocarbons that could be purified to obtain at least 10mg; lycopadiene, lycopatriene and lycopapentaene. These hydrocarbons were separately dissolved in dichloromethane and submitted to ozone cleavage for 5 min at -78ºC. Each product from reductive cleavage of the resulting ozonide was directly subjected to GC-MS (electron ionization) analyses without purification. GC-MS analysis was carried out using a GCMS-QP2010 Ultra spectrometer (Shimadzu, Kyoto, Japan) equipped with a capillary column (InertCap 1MS, GL Science; 60 m x 0.25 mm, film thickness: 0.25 μm). The column temperature was programmed as follows: 50ºC for 1 min, raised at 10ºC/min from 50ºC to 220ºC,

then at 2ºC/min from 220ºC to 260 ºC, and held for 22 min at the final temperature. Helium was used as a carrier gas at a flow rate of 41.2 cm/sec. Temperatures of injection port, interface, and ion source were, 260 ºC, 250 ºC and 200 ºC, respectively.

2.4i Preparation of algal cell lysate

Algae from the rapid growth phase were harvested, snap frozen with liquid nitrogen and stored at -80°C for future use. In a typical preparation, 200 mg of frozen algae was added to 8 individual tubes containing 0.8 mL extraction buffer (50 mM Mops (pH 6.8), 20 mM $MgCl_2$, 5mM β-mercaptoethanol, 5 mM EGTA and 20% (v/v) glycerol) with 200 µL of stainless steel metal beads (mixture of 0.9 - 2 mm diameter beads), and then homogenized using a Bullet Blender Storm 24 (Next Advance, Inc., USA) for 10 min at a speed of 10. The algal homogenates were centrifuged at 9,000 × *g* for 10 min at 4°C. The 9,000 × *g* supernatants were pooled and then centrifuged at 100,000 × *g* for 1 hour to obtain the soluble fraction. The pellet was washed with 4 mL: extraction buffer and centrifuged again at 100,000 × *g*. This process was repeated twice, and the pellet was resuspended in 0.5 mL extraction buffer to obtain a microsomal fraction. An aliquot of each protein fraction was used for enzyme assays.

2.4j Enzyme assays

Radioactive enzyme assays were conducted in a 50 µL total reaction volume as described previously[58, 105] and as follows. Reactions were initiated by adding 10 µg algal lysate or 1 µg of recombinant enzyme purified from *E. coli* to a reaction buffer containing 50mM MOPS (pH 6.8), 2.5 mM β-mercaptoethanol, 20 mM $MgCl_2$, 2 mM NADPH or NADH and 10 µM [1-^{3}H]-prenyl-PP substrate (specific activity = 0.25

Ci/mmol). Enzyme assays were incubated at 37°C for 60 minutes and terminated by adding 60 µL of *n*-hexane, followed by brief vortexing and centrifugation. Thirty µL of the organic layer was spotted onto silica gel 60 TLC plates along with authentic standards of squalene (R_f = 0.17), C_{30}-botryococcene (R_f = 0.2), lycopaoctaene (R_f = 0.09) and lycopadiene (R_f = 0.5), and then developed with *n*-hexane. Hydrocarbon standards were visualized by iodine vapor and spots corresponding to authentic standards were scraped and analyzed by liquid scintillation counter to determine radioactive incorporation into the respective hydrocarbons. Squalene synthase and botryococcene synthase assays were conducted using ^{3}H-FPP, whereas lycopaoctaene synthase and lycopadiene synthase assays were conducted using ^{3}H-GGPP and ^{3}H-PPP, respectively. For GC-MS analysis of reaction products, enzyme assays were conducted with nonradiolabeled substrates in glass vials by scaling the total reaction volume and contents to 500 µL. Two separate 500 µL reactions were each extracted twice with 1 mL *n*-hexane, the extracts pooled, dried under a stream of nitrogen gas, resuspended in 100 µL *n*-hexane and a 5 µL aliquot analyzed by GC-MS.

2.4k Hydrocarbon accumulation and LOS activity over growth cycle

Several flasks of race L culture were grown over a 39-day culture period. Twenty-five mL of culture was collected at indicated times on pre-weighed Whatman GF/C filters by vacuum filtration. The filters with algae were dried in an 80°C oven, and total hydrocarbons were extracted using the protocol described above. Similarly, 500 mL of algae cells were also harvested at indicated times by filtering through a 10 µm nylon-

mesh and the samples used to determine LOS enzyme activity over the culture period. LOS enzyme assays were conducted using 10 μg of 9,000 × *g* supernatant.

2.4l ^{14}C-GGOH phosphorylation assays

Geranylgeraniol (GGOH) phosphorylation assays were conducted as described previously[111]. In brief, a 50μL reaction was initiated by adding 50μg of microsomal protein to a reaction buffer containing 5mM Tris-HCl (pH=7.4), 5mM $MgCl_2$, 0.05% CHAPS, 20mM sodium orthovanadate, 100μM ATP, 100μM CTP, 100μM UTP, 100μM GTP and 136μM ^{14}C-GGOH (specific activity 0.015 Ci/mmol). The reactions were incubated for 16 min at 37°C and terminated by adding 200μL of methanol, followed by brief vortexing and centrifugation to collect the supernatant. Reaction products were further extracted with 200μL methanol:water (4:1). The supernatants were pooled, dried under a stream of nitrogen gas, resuspended in 50μL of water-saturated butanol and then analyzed on silica gel 60 TLC plates using isopropanol:NH_4OH: H_2O (6:3:1) as mobile phase. ^{14}C incorporation into products was determined by scraping the spots corresponding to authentic standards of GGP (R_f= 0.59) and GGPP (R_f= 0.26) followed by analysis on a liquid scintillation counter.

2.4m Pulse-chase experiment

Feeding experiments in *B. braunii* race L were conducted as described previously[55, 112, 133]. In brief, a solution containing 9μCi of [1-^{14}C]-geranylgeraniol (GGOH) (specific activity, 0.3 Ci/mol) was equally divided and fed to three separate 625ml 4-day-old race L cultures. After a 45 hour incubation (pulse) with ^{14}C-GGOH, algae cells were filtered using a 10μm nylon-mesh, washed thoroughly with fresh media,

transferred into fresh media containing 19μM non-radiolabelled GGOH, and incubated for three days (chase). During the three days chase, algae cells were harvested at different time points by vacuum filtration using Whatman G/FC filters. The filters with algae cells were freeze-dried and the total hydrocarbons were extracted using the protocol described above. The total hydrocarbon fraction was spotted onto preparative silica gel 60 TLC plates along with the authentic standards indicated above plus those of lycopahexaene (R_f = 0.32), lycopapentaene (R_f = 0.24), lycopatetraene (R_f = 0.32), and lycopatriene (R_f = 0.42). The TLC plate was developed with *n*-hexane and visualized by iodine vapor. Incorporation of ^{14}C into each hydrocarbon molecule was determined by scraping the corresponding spot on the TLC plate followed by analysis with a liquid scintillation counter.

2.4n RNA isolation

Total RNA was isolated as previously described[63] and as follows. For *B. braunii*, liquid-nitrogen frozen samples were pulverized using a Tissuelyser II (Qiagen, Valencia, CA). Approximately 200 mg of this frozen pulverized *B. braunii* tissue, or frozen leaf tissues of *Arabidopsis thaliana* and *Solanum lycopersicum*, were ground in a mortar and pestle with liquid nitrogen. The samples were then added to 1 mL of TRIzol (Life Technologies, Grand Island, NY) and the total RNA was isolated following the manufacturer instructions. The RNA from *A. thaliana* and *S. lycopersicum* were used directly for gene cloning (see below). For the *B. braunii* RNA, contaminating polysaccharides were removed prior to use for gene cloning as follows. The RNA pellet was resuspended in 0.5 mL 2M LiCl, incubated for 5 min, and centrifuged at 12,000 × *g*

for 15 min at 4°C to pellet the total RNA. The polysaccharides remain in the supernatant. This process was repeated until the size of RNA pellet remained constant. The RNA pellet was then dissolved in 0.5 mL 1x TE, extracted with an equal volume of phenol/chloroform/isoamyl alcohol mix (25:4:1), and centrifuged at 12,000 × *g* for 15 min at 4°C. The aqueous supernatant phase was removed and extracted with 0.5 mL of chloroform followed by centrifugation at 12,000 × *g* for 15 min at 4°C. The RNA was precipitated form the aqueous supernatant by adding 0.1 volume 3M sodium acetate and 2.5 volumes of 100 % ethanol, incubated for 20 min at -20°C, and centrifuged at 12,000 × *g* for 15 min at 4°C. The RNA pellet was washed twice with 0.5 mL 70% ethanol, centrifuged at 12,000 × *g* for 15 min at 4°C, the pellet dried using a speedvac, and the RNA resuspended in 50 uL of RNAase free water.

2.4o LSS, LOS, GGR-1 and GGR-2 cloning

RNAseq analysis on the L race of *B. braunii* was carried out as previously reported[63] and as follows. RNA isolated from days 0, 3, 7, 14, 21, and 28 over the 4-week culture cycle were kept separate, paired-end libraries were prepared from each RNA sample, and each sample was sequenced under the Illumina platform. A transcriptome contig library was created using the Trinity software suite. Squalene synthase-like sequences were computationally screened for using this transcriptomic database with race B squalene synthase (BSS)[57] as a query. Similarly, geranylgeranyl reductase (GGR)-like sequences were screened for using the *Arabidopsis thaliana* GGR as a query. Two SS-like (LSS and LOS) and two GGR-like (GGR-1 and GGR-2) cDNAs, were identified. Total RNA was extracted from a Day-3 race L culture and first

strand cDNA was prepared using the SuperScript III first strand synthesis kit (Invitrogen). Primers specific to each cDNA based on the transcriptome sequence were used to amplify the PCR product from first strand cDNA using GoTaq DNA polymerase mix (Promega) for LSS and LOS, and with Phusion DNA polymerase for GGR-1 and GGR-2 followed by cloning into the pGEM-T vector (Promega). Gene specific primers were as follows: For LSS, forward primer 5'-ATGGGGAAGCTACAGGAGGTTTTGAAGC-3' and reverse primer 5'-TCAGGCAAGGCCGCCGCGAAG-3'; for LOS, forward primer 5'-ATGAAGTACACAGATTTCCTTGCGC-3' and reverse primer 5'-TCACACAGTCTTGAGGGCGAG-3'; for GGR-1, forward primer 5'-ATGGCTTCCACAGGCTTGAAC-3'and reverse primer 5'-TTAAGCGGCCACCTTCTCCTC-3'; for GGR-2, forward primer 5'-ATGAGTAGCACTTCTGATAATAATGTGCTCC-3'and reverse primer 5'-CTAATCAGAGGGCTTTACGTAAGCGC-3'.

2.4p *At*GGPPS11 and *Sl*SS cloning

The cDNA sequences for *At*GGPPS11 (NR_142395.1)[30] and *Sl*SS (ACY25092.1) were obtained from the NCBI nucleotide database. Total RNA isolated from the leaves of *Arabidopsis thaliana* and *Solanum lycopersicum* were used for first strand cDNA synthesis and PCR products corresponding to each cDNA were amplified using gene specific primers followed by cloning into the pGEM-T vector. Gene specific primers were as follows: For *At*GGPPS11, forward primer 5'-ATGGCTTCAGTGACTCTAGGTTC-3' and reverse primer 5'-

TCAGTTCTGTCTATAGGCAATG-3'; for *Sl*SS, forward primer 5'-ATGGGAACATTGAGGGCA-3' and reverse primer 5'-CTAAGACCGGCTGCCAAAAAGTTG-3', were used for PCR amplification.

2.4q Protein expression and purification

Except for SSL-1 and SSL-3, DNA sequences encoding the predicted transmembrane domain at the C-terminus of each SS or SS-like protein were deleted, and then cloned to pET28a to encode for N-terminal 6xHis-tagged proteins. DNA templates of BSS, SSL-1, SSL-2 and SSL-3 were received from previous studies[25,38]. The pET28a expression constructs were made using appropriate restriction sites; $LOS^{\Delta 392\text{-}444}$ with NheI and HindIII; $LSS^{\Delta 399\text{-}462}$, $SlSS^{\Delta 387\text{-}411}$, $BSS^{\Delta 399\text{-}461}$, SSL-1, and SSL-3 with *Nde*I and *Hind*III; SSL-$2^{\Delta 392\text{-}465}$ with *Nde*I and *Sal*I. The expression constructs were transformed into *E. coli* BL21(DE3), grown at 37°C to OD_{600} = 0.8, and protein expression for each gene induced by adding 1mM isopropyl-β-D-1-thiogalactopyranoside. The induced cultures were then grown for an additional 6 hours at 25°C.

His-tagged proteins were purified at 4°C by standard procedures. In a typical purification, pellets from a 100mL culture were resuspended in 10mL extraction buffer (50 mM sodium phosphate buffer (pH=7.8), 300 mM NaCl, 10 mM imidazole, 1x general protease inhibitor cocktail (Sigma), 1 mM $MgCl_2$, and 1% glycerol (v/v)) and then sonicated four times for 15 sec at 70% maximum power with 2 min interval between each sonication. The sonicated samples were centrifuged at 16,000 × *g* for 10 min at 4°C. The supernatants were applied to a gravity-fed column containing Ni-NTA

agarose (QIAGEN) and the His-tagged proteins were purified according to the manufacturer's recommendations. The wash buffer (50 mM sodium phosphate buffer (pH=7.8), 300 mM NaCl, 60 mM Imidazole, 1 mM $MgCl_2$, and 1% glycerol (v/v)) and the elution buffer (50 mM sodium phosphate buffer (pH=7.8), 300 mM NaCl, 400 mM imidazole, 1 mM $MgCl_2$, and 1% glycerol (v/v)) were used for protein purification. The eluted fractions were dialyzed using storage buffer (300 mM NaCl, 20 mM Tris-HCl (pH=7.5), 5 mM DTT and 2 mM $MgCl_2$), concentrated with an Amicon Ultra centrifugal filter (0.5 mL, 30 kDa cutoff; EMD Millipore) to desired protein concentration, an equal amount of 100% glycerol added, and stored at -20°C for 1-2 months without a loss of enzyme activity.

2.4r LOS steady-state kinetic experiments

Michaelis-Menten enzyme kinetics experiments with the LOS enzyme were set up using the protocol described previously[64] and as follows. Pilot experiments were initially conducted using radioactive enzyme assays to determine the reaction conditions where reaction velocity is linear (less that 10% turnover) with respect to enzyme concentration. For steady-state kinetics experiments, enzyme assays (50 µL total reaction volume) were conducted with 100 nM of purified LOS enzyme at the indicated concentration of ^{3}H-GGPP, ^{3}H-FPP, or ^{3}H-PPP for a set time interval. Kinetics for the NADPH cofactor were not determined and this factor was held constant at 2 mM in all assays. The reaction velocities were plotted against substrate concentrations to generate Michaelis-Menten curves and the kinetics parameters of LOS enzyme for GGPP, FPP, PPP were determined by analyzing the data using GraphPad Prism 6 software.

2.4s Yeast expression

The DNA sequence encoding the 56 N-terminal amino acids for the plastid-targeting signal of *At*GGPPS11 were deleted resulting in *At*GGPPS11$^{\Delta 57}$, which was cloned into pESC-TRP using BamHI and SalI restriction sites for expression under the inducible GAL1 promoter. For coexpression studies, the DNA sequences encoding the predicted C-terminal transmembrane domain of LOS was deleted resulting in LOS$^{\Delta 392\text{-}444}$, which was cloned into the second multiple cloning site of *At*GGPPS11$^{\Delta 57}$:pESC-TRP using EcoRI and SpeI restriction sites for expression under the inducible GAL10 promoter. The yeast expression constructs, *At*GGPPS11$^{\Delta 57}$:pESC-TRP and *At*GGPPS11$^{\Delta 57}$ + LOS$^{\Delta 392\text{-}444}$:pESC-TRP, were introduced into yeast strain CKY457 (*MAT***a**, leu2Δ1, *ura3-52*, *trp1Δ63*, *his3Δ200*, *lys2-128∂*) via lithium acetate transformation followed by selection on yeast synthetic drop-out medium (SC-TRP). Positive transformants were grown at 30°C in 150 mL SC-TRP media to mid-log phase, induced with 2% final galactose concentration, and grown for an additional 130 hours. Yeast cells were harvested, freeze-dried, and extracted with *n*-hexane for 2 hours by stirring at room temperature. The organic extracts were centrifuged at 1,000 × *g*, the supernatant dried using a rotary evaporator, resuspended in 500 µL *n*-hexane, and a 5µL aliquot analyzed by GC-MS to evaluate for the production of lycopaoctaene.

2.4t Purification of lycopaoctaene standard

A 3-liter culture of yeast strain CKY457 expressing *At*GGPPS11$^{\Delta 57}$ + LOS$^{\Delta 392\text{-}444}$:pESC-TRP was grown and the metabolites were extracted with *n*-hexane using the protocol described above. The organic extracts were applied to a silica gel gravity-fed

column, and metabolites eluted sequentially with two different solvent systems, *n*-hexane and 1% diethyl ether in *n*-hexane. The 1% diethyl ether in *n*-hexane eluent fraction containing lycopaoctaene and other metabolites was concentrated and resuspended in a small volume of *n*-hexane before injecting the samples onto a 10 mm × 250 mm Cosomil $5C_{18}$-AR-II HPLC column. Lycopaoctaene was purified by RP-HPLC using methanol:acetone (60:40) as a mobile phase at a flow rate of 3 mL/min and detection at 210 nm. Identity of lycopaoctaene molecule was confirmed by GC-MS (Figure 13) and NMR (Table 1 & Figure 25).

2.4u LSS and LOS complementation of yeast SS knockout strain

The yeast strain ZX 178-08 (*MAT***a**, *his3, leu2, met15, ura3, SUE, erg9Δ::HPH*) which has a knockout of the endogenous squalene synthase gene (ERG9) was used for complementation experiments[65]. The expression vector XURA used in our study was made by replacing the GAL1 and GAL10 promoters in pESC-URA vector backbone with constitutive promoters TEF1 and GPD, respectively[66]. Full-length LSS was cloned into XURA using Not1 and SpeI restriction sites. Full-length LOS was cloned into EcoRI and SpeI sites of XURA vector. The expression constructs, LSS:XURA, LOS:XURA and XURA were transformed into ZX 178-08 yeast strain via lithium acetate transformation followed by selection on yeast synthetic drop-out medium supplement with ergosterol (SCE-URA). The positive transformants were further characterized by streaking on selection media with or without ergosterol to test the ability of individual genes to restore ergosterol prototrophy of ZX 178-08 yeast.

2.4v GC-MS analysis of prelycopaoctaene alcohol (PLOH)

The enzyme assay for prelycopaoctaene diphosphate (PLPP) was conducted following the protocol described previously for analysis of presqualene diphosphate[38,51] and as follows. The PLPP assay contained 50 mM MOPS (pH 6.8), 2.5 mM β-mercaptoethanol, 20 mM $MgCl_2$, 180 μM GGPP and 750 μg of purified LOS enzyme in a 2.5 mL total reaction volume. The reaction mixture was incubated at 37°C for 2 hours and terminated by snap freezing with liquid nitrogen followed by lyophilization. For hydrolysis of PLPP to PLOH, the white residue obtained after lyophilization was resuspended in 2 mL phosphatase solution (20% 1-propanol, 100mM sodium acetate pH 4.7, 0.1% Triton-X-100, 50 units of sweet potato acid phosphatase) and then incubated for 16 hours in a 28°C shaker. The dephosphorylation reaction mixture was extracted three times with 4 mL of MTBE, the extracts pooled, dried under a stream of nitrogen gas, resuspended in 100 μL *n*-hexane and a 5 μL aliquot analyzed by GC-MS as under the conditions described above with the following differences. Chemical ionization using methane gas at 20 psi was employed for analysis of the PLOH molecule as electronic ionization did not result in a molecular ion of PLOH. Initial oven temperature was 70°C, held for 1 min, then increased to 200°C at the rate of 8°C/min, and then ramped to 300°C at a rate of 20°C/min, and held for 20.75 min. Temperatures of injection port, interface, and ion source were, 260ºC, 250ºC and 200ºC, respectively.

2.4w Hydrocarbon production analysis in *E.coli*

The expression constructs, $LOS^{\Delta 392\text{-}444}$:pET28a and *At*$GGPPS11^{\Delta 57}$:pET22b were made using the NheI and HindIII, and NdeI and SalI restriction sites, respectively.

The DNA sequences encoding the 57 N-terminal amino acids of GGR-1 for the chloroplast targeting signal was deleted resulting in GGR-$1^{\Delta 58}$, which was then cloned into the pACYCDuet-1 vector using the BamHI and HindIII restriction sites. Similarly, the DNA sequences corresponding to a signal peptide and the N-terminal transmembrane domain were deleted from GGR-2 to yield GGR-$2^{\Delta 33}$, which was then cloned into the pACYCDuet-1 vector using the BamHI and HindIII restriction sites. The expression constructs, pET28a, *At*GGPPS11$^{\Delta 57}$:pET22b, LOS$^{\Delta 392\text{-}444}$:pET28a, GGR-$1^{\Delta 58}$:pACYCDuet-1 and GGR-$2^{\Delta 33}$:pACYCDuet-1 were transformed into *E.coli* BL21 (DE3) cells and the positive transformants were selected using standard procedures. The *E.coli* line harboring the desired construct(s) was grown in 100 ml of terrific broth media at 37°C to OD_{600} = 0.8 and expression of the corresponding genes was induced by adding 1mM isopropyl-β-D-1-thiogalactopyranoside (IPTG). After induction, cells were grown for 6 hrs at 25°C, harvested by centrifugation, snap frozen with liquid nitrogen and lyophilized. Freeze-dried cells were extracted with *n*-hexane by stirring with a magnetic stir bar in a glass beaker for 2 hrs. The *n*-hexane extracts were centrifuged at 1,000 × *g*, the supernatant dried, resuspended in 200µL *n*-hexane, and a 5µL aliquot analyzed by GC-MS to evaluate the production of hydrocarbons.

CHAPTER III

SUBSTRATE AND PRODUCT SPECIFICITY FOR A TETRATERPENOID SYNTHASE FROM *BOTRYOCOCCUS BRAUNII* RACE L

3.1 Background and rationale

Isoprenoids (a.k.a. terpenes) are one of the largest groups of structurally diverse natural products produced by all domains of life[63]. All isoprenoids are biosynthesized *via* either the mevalonate pathway or the methylerythritol phosphate (MEP) pathway depending on the organism in question, and based on the number of C_5 isoprene units present isoprenoids are classified as monoterpenes (C_{10}), sesquiterpenes (C_{15}), diterpenes (C_{20}), triterpenes (C_{30}) or tetraterpenes (C_{40})[63]. Beside critical roles in primary and secondary metabolism, isoprenoids and their derivatives have a wide range of applications including use as medicines, nutraceuticals, agricultural chemicals, fragrances, flavorings, and industrial chemicals[65]. More recently, there has been significant interest in the use and development of isoprenoid-based chemicals as a direct substitute for petroleum-derived fuels[34, 134]. The highly branched and cyclic nature of isoprenoids provides these molecules with properties such as high energy density, high ignition stability, and a low freezing temperature, which are physiochemical characteristics desired in an ideal combustion fuel[34, 134]. However, low yields in natural hosts have directed isoprenoid-based biofuel research toward metabolic engineering for overproduction of isoprenoid chemicals in industrial microorganisms such as yeast and *E. coli*[135]. For example, the sesquiterpene farnesene, a potential diesel and jet-fuel

alternative, has been commercially produced by utilizing metabolically engineered yeast and sugarcane-derived carbohydrates as an energy source[134].

The colony-forming green microalga *Botryococcus braunii* is an exception to low isoprenoid yielding organisms as it naturally produces large amounts, 30-50% of dry weight, of liquid hydrocarbon oils, which are stored in the colony extracellular matrix (Figure 1)[36]. Furthermore, these hydrocarbons can be catalytically cracked to produce petroleum-equivalent combustion engine fuels[4, 136], thus making *B. braunii* hydrocarbons a promising renewable biofuel feedstock[4, 5]. There are three chemical races of *B. braunii*, each producing distinct types of hydrocarbons (Figure 2)[36]. Race A produces fatty acid derived C_{23}-C_{33} alkadienes and alkatrienes (Figure 2)[41]. Race B and race L produce isoprenoid hydrocarbons derived from the MEP pathway[36], with race B producing the triterpene C_{30}-C_{37} botryococcenes and C_{31}-C_{34} methylsqualenes[39-41], and race L, the focus of this study, producing the C_{40} tetraterpenoid hydrocarbon lycopadiene as the predominant hydrocarbon (Figure 2)[42, 132].

The lycopadiene biosynthetic pathway has been recently elucidated[59]. Early studies suggested lycopadiene could be generated *via* a squalene synthase (SS) like reaction since lycopadiene resembles a C_{40} version of the triterpene squalene (Figure 43)[42, 132]. The SS enzyme carries out the head-to-head condensation of two C_{15} farnesyl diphosphate (FPP) molecules in a two-step reaction to form squalene for use in sterol production (Figure 43)[97, 137]. The first step of the SS reaction generates the intermediate presqualene diphosphate (PSPP), which is reductively rearranged to squalene in the

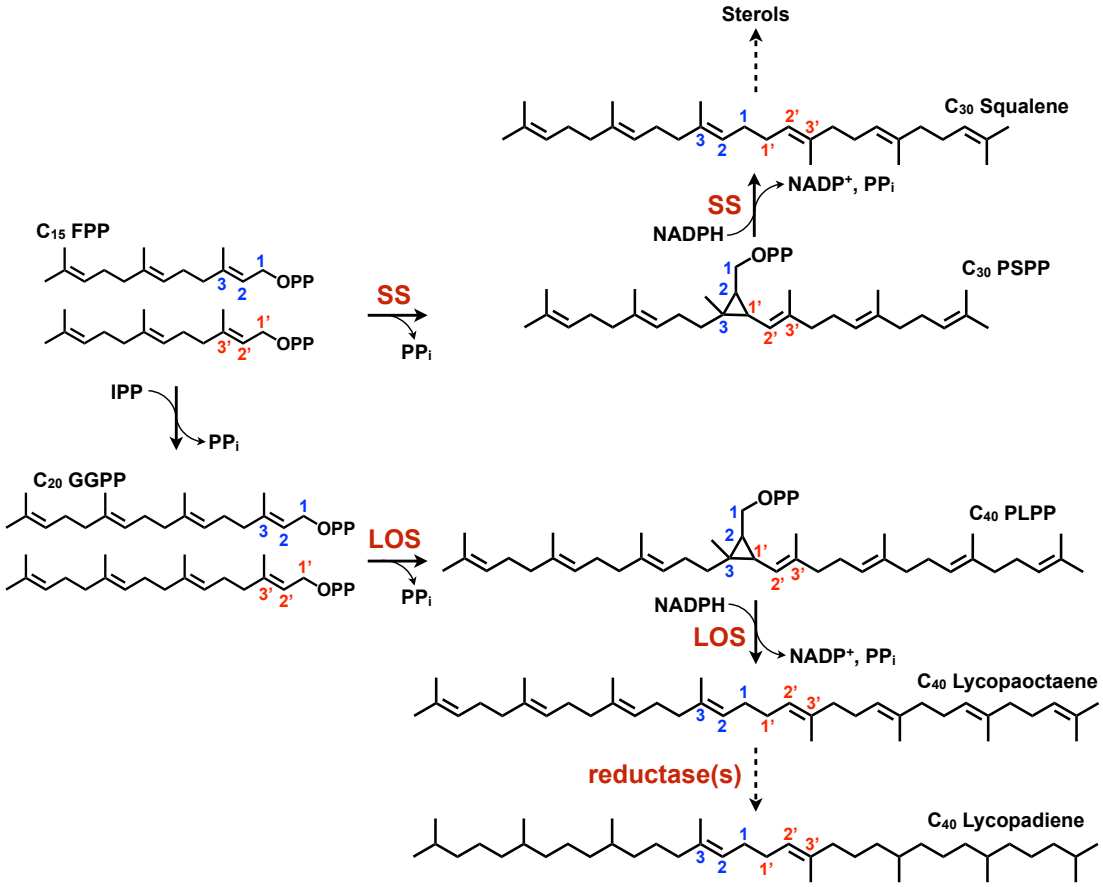


Figure 43. Squalene and lycopadiene biosynthesis in race L. Two molecules of farnesyl diphosphate (FPP) are condensed by squalene synthase (SS) to produce the presqualene diphosphate (PSPP) intermediate followed by reductive rearrangement to produce squalene. Lycopaoctaene synthase (LOS) catalyzes the condensation of two molecules of geranylgeranyl diphosphate (GGPP) to produce the prelycopaoctaene diphosphate (PLPP) intermediate followed by reductive rearrangement to produce lycopaoctaene, which is subsequently reduced to produce lycopadiene.

second reaction step (Figure 43)[97, 137]. An SS-like enzyme called lycopaoctaene synthase (LOS) was identified from the L race of *B. braunii* that, like SS, catalyzes a two-step reaction where two molecules of C_{20} geranylgeranyl diphosphate (GGPP) undergo head-to-head dimerization to produce the prelycopaoctaene diphosphate (PLPP) intermediate in the first step, followed by reductive arrangement to produce C_{40} lycopaoctaene in the second step (Figure 43)[59]. Production of lycopaoctaene represents the committed step in lycopadiene biosynthesis, and is subsequently reduced to lycopadiene by a yet to be identified enzyme(s) (Figure 43)[59].

The LOS enzyme also shows unusual substrate promiscuity not previously observed in eukaryotic SS or SS-like enzymes[59]. LOS is capable of utilizing three prenyl diphosphate substrates FPP, GGPP, and C_{20} phytyl diphosphate (PPP) individually or in combination to produce six different hydrocarbons with C_{30}, C_{35}, and C_{40} chain lengths (Figure 32)[59]. For example, when FPP and GGPP are provided as equimolar substrates LOS produces three of these hydrocarbons: squalene from combining two FPP molecules, $C_{35}H_{58}$ from joining one FPP and one GGPP, and lycopaoctaene from the condensation of two GGPP molecules (Figure 32)[59]. The remaining three hydrocarbons are produced when LOS is incubated with FPP and PPP, or GGPP and PPP (Figure 32)[59]. This discovery provides a unique opportunity to engineer the LOS enzyme for robust production of individual hydrocarbons with specific chain lengths for possible industrial uses such as fuel production.

In order to accomplish this, the mechanism underlying LOS substrate promiscuity needs to be understood. It has been proposed that LOS and the *B. braunii* L race SS

(LSS) arose from the duplication of an ancient *SS* gene in this alga[59]. LSS maintained the original SS activity to provide carbon flux towards sterol biosynthesis, while LOS evolved the novel catalytic function to produce lycopaoctaene leading to hydrocarbon production[59]. The promiscuous activity of LOS is likely due to preservation of the original SS activity following the gene duplication event[59]. Thus, conversion of LOS to an SS enzyme, or an enzyme that produces only one C_{30}, C_{35}, or C_{40} hydrocarbon product for potential industrial use could be accomplished by understanding how LOS binds substrate for product formation. In this chapter, we pursued X-ray crystallography to obtain the structure of LOS protein for structure-function study. Furthermore, we used protein alignments between LOS and SS proteins, as well as a comparison of a predicted LOS 3D-strucure with the human SS (HSS) crystal structure to conduct rational site-directed mutagenesis in order to identify the critical LOS residues important for conferring substrate and product specificity.

3.2 Results and discussion

3.2a Purification of LOS enzyme for protein crystallization

3.2a1 Structural comparison of LOS with HSS

Among many eukaryotic SS enzymes, HSS has been extensively studied due to consideration as a drug target for reducing cholesterol in humans[92-94, 138]. Initial attempts to obtain soluble and catalytically active form of HSS faced many challenges as HSS is an endoplasmic reticulum associated enzyme, and contains a C-terminal membrane spanning domain[138]. Although deletion of the C-terminal membrane domain of HSS yielded catalytically active protein in *E.coli* extracts, purification of highly soluble,

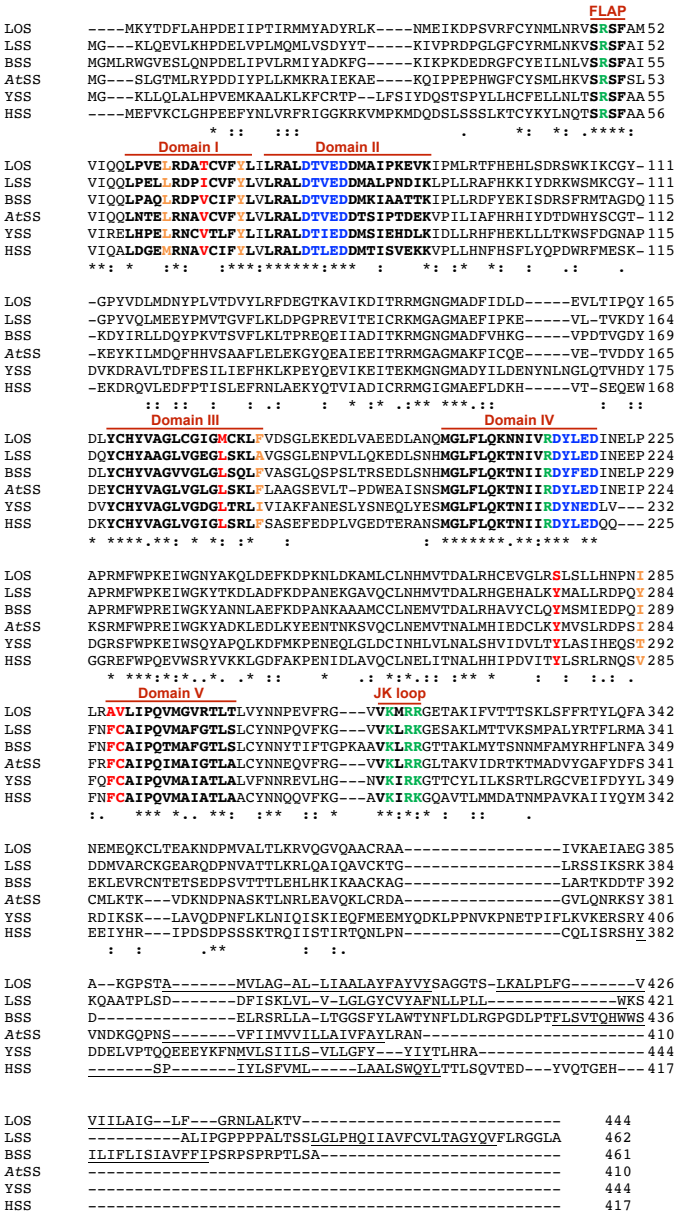


Figure 44. Amino acid sequence alignment of LOS with other SS proteins. Conserved domains and regions found in a typical squalene synthases are highlighted indicated above the sequence. The substrate binding motifs are shown in blue, NADPH binding residues are highlighted in green, and the residues of LOS chosen for mutagenesis and the corresponding residues in the SS proteins are shown in red and orange. Transmembrane domain predicted by TMpred for each protein are underlined. SS, squalene synthase; LOS, lycopaoctaene synthase; LSS, SS from *Botryococcus braunii* race L; BSS, SS from *Botryococcus braunii* race B; *At*SS, SS from *Arabidopsis thaliana*; YSS, SS from *Saccharomyces cerevisae*; and HSS, Human SS.

pure and active HSS required removal of 30 amino acids at the N-terminus and 47 amino acids at the C-terminus[138]. Several crystal structures of HSS have now been solved, but the crystallization of double truncated form of HSS (HSS^{31-370}) protein was only possible in the presence of either farnesyl thiodiphosphate (FSPP; an analog of substrate FPP), PSPP intermediate or inhibitors of the SS reaction[92-94, 138]. As mentioned earlier, LOS is an SS-like enzyme and shares high amino acid sequence similarity at the conserved catalytic regions of HSS enzyme (Figure 44). Thus, our approach to obtain protein crystal of LOS is similar to that used for the HSS^{31-370} protein. LOS encodes 444 amino acids with two predicted transmembrane domains at the C-terminus, one of the domain corresponds to amino acids 392 to 409 and the other corresponds to amino acids 420 to 441 (Figure 44). The three-dimensional structure of LOS was generated using the Iterative Threading ASSEmbly Refinement (I-TASSER) bioinformatics program[139-141]. The predicted 3-D structure shows LOS as an α-helical protein with several regions matching previous SS structures such as a large central catalytic channel running through the center of the protein (dashed box in Figure 45A), an N-terminal flexible region (green in Figure 45A) and a C-terminal transmembrane domain (magenta in Figure 45A).

As we previously reported[59], removal of the 53 C-terminal amino acids encoding the transmembrane domain, producing LOS^{1-391}, was sufficient to achieve soluble, pure and catalytically active form of this enzyme. In order to match the HSS^{31-370} structure, the first 32 N-terminal amino acids of LOS^{1-391} were truncated plus an additional nine C-

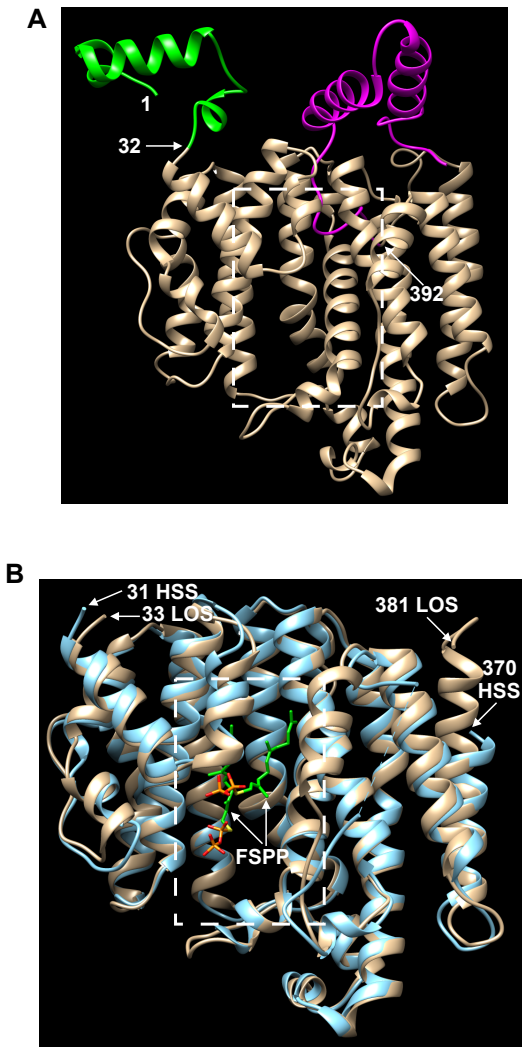


Figure 45. 3D-Structures of LOS and HSS. A. A model of LOS structure as predicted by the I-TASSER program. The N-terminal flexible region is shown in green and a double pass transmembrane domain predicted by TMPred software is shown in magenta. **B.** Comparison of the crystal structure of $HSS^{31\text{-}370}$ (light blue) bound to two molecules of farnesylthiol diphosphate (FSPP) and a computationally truncated model of $LOS^{33\text{-}381}$(light brown). LOS = lycopaoctaene synthase and HSS = human squalene synthase.

terminal amino acids were removed to generate LOS^{33-381} (Figure 45B). The comparison of the FSPP bound HSS^{31-370} structure with that of the predicted LOS^{33-381} structure showed that both structures aligned well with very close matches of all α-helices and the active site (Figure 45B). To check whether LOS^{33-381} is biologically active, two constructs of LOS designed to encode for 6xHis-LOS^{1-391} and 6xHis-LOS^{33-381} were made, expressed in *E.coli*, protein samples purified and the enzymatic activity tested. As shown in Figure46A, both LOS^{1-391} and LOS^{33-381} were successfully purified, but their apparent protein size as analyzed by SDS-PAGE is smaller than the predicted molecular weight. It should be noted that purified HSS also migrated lower than its predicted molecular weight when analyzed by SDS-PAGE[138]. When the LOS enzyme assay was conducted with purified LOS^{1-391} and LOS^{33-381}, both samples successfully produced lycopaoctaene using GGPP as a substrate (Figure 46B). Similarly, when activities were tested using the FPP/GGPP mixed substrate assay, both LOS protein samples produced hydrocarbons of all possible combinations; squalene, $C_{35}H_{58}$ and lycopaoctaene (Figure 46C). Although the absolute values of enzymatic activities of LOS^{33-381} are lower when compared to LOS^{1-391}, both LOS constructs can be purified and are catalytically active, and will be used for future crystallization experiments (Figure 46).

3.2a2 Purification of 6xHis-LOS^{1-391} protein for crystal screening

Unlike enzyme assays, protein crystallization requires large amounts (~10 mg) of highly pure protein as most of the protein sample will be used to screen conditions optimal for obtaining crystals of the desired protein. For optimization of large-scale

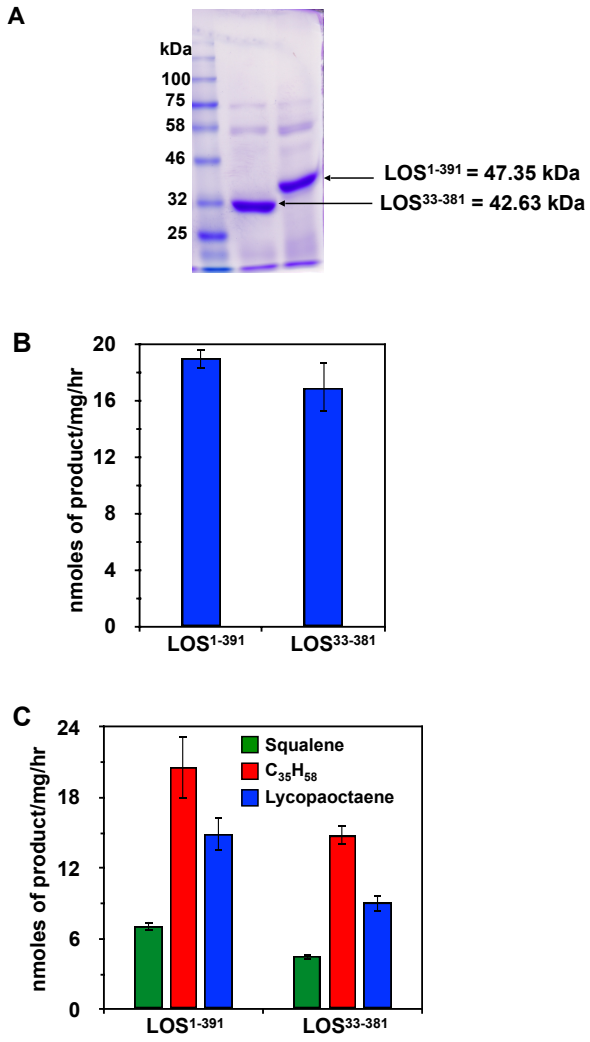


Figure 46. Characterization of LOS$^{1-391}$and LOS$^{33-381}$ proteins. A. Approximately 5µg of purified protein for each sample was analyzed by SDS-PAGE. Predicted molecular weight for 6xHis-LOS$^{1-391}$and 6xHis-LOS$^{33-381}$ is shown. **B.** LOS enzyme assay for lycopaoctaene production using LOS$^{1-391}$ and LOS$^{33-381}$ with GGPP given as a substrate. **C.** Enzyme activities of LOS$^{1-391}$ and LOS$^{33-381}$ in the FPP/GGPP mixed substrate assay showing production of squalene, $C_{35}H_{58}$, and lycopaoctaene. GGPP = geranylgeranyl diphosphate and FPP = farnesyl diphosphate.

protein purification, we decided to focus on the construct encoding 6xHis-LOS$^{1-391}$. Initial purification of the LOS$^{1-391}$ protein was conducted to determine the optimal imidazole concentration required in the washing step of protein purification. Four different imidazole concentrations (20, 30, 40 and 50 mM) were tested, and the purification for each sample was conducted using the *E.coli* cell pellet from a 200 mL culture expressing the LOS$^{1-391}$ construct (Figure 47A). Samples washed with 40 mM and 50 mM imidazole yielded protein with higher purity despite a significant loss in protein yield when compared to treatment with 20 mM and 30 mM imidazole (Figure 47A). Although there is no significant difference in the purity between the 40 mM and 50 mM imidazole washes, protein yield was reduced by more than 60% when treated with the 50 mM imidazole wash (Figure 47A). In order to obtain a good yield along with pure protein, the 40 mM imidazole wash step was used in future purification experiments.

Next, the protein purification was scaled up using the cell pellet from a 750 mL *E.coli* culture. About 2.9 mg of purified protein from nickel affinity chromatography was loaded into a gel filtration column (Figure 47B). Two peaks were observed, the large peak with earlier elution time corresponded to aggregated protein, and the small peak corresponded to the LOS monomer (Figure 47B). Analysis by SDS-PAGE showed the aggregation sample contained LOS protein as the major component and a small amount of other contaminants. In case of the monomer peak, the LOS protein is relatively pure except for the presence of a 25kDa protein band (Figure 47B). Protein recovery for the

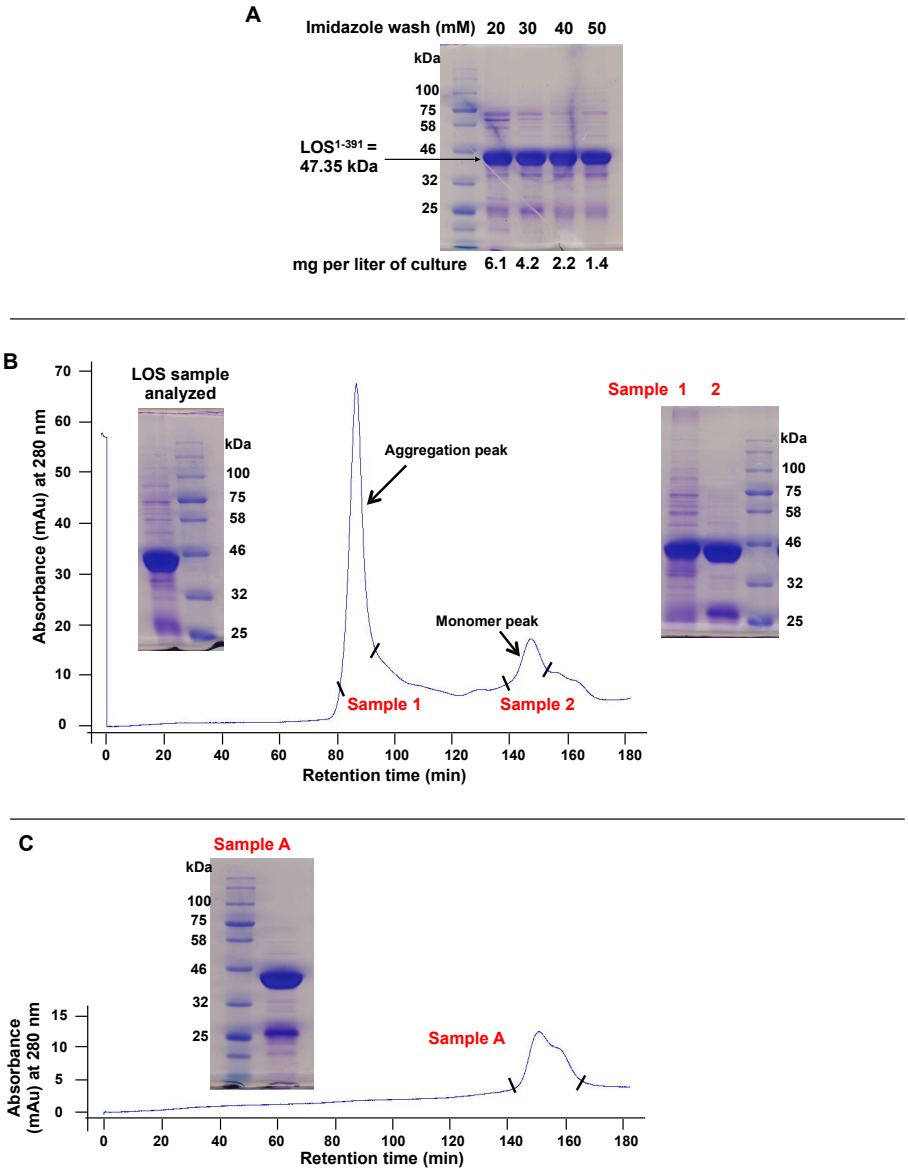


Figure 47. Purification of 6xHis-LOS$^{1-391}$. Approximately 20μg of purified protein was analyzed by SDS-PAGE. **A.** LOS protein purification optimization using different concentrations of imidazole in the wash buffer. The amount of purified protein recovered after each wash treatment is expressed as mg of protein per liter of *E.coli* culture. **B.** Protein purified from nickel affinity chromatography as analyzed by SDS-PAGE on the left was loaded into the gel filtration column. Two samples were collected and analyzed by SDS-PAGE as shown in right. **C.** Sample 1 from **B** was loaded onto the gel filtration column, sample A was collected and then analyzed by SDS-PAGE gel. LOS = lycopaoctaene synthase.

LOS monomer sample from the gel filtration column was 20%. To check whether the LOS monomer sample was stable over time, sample 2 from figure 47A was stored at -80°C for two weeks, and then loaded onto a gel filtration column (Figure 47C). The LOS monomer sample did not show any aggregation peak suggesting it is stabile over time (Figure 47C). Interestingly, a shoulder peak was observed for sample A in the gel filtration analysis and may correspond to the small protein band observed around 25 kDa in the SDS-PAGE analysis (Figure 47C). Further analysis of sample A by western blot using an α-His antibody detected this 25 kDa band suggesting it is a C-terminal degradation product of the 6xHis-LOS$^{1-391}$ protein (data not shown). It should be noted that the 25 kDa band is present in all purification steps and becomes more prevalent after purification through gel filtration (Figure 47). When purification was scaled up using the cell pellet from a ~1.5 liter *E.coli* culture, LOS purity decreased (data not shown). Due to this reason, we repeated purification using the cell pellet from a ~750 mL *E.coli* culture. This was carried out five independent times and generated 3.3 mg of the LOS$^{1-391}$ monomer sample, which was given it to our collaborator Su Tang (Dr. James Sacchettini's lab at Texas A&M University) for crystal screening. Unfortunately, screening under 480 crystallization conditions did not yield protein crystals. It is possible that the small band of LOS protein could have interfered the crystallization process.

The LOS enzyme shows optimal enzyme activity in a MOPS based reaction buffer[59]. However, the large scale purification of LOS$^{1-391}$ conducted so far utilized a sodium phosphate based buffer. The detail information on the composition of sodium phosphate and MOPS based buffer is provided in the method section of this chapter. In

order to test the effect of a MOPS based buffer on protein purification, we conducted nickel affinity chromatography purification using a cell pellet from a ~750 mL *E.coli* culture and analyzed the purified protein using gel filtration (Figure 48A). As shown in the gel filtration chromatograph (Figure 48A), the MOPS based buffer did not change the profile of aggregated versus LOS monomer peaks, but shifted the shoulder peak of the LOS monomer peak towards left (sample 1) when compared to the shoulder peak observed on the right in the previous purification with the sodium phosphate based buffer (compare figures 47C and 48A). We also collected the fractions corresponding to the left shoulder (sample 1) and the main peak (sample 2), and then analyzed the protein samples by SDS-PAGE (Figure 48A). Interestingly, sample 1 contained 6xHis-LOS$^{1\text{-}391}$ as a major protein band and the 25 kDa small protein band corresponding to a C-terminal degraded or truncated form of LOS (Figure 48A). It seems that the 25 kDa protein band of LOS interacts with 6xHis-LOS$^{1\text{-}391}$ protein, and eventually results in the shoulder peak and earlier elution of sample 1 from the gel filtration column (Figure 48A). In the case of sample 2, the 25 kDa protein band was not present and the 6xHis-LOS$^{1\text{-}391}$ accounted for most of the protein in the sample (Figure 48A). Based on this data, only sample 2 was collected and used for future crystal screening experiments.

Further purification optimization using 0.1% Triton X-100 in the MOPS based lysis buffer increased the yield of the purified protein from affinity chromatography by 2.5-fold; 10.3 mg versus 4.1 mg per ~750 mL *E.coli* culture with or without Triton X-100, respectively (Figure 48B). This result is consistent with the hydrophobic nature of LOS and suggests increased solubility of the LOS protein by the non-ionic detergent

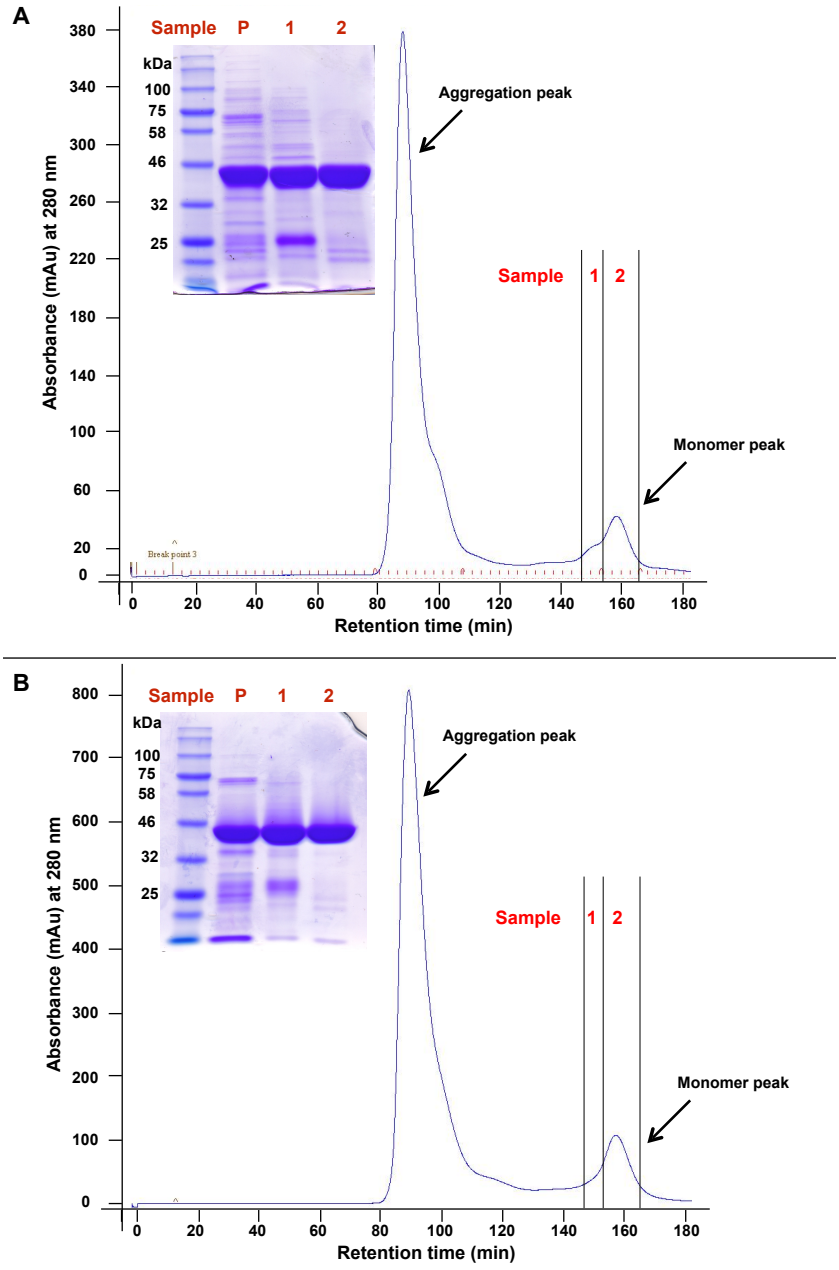


Figure 48. Purification of 6xHis-LOS$^{1-391}$ using MOPS based buffer. Approximately 20 μg of purified protein was analyzed by SDS-PAGE. Sample P = protein purified from nickel affinity chromatography was loaded onto the gel filtration column. Sample 1 and sample 2 were collected from the LOS monomer peak. **A.** Purification using the MOPS based buffer. **B.** Purification using the MOPS based buffer containing 0.1% Triton X-100 in the lysis buffer. LOS = lycopaoctaene synthase.

Triton X-100. However, analysis of the 10.3 mg of purified protein by gel filtration did not show any change in the chromatograph profile (compare figures 48A and 48B). Approximately 2.5 mg of sample 2 (LOS monomer) was recovered from the gel filtration column and analysis by SDS-PAGE showed a highly pure protein band for 6xHis-LOS$^{1-391}$ (Figure 48B). This process was repeated several times to generate 10.6 mg of highly pure LOS$^{1-391}$ monomer (sample 2 in figure 48B). As mentioned earlier, generating protein crystals of HSS was only possible when HSS was co-incubated with either FSPP, PSPP, or the inhibitor molecules[92-94, 138]. Since LOS can utilize FPP to produce squalene, crystal screening was conducted with the LOS$^{1-391}$ monomer sample by incubating with FSPP. However, crystal screening under 600 different solution conditions with 4 mM FSPP did not result in any protein crystals. There could be many factors contributing to the unsuccessful crystallization of the LOS$^{1-391}$ protein. It should be noted the 6xHis tag was not cleaved from the LOS$^{1-391}$ monomer sample used in the crystal screening experiment. It was assumed the 6xHis tag may not affect the LOS$^{1-391}$ crystallization since the enzyme activity of LOS$^{1-391}$ was not compromised by the 6xHis tag and the monomer sample of 6xHis-LOS$^{1-391}$ was stable and did not aggregate over time. Furthermore, several protein structures containing 6xHis tags have been reported in the literature[142]. However, many crystallographers do not recommend the use of 6xHis or any other tags during crystallization as they are known to increase the aggregation of the purified protein and hinder the protein crystallization process[142]. Although highly pure and stable LOS$^{1-391}$ monomer was obtained, unsuccessful crystallization of this protein could have resulted due to the presence of the 6xHis tag at the N-terminus.

3.2a3 Purification of 6xHis-SUMO-TEV-LOS$^{1-391}$ protein for crystal screening

In order to purify a tag free version of LOS, a new LOS construct was designed to encode for 6xHis-SUMO-TEV-LOS$^{1-391}$. The N-terminal 6xHis tag was added in order to purify protein using nickel affinity chromatography, whereas the small ubiquitin-like modifier (SUMO) protein was fused to LOS as it is known to significantly increase the solubility and stability of fusion proteins[143]. The cleavage site for tobacco etch virus nuclear-inclusion-a endopeptidase (TEV protease) was also added after the 6xHis-SUMO tag to facilitate the removal of both tags from the LOS$^{1-391}$ fusion protein. Purification of 6xHis-SUMO-TEV-LOS$^{1-391}$ using a cell pellet from a ~750 mL *E.coli* culture yielded approximately 32.5 mg of pure protein after nickel affinity chromatography (sample 1, Figure 49). This is three times higher yield than previously observed for 6xHis-LOS$^{1-391}$. Next the 6xHis-SUMO-TEV-LOS$^{1-391}$ protein was treated with 6xHis-TEV protease to remove the 6xHis-SUMO tag, the protein sample incubated with nickel resin to bind any uncleaved LOS fusion protein and the 6xHis-TEV protease, and the flow through (sample 4, Figure 49) containing LOS$^{1-391}$ analyzed by gel filtration (Figure 49). Out of the 32.5 mg of protein used in the cleavage reaction with 6xHis-TEV protease, only 4.3 mg of tag free LOS$^{1-391}$ was recovered in the flow through (sample 4, Figure 49) whereas most of the tag free LOS$^{1-391}$ remained on the nickel column as an aggregate (sample 3, Figure 49).

Interestingly, the gel filtration chromatograph of LOS$^{1-391}$ showed a change in the profile of the aggregation and monomer peaks when compared to the profile

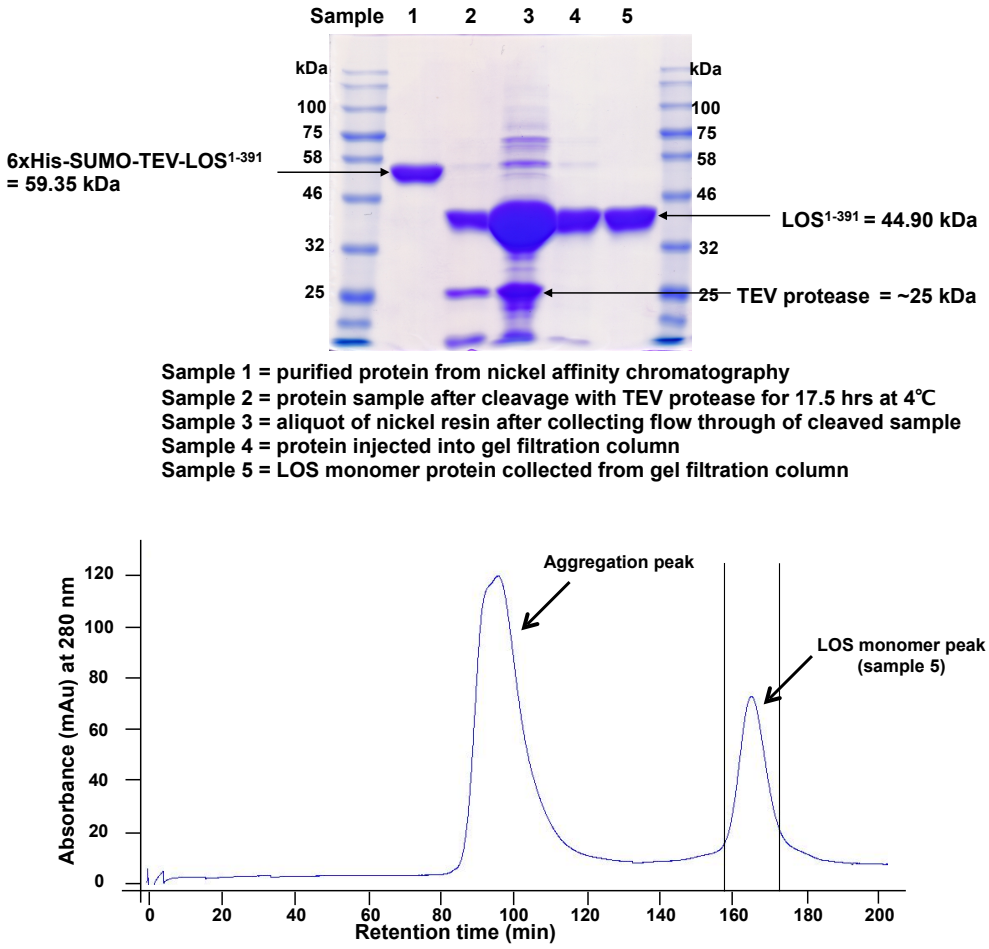


Figure 49. Purification of 6xHis-SUMO-TEV-LOS$^{1-391}$. Except for sample 3, approximately 10µg of protein sample was analyzed by SDS-PAGE. The predicted molecular weight of each protein is shown. Protein purification was done using MOPS based buffer containing 0.1% Triton X-100 in the lysis buffer. LOS = lycopaoctaene synthase.

observed for 6xHis-LOS$^{1-391}$; the peak size of the aggregated protein decreased whereas the peak size of LOS monomer increased (compare figures 48B and 49). Analysis of the LOS monomer sample by SDS-PAGE showed a highly pure protein band corresponding to the LOS$^{1-391}$ protein (sample 5, Figure 49). Unlike purification of 6xHis-LOS$^{1-391}$, the LOS$^{1-391}$ monomer peak does not contain the small protein band around 25 kDa (compare Figures 48B and 49). It is possible the SUMO tag increased protein solubility as well as the stability of LOS$^{1-391}$. Protein purification for 6xHis-SUMO-TEV-LOS$^{1-391}$ was repeated several times to generate 10.3 mg of tag free LOS$^{1-391}$ monomer protein. Unfortunately, crystal screening using the LOS$^{1-391}$ protein conducted with or without geranylgeranylthiol diphosphate (GGSPP), a GGPP substrate analog, under many crystallization conditions did not yield protein crystals. Extensive screening of crystallization conditions should be done in the future in order to obtain a crystal of the LOS$^{1-391}$ protein.

3.2a4 Purification using 6xHis-LOS$^{33-381}$ & 6xHis-SUMO-TEV-LOS$^{33-381}$ constructs

The LOS$^{33-381}$ construct, which was designed based on the published crystal structure of HSS, is soluble and catalytically active (Figure 46). In order to use LOS$^{33-381}$ for crystal screening experiments, test purification using two different constructs was conducted; 6xHis-LOS$^{33-381}$ and 6xHis-SUMO-TEV-LOS$^{33-381}$. Purification of 6xHis-LOS$^{33-381}$ using a cell pellet from a ~750 mL *E.coli* culture yielded approximately 3.3 mg of purified protein after nickel affinity chromatography (sample P, Figure 50). This is 3-fold less protein yield when compared to purification of the 6xHis-LOS$^{1-391}$ protein

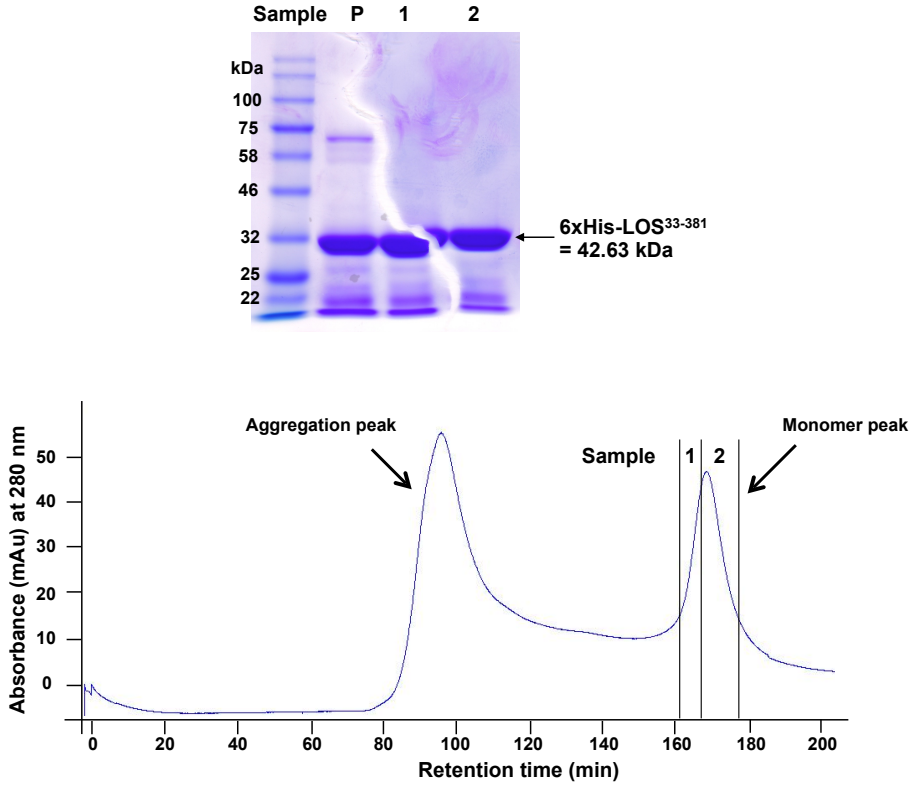
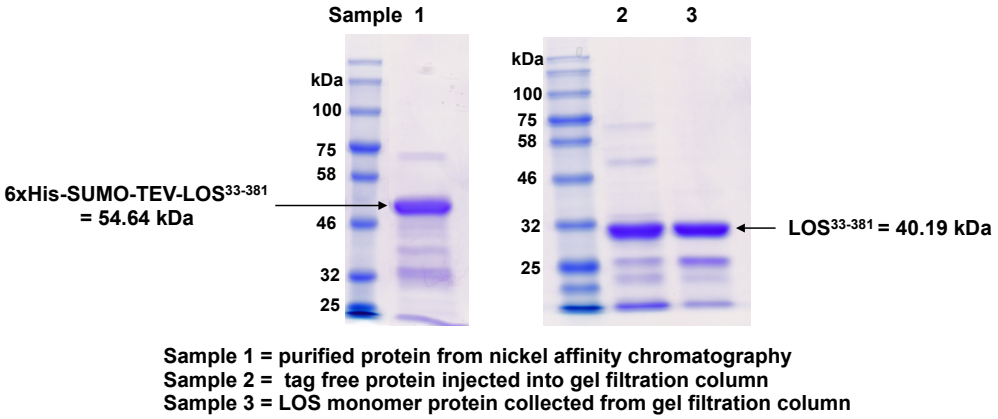


Figure 50. Purification of 6xHis-LOS$^{33-381}$. Approximately 20 μg of purified protein was analyzed by SDS-PAGE. The predicted molecular weight of 6xHis-LOS$^{33-381}$ protein is shown. Sample P = protein purified from nickel affinity chromatography was loaded onto the gel filtration column. Sample 1 and sample 2 were collected from the LOS monomer peak. Protein purification was done using MOPS based buffer containing 0.1% Triton X-100 in the lysis buffer. LOS = lycopaoctaene synthase.



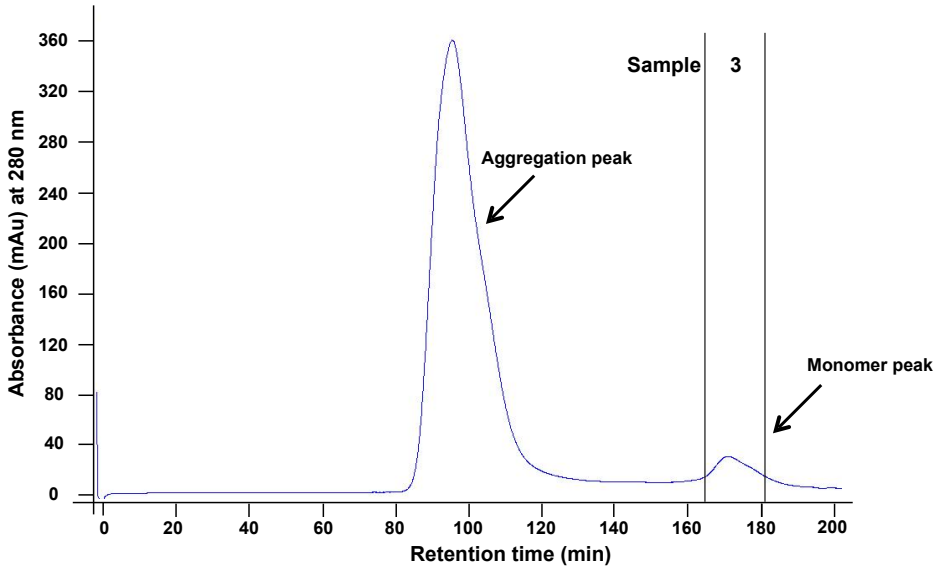


Figure 51. Purification of 6xHis-SUMO-TEV-LOS$^{33-381}$. Approximately 10µg of each protein sample was analyzed by SDS-PAGE. The predicted molecular weight of each protein is shown. Protein purification was done using MOPS based buffer containing 0.1% Triton X-100 in the lysis buffer. LOS = lycopaoctaene synthase.

under same conditions. In contrast to 6xHis-LOS$^{1-391}$, analysis of purified 6xHis-LOS$^{33-381}$ protein by gel filtration showed increased peak size of the LOS monomer whereas the peak size of the aggregated LOS protein is decreased (compare figures 48B and 50). Although a shoulder peak was not observed for the LOS monomer, two fractions (sample 1 and 2) from the monomer peak were collected. Analysis of these samples by SDS-PAGE showed a major protein band corresponding to 6xHis-LOS$^{33-381}$ along with several small protein bands of sizes below 25kDa (Figure 50).

In the case of 6xHis-SUMO-TEV-LOS$^{33-381}$, protein purification using a cell pellet from a ~750 mL *E.coli* culture yielded approximately 11.4 mg of purified protein after nickel affinity chromatography (sample 1, Figure 51). This is 3-fold less protein yield when compared to purification of 6xHis-SUMO-TEV-LOS$^{1-391}$ under the same conditions. Next the 6xHis-SUMO tag of 6xHis-SUMO-TEV-LOS$^{33-381}$ was cleaved using TEV protease, and the tag free LOS$^{33-381}$ was loaded onto the gel filtration column (Figure 51). Surprisingly, most of the LOS$^{33-381}$ protein aggregated as shown by the prominent aggregation peak and a small peak was observed for the LOS$^{33-381}$ monomer (Figure 51). Further analysis of the monomer peak (sample 3) by SDS-PAGE showed decreased purity of LOS$^{33-381}$ as it contained several protein contaminants of smaller sizes (Figure 51). It is possible that the small protein bands seen in sample 3 could result from the degradation of the LOS$^{33-381}$ protein. In summary, protein purity of the double truncated constructs (6xHis-LOS$^{33-381}$ and 6xHis-SUMO-TEV-LOS$^{33-381}$) is less when compared to the C-terminal truncated constructs (6xHis-LOS$^{1-391}$ and 6xHis-SUMO-TEV-LOS$^{1-391}$). Further optimization of LOS$^{33-381}$ protein purification should be

conducted in order to obtain highly pure and stable monomer samples suitable for crystal screening experiments.

3.2b Mutagenesis experiment with LOS enzyme

3.2b1 Identification of LOS residues for mutagenesis

Our attempts to obtain the crystal structure of LOS for structure-function studies were unsuccessful. Thus, for rational mutagenesis experiments LOS was compared to LSS and four other SS enzymes from other organisms to determine regions of similarity or dissimilarity and to identify mutational targets for understanding LOS substrate specificity. The reaction catalyzed by SS has been extensively investigated and the role of important catalytic residues verified using protein crystal structures, site-directed mutagenesis, and enzyme assays[81, 92-94]. From these studies seven functional SS domains have been identified: domains I-V for substrate binding and catalysis, and a FLAP domain and a J-K loop involved in NADPH binding (Figure 44)[79, 81, 94]. An amino acid alignment of LOS with the five SS proteins shows conservation of all these domains in LOS as well as two domain II and IV DXXED motifs involved in FPP/PSSPP diphosphate binding[81, 94], and the positively charged residues R52, R218, K315, R317 and K318 that are involved in NADPH binding[94] (Figure 44). From this analysis it is obvious LOS shows high sequence conservation with SS proteins, making it difficult to use these alignments to identify residues that may confer LOS substrate specificity. Additionally, residues that dictate the novel catalytic activity and promiscuity of LOS may lie outside these conserved regions.

Thus, we found identifying the amino acid differences that control LOS substrate binding to be most straightforward using a structural comparison. Based on the HSS^{31-370} - LOS^{33-381} structural comparison, five LOS amino acids appear to be important for substrate interaction: T65, M180, S276, A288 and V289 (Figure 52A). Except for S276, all residues are located within the conserved functional SS enzyme domains I, III, and IV, and all five LOS residues are different from their corresponding amino acids in the other SS proteins (Figure 52A). For instance, the corresponding amino acids in HSS are V69, L183, Y276, F288 and C289 (Figure 52A), and are located in the hydrophobic region that forms the bottom of the central cavity for the active site of HSS (Figure 52B)[94]. Previous HSS structure studies have shown that these five amino acids form a hydrophobic floor in the substrate binding pockets for interaction with the non-polar tail of the FSPP substrates[94]; the hydrophobic side chains of V69, L183, and F288 form the hydrophobic floor for the substrate 1 (S1) pocket, whereas the hydrophobic portions of the Y276 and C289 side chains form the hydrophobic floor for the substrate 2 (S2) pocket (Figure 52B). The corresponding five LOS residues are also located in the same hydrophobic region of the LOS active site, with the T65, M180, and A288 side chains contributing to the hydrophobic floor for the S1 substrate pocket, and V289 and S276 contributing to the hydrophobic floor for the S2 substrate pocket (Figure 52C). However, the replacement of HSS F288 with A288 in LOS and HSS Y276 with S276 in LOS appears to enlarge the S1 and S2 substrate binding pockets, respectively, to accommodate the larger GGPP substrate (Figure 52C). We therefore hypothesized that LOS A288 and S276 were key mutations that arose during gene duplication to allow

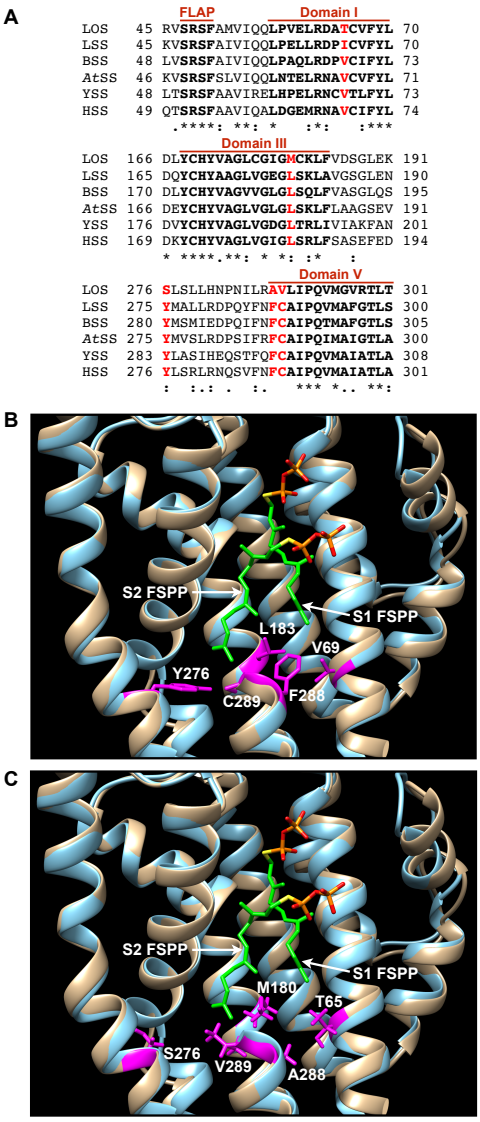


Figure 52. Alignment of amino acid and structures. For clarity, structures are focused on the active site of both enzymes. **A.** Amino acid sequence alignment of LOS with SS proteins focusing on the regions around the residues analyzed in this study. The five LOS residues chosen for mutagenesis and the corresponding residues in the SS proteins are shown in red. LOS, lycopaoctaene synthase; LSS, *B. braunii* L race SS; BSS, *B. braunii* B race SS; *At*SS, *Arabidopsis thaliana* SS; YSS, yeast SS; HSS, human SS. **B.** Superposition of HSS (light blue) and LOS (light brown) showing the five amino acids of HSS located on the hydrophobic floor of the two substrate binding sites highlighted in magenta. **C.** Superposition of HSS (light blue) and LOS (light brown) showing the five amino acids of LOS located on the hydrophobic floor of two substrate binding sites highlighted in magenta.

binding of GGPP while still retaining FPP binding. Thus, A288 and S276 plus the three other residues (T65, M180, V289) forming the LOS S1 and S2 substrate pocket hydrophobic floors were chosen for mutational studies to decipher their roles in substrate binding and product specificity.

3.2b2 Characterization of LOS mutants *in vitro*

In order to determine the roles of LOS T65, M180, S276, A288 and V289 in substrate binding and product formation these amino acids were mutated to their corresponding residues in HSS individually and in combinations (Table 2). The mutated proteins were purified, enzyme activities characterized in a mixed substrate assay (Figure 53A) containing equimolar ^{3}H-FPP and ^{3}H-GGPP, and the enzymatic activity quantified based on ^{3}H incorporation into the products lycopaoctaene, $C_{35}H_{58}$, and squalene (Figure 53B). The FPP/GGPP mixed substrate assay was chosen over the FPP/PPP or GGPP/PPP mixed substrate assays because the FPP/GGPP assay is more robust in terms of product formation[59].

The results from these assays indicate that for the M1 to M5 single amino acid mutations, M1, M2, and M5 retained their ability to produce lycopaoctaene, whereas the replacement of S276 and A288 with the aromatic amino acids Tyr and Phe in M3 and M4, respectively, lead to a significant loss of lycopaoctaene production activity (Figure 53B). When compared to wild-type (WT), M3 retained 47% of lycopaoctaene production, and M4 had a substantial reduction retaining only 29% of lycopaoctaene production activity (Figure 53B). In the case of $C_{35}H_{58}$ production, all single mutants at least retained WT activity (Figure 53B). Interestingly, the M3 mutant showed a

substantial increase in $C_{35}H_{58}$ production to 183% of that of WT LOS (Figure 53B). Except for M3, all single mutants retained at least WT activity levels for squalene production (Figure 53B). The observed activity for M3 and M4 suggests the increased residue size in these mutants blocks GGPP biding in the S1 pocket by the M4 A288F mutation and in the S2 pocket by the M3 S276Y mutation. This indicates the importance of LOS S276 and A288 in binding of GGPP substrate. Further support for this is seen in the increased $C_{35}H_{58}$ production in M3 (Figure 53B) suggesting favorable binding of one GGPP in the S1 pocket and one FPP in the S2 pocket. Additionally, the decreased lycopaoctaene production, but no change in $C_{35}H_{58}$ and squalene production in M4 (Figure 53B) suggests limited binding of two GGPP molecules and preferable binding of one FPP in the S1 pocket and one GGPP in the S2 pocket (Figures 53A and B).

Next, the LOS double mutants M6 to M9 (Table 2) were characterized for enzymatic activity using the same mixed substrate assay (Figure 53A). As shown in Figure 53B, the enzyme activity of M6 is similar to that of WT. When comparing the M7 double mutant to the parent M4 mutant background, M7 did not show a difference in lycopaoctaene production, but $C_{35}H_{58}$ production was slightly increased and squalene formation was significantly reduced (Figure 53B). For the M9 double mutant, which has a V289C substitution in the S276Y M3 background, a small increase in $C_{35}H_{58}$ production was observed, whereas the activities for lycopaoctaene and squalene production remained the same as that of M3 background (Figure 53B). Most interestingly, the M8 double mutant, which has both the M3 S276Y mutation and the M4

Mutant name	Mutation
M1	T65V
M2	M180L
M3	S276Y
M4	A288F
M5	V289C
M6	T65V M180L
M7	T65V A288F
M8	S276Y A288F
M9	V289C S276Y
M10	T65V M180L A288F
M11	T65V M180L S276Y A288F
M12	T65V M180L S276Y A288F V289C

Table 2. List of lycopaoctaene synthase (LOS) mutants characterized in this study.

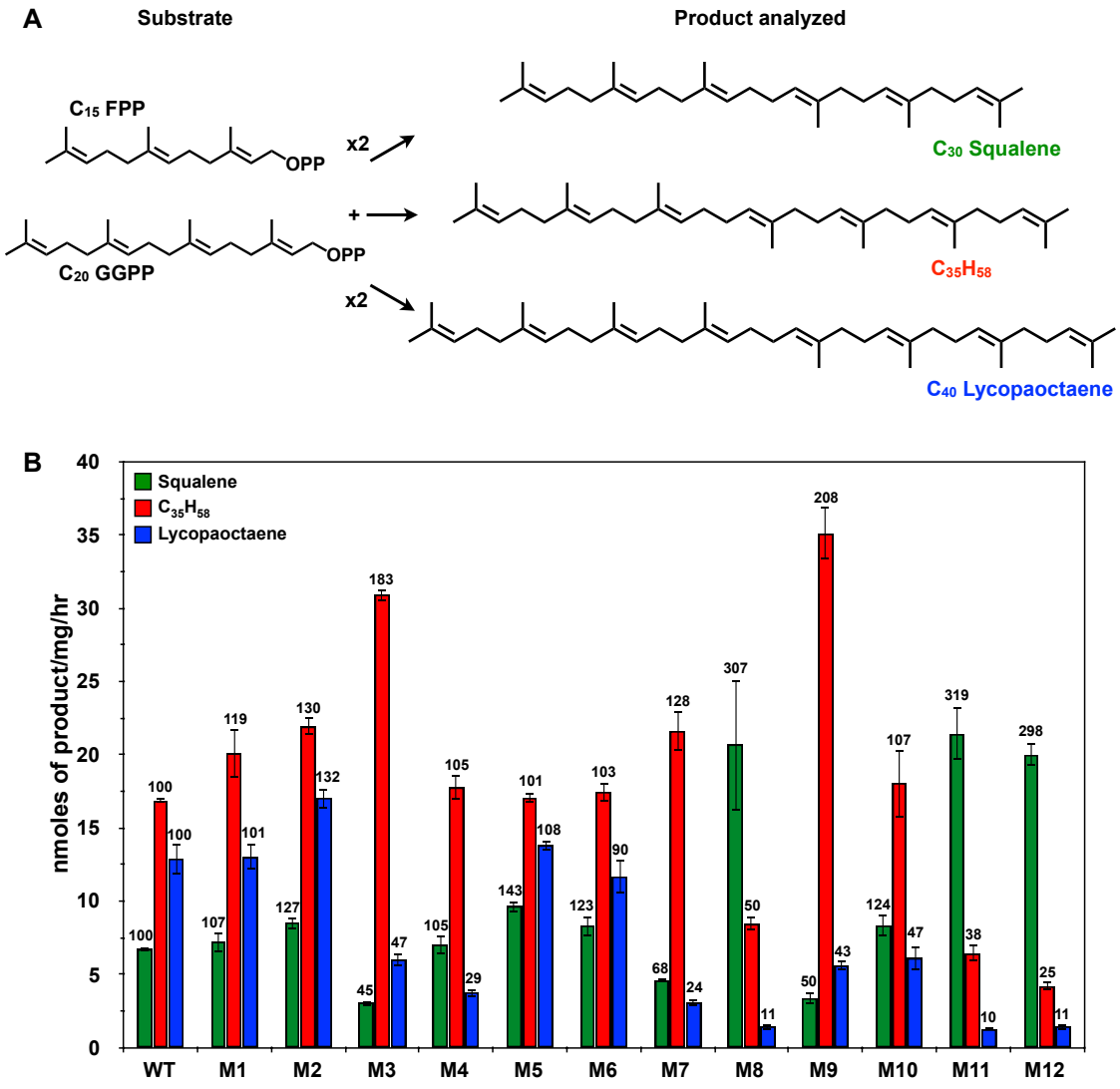


Figure 53. Characterization of LOS mutants *in vitro*. A. Reaction scheme showing LOS products from the FPP/GGPP mixed substrate assay. **B.** Enzyme activities of the LOS mutants in the FPP/GGPP mixed substrate assay showing production of squalene, $C_{35}H_{58}$, and lycopaoctaene. Numbers on top of each bar indicate the activity percentage based on WT LOS activity as 100%. Data shown are from three independent experiments (n = 3).

A288F mutation, showed a drastic reduction in the ability to produce lycopaoctaene and $C_{35}H_{58}$ when compared to WT (Figure 53B). Moreover, the SS activity for M8 is substantially increased more than 3-fold over the WT enzyme (Figure 53B). These results for M8 further support our hypothesis that S276 and A288 are important for controlling LOS substrate binding, and replacing these residues with bulky aromatic amino acids blocks the binding of GGPP in both substrate binding pockets. This is further supported by a previous study where F288 was shown to be essential in the SS reaction[94]. The HSS^{F288A} crystal structure showed mutation to the smaller Ala resulted in a deeper cleft in the S1 substrate pocket[94]. Thus, having Ala at position 288 in LOS would expand the S1 substrate pocket to allow for GGPP binding. Additionally, F288 was proposed to be involved in PSPP diphosphate ionization during the second half reaction, and the HSS^{F288A} structure indicated the expanded S1 substrate pocket caused misorientation of the PSPP intermediate in the active site[94]. Accordingly the F288A mutation in HSS significantly reduced the HSS enzyme activity to 22.7% of WT activity[94]. Since the WT LOS with A288 is fully active this would suggest there are additional LOS residues that are different from SS that would allow for proper PLPP diphosphate orientation and ionization.

Finally, the M10 triple mutant, the M11 quadruple mutant, and the M12 quintuple mutant (Table 2) were analyzed for enzymatic activity using the mixed substrate assay. The M10 mutant showed activity similar to that of the M4 parent background (Figure 53B). For the M11 and M12 mutants, the activity is very similar to each other. But in comparison to the M8 parent background, both M11 and M12 showed

a further reduction in $C_{35}H_{58}$ production and no significant change in lycopaoctaene and squalene formation (Figure 53B). These results suggest that mutation of the five LOS amino acids studied here to their corresponding HSS residues is sufficient to change LOS substrate binding specificity to mainly FPP binding and thus product specificity to squalene production (compare WT and M12 in Figure 53B). It should be noted the M12 mutant retained residual activity for lycopaoctaene and $C_{35}H_{58}$ production (Figure 53B), indicating low levels of GGPP binding and suggesting further modification is required in the M12 mutant background to completely abolish GGPP binding (Figure 53B).

3.2b3 Analysis of LOS and LOS mutants *in vivo*

To further support our *in vitro* assay data we attempted to characterize the activity of WT LOS and the LOS mutants using *Escherichia coli* as an *in vivo* expression system since the native host, *B. braunii*, cannot be transformed. Like *B. braunii*, *E. coli* utilizes the MEP pathway for isoprenoid biosynthesis[144]. Most importantly, *E. coli* does not naturally produce the lycopaoctaene, $C_{35}H_{58}$, or squalene hydrocarbons under investigation in our study (Figure 54, panel I). Thus, the *E. coli* expression system allows for a straightforward platform to characterize LOS and the LOS mutants *in vivo*.

As mentioned earlier, the *in vitro* mixed substrate assay used for characterization of the LOS mutants contained equal concentrations of FPP and GGPP. Thus, an *E. coli* line generating equal and high amounts of FPP and GGPP is desired for characterization of LOS and the LOS mutants *in vivo*. However, when LOS was overexpressed in *E. coli*, only a small amount of squalene was produced with no detection of $C_{35}H_{58}$ or lycopaoctaene (Figure 54, panel II). This result is consistent with the limited FPP

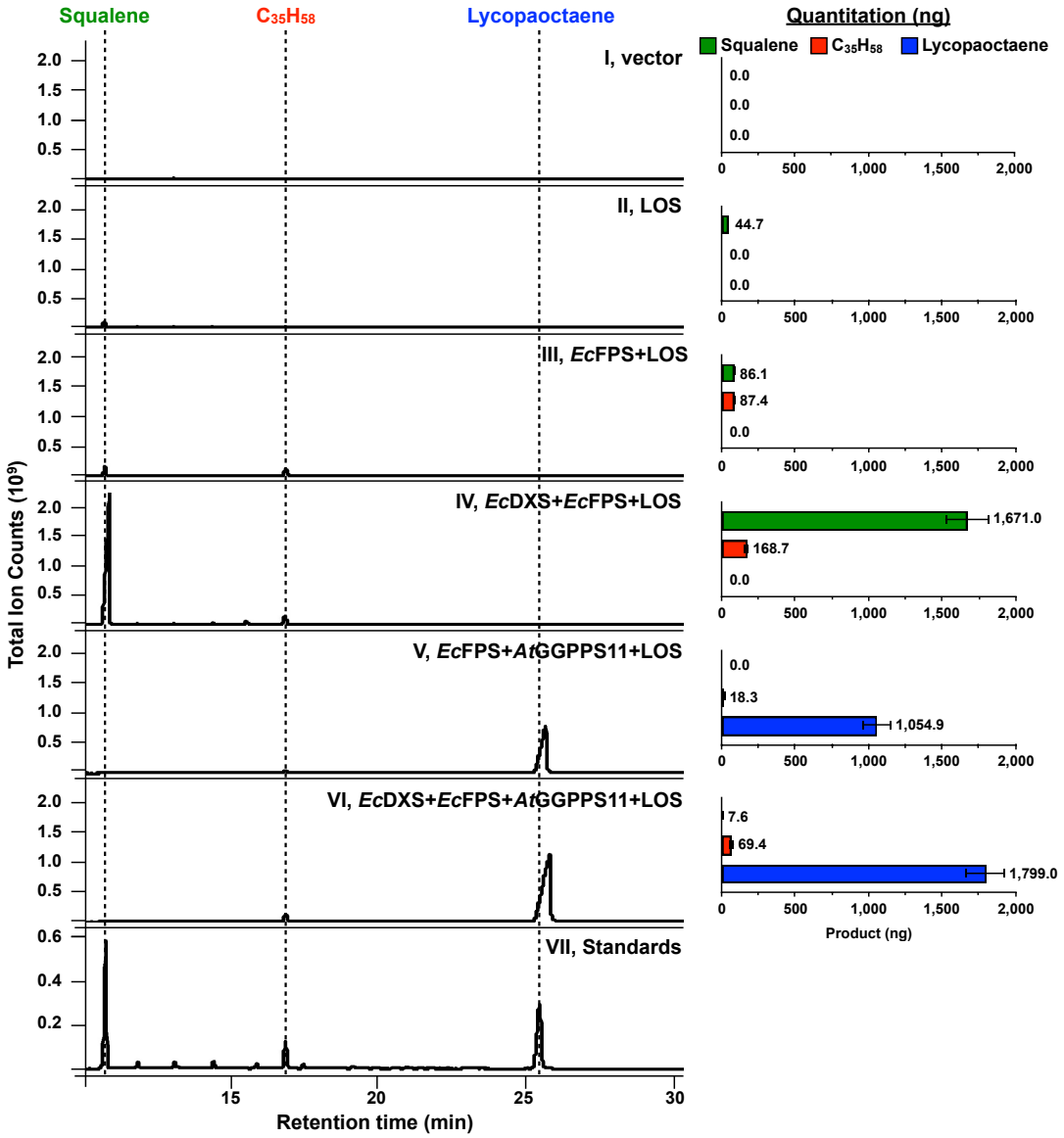


Figure 54. GC-MS analysis of *E. coli* cells expressing different gene constructs. (I) Empty vector expression. (II) LOS expression. (III) *Ec*FPS and LOS expression. (IV) *Ec*DXS, *Ec*FPS, and LOS expression. (V) *Ec*FPS, *At*GGPPS11, and LOS expression. (VI) *Ec*DXS, *Ec*FPS, *At*GGPPS11, and LOS expression. (VII) Hydrocarbon standards. Shown to the right of the GC-MS profiles is quantitation of each hydrocarbon based on the GC-MS data and expressed in nanograms. The GC-MS data shown are representatives from three independent experiments (n = 3) and each experiment was used for quantitation. Numbers shown in the quantitation are the average value for each hydrocarbon. The large amount of squalene and lycopaoctaene in panels IV, V, and VI caused the altered migration of these peaks compared to the standards.

amounts and even lower GGPP levels present in *E. coli*[126]. To overcome this limitation, we overexpressed FPP synthase from *E. coli* (ispA, referred to here as *Ec*FPS) to increase FPP levels. Overexpression of *Ec*FPS with LOS showed squalene production almost doubled and a roughly equal amount of $C_{35}H_{58}$ was detected (Figure 54, panel III). This would suggest *Ec*FPS overexpression slightly increased intracellular FPP levels, and GGPP levels were high enough to contribute to $C_{35}H_{58}$ production. However, lycopaoctaene was not detected in this line (Figure 54, panel III) indicating the available GGPP pool is small and not sufficient for LOS-mediated lycopaoctaene production.

To further boost FPP and GGPP production, deoxyxylulose-5-phosphate synthase from *E. coli* (*Ec*DXS) was overexpressed along with *Ec*FPS and LOS. The DXS enzyme catalyzes the first reaction of the MEP pathway, and is considered a rate limiting step of the MEP pathway[61]. Expression of *Ec*DXS with *Ec*FPS and LOS generated high levels of squalene and increased the amount of $C_{35}H_{58}$ (Figure 54, IV), suggesting an increase in the carbon flux through the MEP pathway to further boost FPP production. However, production of lycopaoctaene was still not detected (Figure 54, panel IV), arguing that while *Ec*DXS and *Ec*FPS overexpression significantly increased intracellular FPP pools, GGPP pools were not increased to a level required for lycopaoctaene biosynthesis.

In order to enhance the GGPP pool, GGPP synthase from *Arabidopsis thaliana* (*At*GGPPS11[115]) was overexpressed along with *Ec*FPS and LOS. Analysis of hydrocarbon production in this line showed detection of a prominent peak for lycopaoctaene, whereas only a trace amount of $C_{35}H_{58}$ was observed (Figure 54, panel V). Surprisingly, no squalene was produced in this line (Figure 54, panel V), indicating

*At*GGPPS11 converted most of the FPP pool into GGPP, which was utilized by LOS to produce lycopaoctaene, and the available FPP was used to produce the low level of $C_{35}H_{58}$ (Figure 54, panel V). To further boost FPP production, we overexpressed *Ec*DXS, *Ec*FPS, and *At*GGPPS11 in one line with LOS and analyzed the hydrocarbon profile in this line. The analysis showed lycopaoctaene production was further enhanced and remained the predominant product (Figure 54, panel VI). A trace amount of squalene was detected and $C_{35}H_{58}$ production was slightly increased (Figure 54, panel VI), suggesting an enhanced level of intracellular FPP pools as a result of *Ec*DXS overexpression in this line. However, the level of FPP may still not be sufficient to support increased $C_{35}H_{58}$ or squalene production. Low FPP pool levels in this line can be supported by two theories; the *At*GGPPS11 is a very efficient enzyme that rapidly converts available FPP to GGPP, or the use of GGPP by LOS creates a pull on the MEP pathway that rapidly converts the FPP pool to GGPP by *At*GGPPS11.

Interestingly, this product specificity of LOS for lycopaoctaene over $C_{35}H_{58}$ and squalene as observed in *E. coli* (Figure 54, panel VI) could mirror hydrocarbon metabolism in *B. braunii* race L. Our past studies have shown that while lycopaoctaene is rapidly converted to lycopadiene there is no detectable squalene production and low levels of $C_{35}H_{64}$, a reduced product of $C_{35}H_{58}$[59]. This would suggest in the L race there is a small FPP pool available to LOS for $C_{35}H_{64}$ production, and mainly a large GGPP pool available for lycopaoctaene production. In *B. braunii* race L this may come about by metabolic channeling[145] of the MEP pathway towards lycopadiene biosynthesis via

lycopaoctaene production by LOS. As we saw in *E. coli*, the use of GGPP by LOS would create a pull that would reduce FPP pools as FPP is used to generate more GGPP.

Next we investigated the *in vivo* activity of several select LOS mutants by coexpression in the *Ec*DXS, *Ec*FPS, and *At*GGPPS11 expressing *E. coli* line. Our initial goal was to create an *E. coli* screening platform producing approximately equal FPP and GGPP levels that would give rise to roughly equal levels of lycopaoctaene, $C_{35}H_{58}$ and squalene from WT LOS expression similar to WT LOS in the *in vitro* mixed substrate assay. This line could then be used for comparing the LOS mutants *in vivo*. Even though the *Ec*DXS, *Ec*FPS, and *At*GGPPS11 expressing *E. coli* line does not appear to produce equal FPP and GGPP levels, we still expressed the LOS mutants in this *E. coli* line to assess *in vivo* function of these mutants. This allowed us to generate valuable information as discussed below that can support some aspects of the *in vitro* data.

The mutants selected for expression in the *Ec*DXS, *Ec*FPS, and *At*GGPPS11 expressing *E. coli* line were M3, M4, and M8 (Table 2). Each of these mutants are focused on the LOS S276 and A288 residues, which appear to have the biggest influence on substrate binding and product formation based on the *in vitro* studies (Figure 53B). Additionally, the M12 mutant (Table 2) was chosen because M12 contains the full set of five LOS mutations based on the HSS sequence. As previously seen in Figure 54, WT LOS expression in this *E. coli* line predominantly produced lycopaoctaene, a small amount of $C_{35}H_{58}$, and trace amounts of squalene (Figure 55, panel I). Expression of the S276Y M3 mutant showed a reduced level of lycopaoctaene, a large increase in $C_{35}H_{58}$ production, and no change in squalene levels in comparison to WT LOS (Figure 55,

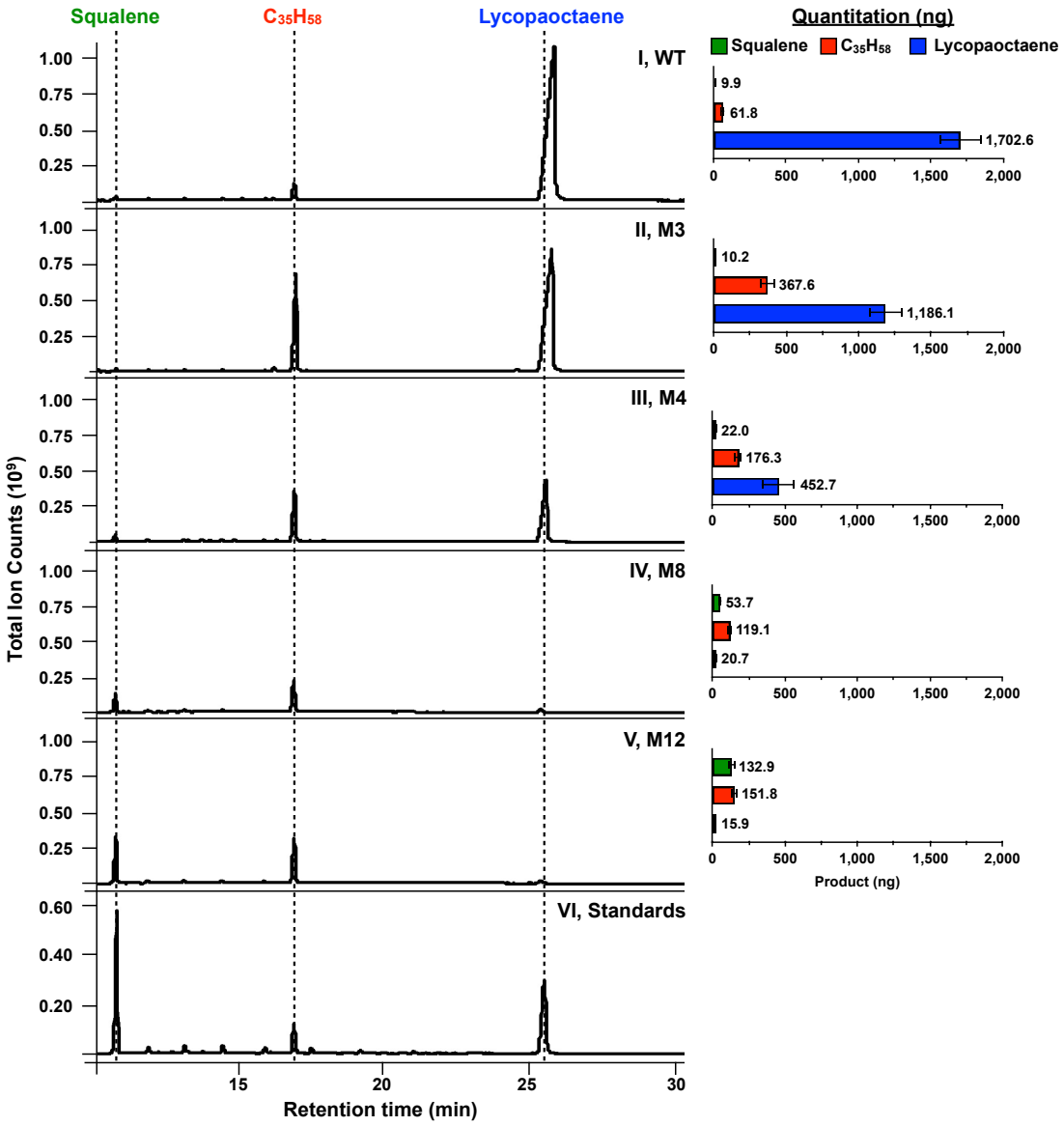


Figure 55. Characterization of LOS mutants *in vivo*. Shown are GC-MS profiles and quantitation of *n*-hexane extractable metabolites from *E. coli* cells expressing WT LOS or select LOS mutants with *Ec*DXS, *Ec*FPS, and *At*GGPPS11. (I) WT LOS expression. (II) LOS M3 mutant expression. (III) LOS M4 mutant expression. (IV) LOS M8 mutant expression. (V) LOS M12 mutant expression. (VI) Hydrocarbon standards. Shown to the right of the GC-MS profiles is quantitation of each hydrocarbon based on the GC-MS data and expressed in nanograms. The GC-MS data shown are representatives from three independent experiments (n = 3) and each experiment was used for quantitation. Numbers shown in the quantitation are the average value for each hydrocarbon. The large amount of lycopaoctaene in panels I and II caused the altered migration of these peaks compared to the standard.

panel II). This is similar to what was seen in the *in vitro* assays for M3 (Figure 53B) where the $C_{35}H_{58}$ levels were increased and the lycopaoctaene levels were reduced compared to WT LOS. Similarly, in comparison to WT LOS the A288F M4 mutant *in vivo* showed a reduction in lycopaoctaene production and increased $C_{35}H_{58}$ and squalene production (Figure 55, panel III). These data are also similar to the *in vitro* data for M4 (Figure 53B), which showed decreased lycopaoctaene levels, maintenance of squalene levels, but no change in $C_{35}H_{58}$ levels. For the M8 S276Y/A288F double mutant, its ability to produce lycopaoctaene *in vivo* was drastically reduced, whereas production of $C_{35}H_{58}$ and squalene was increased compared to WT LOS (Figure 55, panel IV). Again, this is similar to what was seen for M8 *in vitro* (Figure 53B) where lycopaoctaene production was drastically reduced and the squalene levels were increased compared to WT LOS. However, the *in vivo* level of $C_{35}H_{58}$ for M8 went up while the *in vitro* level of $C_{35}H_{58}$ went down compared to WT LOS (Figure 53B). This would indicate some level of GGPP binding by this mutant when high GGPP pools are present *in vivo* (Figure 55, panel IV). As was seen for the *in vitro* assays, *in vivo* the M12 mutant showed a drastic reduction in lycopaoctaene production and increased squalene levels in comparison to WT LOS (Figure 55, panel V). However, the $C_{35}H_{58}$ levels were higher than WT LOS *in vivo* (Figure 55, panel V), while *in vitro* the $C_{35}H_{58}$ levels decreased compared to WT LOS (Figure 53). Again, this could be due to saturating levels of GGPP *in vivo*. As intended, the M12 mutations changed the LOS product specificity from lycopaoctaene to squalene both *in vitro* and *in vivo*. Overall, while the absolute changes in product formation for the mutants in figure 55 do not match that seen for the *in vitro* data in

figure 53, the trends in product formation for the M3, M4, M8, and M12 mutants *in vivo* are consistent with the enzymatic activities observed in the *in vitro* assays.

3.2b4 Structural insights into the role of LOS S276 and A288 in GGPP binding

The studies presented above indicate LOS residues S276 and A288 play a key role in substrate binding and product specificity. To gain more insight into the role of these two amino acids in controlling substrate binding, we compared the HSS structure in complex with FSPP to the predicted LOS structure focusing on HSS Y276 and F288, and LOS S276 and A288 (Figure 56A). The analysis showed the distances in HSS from the aromatic ring of F288 and the phenol group of Y276 to the corresponding S1 and S2 FSPP tails are 3.54Å and 4.35Å, respectively (Figure 56A). This suggests HSS cannot bind GGPP as a substrate because the bulky F288 in the S1 binding site and Y276 in the S2 binding site would interfere with GGPP binding since GGPP is ~4.00Å longer than FPP due to the extra C_5 isoprene unit. This notion is further supported by the observed *in vitro* and *in vivo* enzyme activities of the LOS M8 S276Y/A288F double mutant, which showed substantial losses of lycopaoctaene and $C_{35}H_{58}$ formation, and a large increase in squalene production as a result of replacing the small S276 and A288 residues with Tyr and Phe, respectively (Figure 53B and Figure 55, panel IV). Further support of S276 and A288 controlling GGPP binding in LOS is the 6.76Å distance between the A288 side chain and the S1 FSPP tail, and the 7.66Å distance between the S276 hydroxyl group and the S2 FSPP tail (Figure 56A). These distances should provide enough room in the two LOS substrate binding sites to accommodate the longer hydrophobic tail of GGPP.

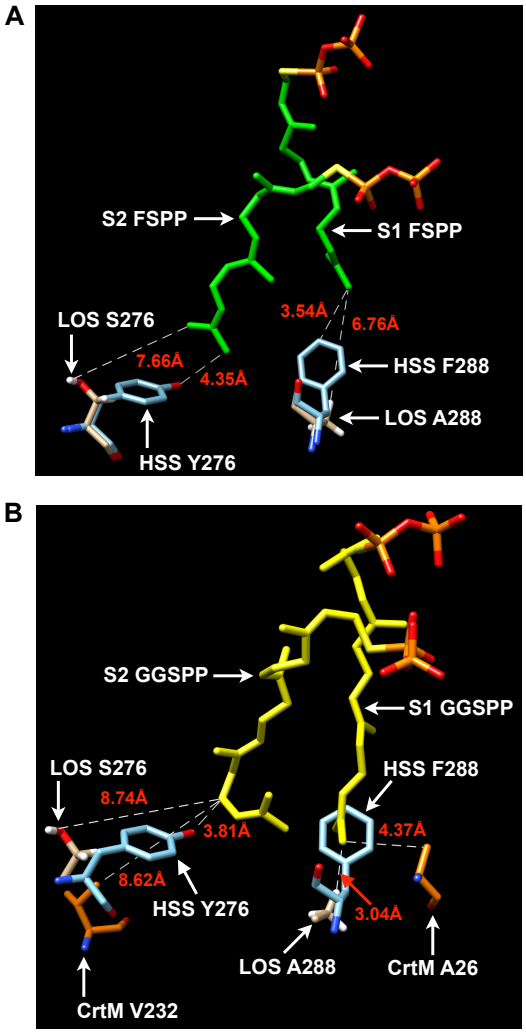


Figure 56. Structure comparison of HSS and LOS. Position comparison of the key residues controlling substrate binding from HSS, LOS, and CrtM in relation to the FSPP (A) and GGSPP (B) substrate analogs. **A.** Superposition of the HSS crystal structure bound to two FSPP molecules and the predicted LOS 3D-structure. **B.** Superposition of the CrtM crystal structure bound to two GGSPP molecules, the HSS crystal structure (FSPP removed), and the predicted LOS 3D-structure.

Additional characterization of the two substrate binding sites was conducted by comparing the HSS structure and the predicted LOS structure with the *Staphylococcus aureus* dehydrosqualene synthase (CrtM) structure bound to the GGPP analog geranylgeranylthiol diphosphate (GGSPP; Figure 56B)[96]. CrtM is an SS-like enzyme catalyzing the condensation of two FPP molecules to PSPP in the first step, followed by NADPH independent rearrangement in the second step to produce dehydrosqualene[96], which contains a *cis* double bond at 1-1' linkage between the two FPP molecules (Figure 57). Except for the NADPH binding residues, CrtM shares sequence similarity with the functionally conserved SS domains[127]. Most importantly, the substrate binding sites in the central cavity of CrtM and HSS are structurally homologous[94, 96]. Although the natural substrate of CrtM is FPP, mutation of the bulky F26 residue to Ala in the S1 substrate binding site was sufficient to allow binding of the longer GGPP substrate[127]. It should be noted the GGSPP bound CrtM crystal structure was not possible with WT CrtM and was obtained only with the F26A mutant[96].

The structural comparison of HSS, LOS, and $CrtM^{F26A}$ complexed with GGSPP reveals additional insights into substrate binding. The analysis showed that HSS F288 would clash with the GGSPP hydrophobic tail in the S1 substrate binding site (Figure 56B). In contrast, the distances from LOS A288 and CrtM A26 to the GGSPP tail are 3.04Å and 4.37Å, respectively, and would provide enough room for binding GGPP in both enzymes (Figure 56B). In the S2 substrate binding site the GGSPP tail, especially the terminal isoprene unit, is in a bent conformation (Figure 56B). The closest distance between the HSS Y276 hydroxyl group and GGSPP is to the first carbon of the terminal

GGSPP isoprene unit at 3.81Å (Figure 56B), and this distance suggests HSS Y276 would not interfere with GGPP binding in the S2 binding site if the S2 GGPP took on the configuration shown in Figure 56B. However, our data suggest the S2 GGPP may assume a more linear conformation in HSS since typical SS enzymes cannot utilize GGPP[59] indicating interference from S276. Additionally, mutation of the small LOS S276 to the bulky Tyr (M3 mutant) resulted in substantial loss of lycopaoctaene formation (Figure 53B), corroborating limited binding of GGPP due to interference from the Tyr as well as a more linear S2 GGPP conformation. The distance between the LOS S276 hydroxyl group and the nearest carbon of GGSPP is 8.74Å (Figure 56B), indicating that even with GGPP in the linear conformation there is sufficient room to accommodate GGPP in the S2 substrate binding site. This assumption is further supported by the 8.62Å distance between GGSPP and CrtM V232 in the S2 substrate binding site (Figure 56B). Altogether, our data support the hypothesis that small structural changes due to the replacement of the bulky Y276 and F288 residues with S276 and A288, respectively, in the substrate binding sites of an ancestral *B. braunii* race L SS enzyme resulted in the neofunctionalized LOS enzyme with the ability to utilize GGPP to produce the tetraterpenoid hydrocarbon lycopaoctaene. Although LOS evolved larger substrate binding pockets to accept GGPP, these larger substrate pockets do not prevent binding of the smaller FPP substrate, thus leading to the substrate promiscuity observed for this enzyme.

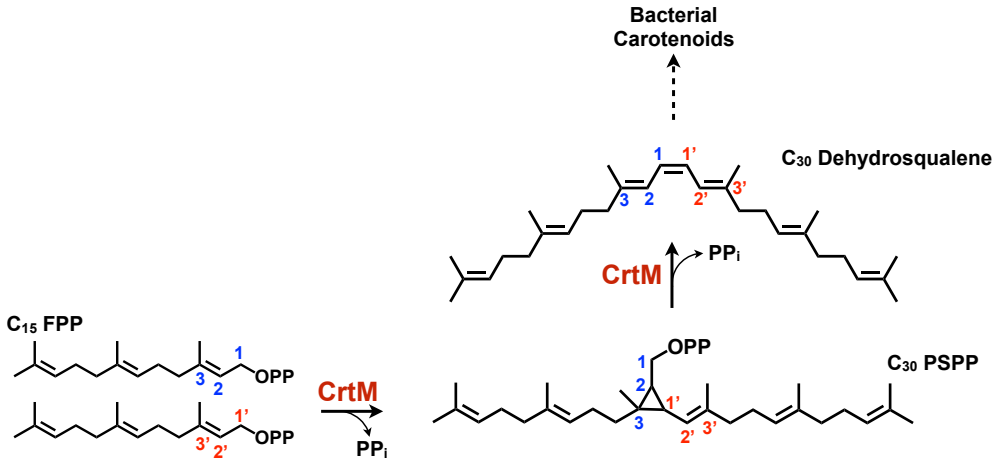


Figure 57. Dehydrosqualene synthase reaction. The two step reaction carried out by *Staphylococcus aureus* dehydrosqualene synthase (CrtM) to produce dehydrosqualene.

3.2b5 Additional mutants of the LOS enzyme

Now that the effect of the LOS M8 mutant (S276Y A288F) on the substrate binding sites has been identified, S276 and A288 of LOS were further characterize by replacing them with amino acids of different properties. Ten more mutants, five for S276 (M13-M17) and five for A288 (M18-M22) were made where the substituted amino acids are either aromatic, polar neutral, acidic, basic or aliphatic in nature (Table 3). Four additional residues L61 and Y69 at the S1 binding site, and F184 and I285 at the S2 binding site for were also selected mutagenesis experiments (Figure 58). These four residues of LOS are also conserved in other SS proteins (Figure 44) and the mutations are designed to increase specificity of GGPP binding in the substrate binding sites rather than eliminating FPP binding. Due to its smaller size compared to GGPP, it will be very challenging to limit FPP binding without affecting LOS affinity for GGPP. The L61Y (M23) mutation was chosen to create a new hydrophobic floor for the S1 GGPP tail, whereas the Y69V (M24) mutation was designed to increase the hydrophobicity of the S1 site (Figure 58). The mutation of F184 to leucine (M25) was made to test if this mutation increases the size of S2 site as well as increase the hydrophobic interaction with the S2 GGPP tail (Figure 58). Another mutation, I285Y, was designed to force the tail of GGPP in the direction of the S2 pocket created by S276 (Figure 58). In addition to single mutants of L61, Y69, F184 and I285 (M23-M26), mutations were also made in combination to yield six double mutants (M27-M32), four triple mutants (M33-M35) and one quadruple mutant M36 (Table 3). Due to time limitations these mutants were not

Mutant name	Mutation
M13	S276F
M14	S276T
M15	S276D
M16	S276K
M17	S276V
M18	A288Y
M19	A288S
M20	A288D
M21	A288H
M22	A288V
M23	L61Y
M24	Y69V
M25	F184L
M26	I285Y
M27	L61Y Y69V
M28	L61Y F184L
M29	L61Y I285Y
M30	Y69V F184L
M31	Y69V I285Y
M32	F184L I285Y
M33	L61Y Y69V F184L
M34	L61Y F184L I285Y
M35	L61Y Y69V I285Y
M35	Y69V F184L I285Y
M36	L61Y Y69V F184L I285Y

Table 3. List of LOS mutants for future study.

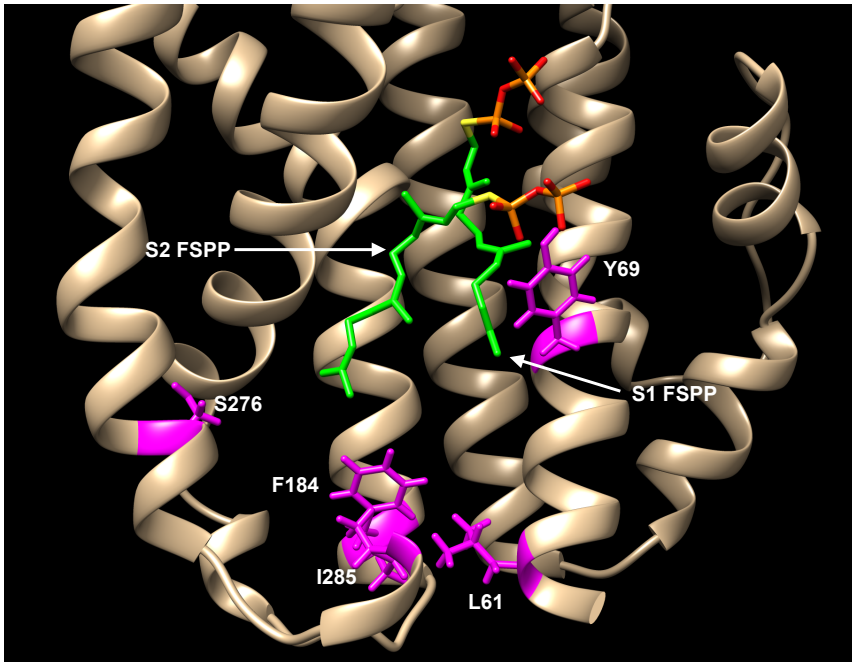


Figure 58. Superposition of the HSS/FSPP and LOS structures. Only FSPP from the HSS/FSPP complex is shown. For clarity, the LOS structure is focused on the enzyme active site. The four amino acids of LOS chosen for mutagenesis study are located in the two substrate binding sites and are highlighted in magenta. FSPP = farnesyl thiodiphosphate, HSS = human squalene synthase and LOS = lycopaoctaene synthase.

analyzed for activity, and future experiments should involve purification and characterization of these mutants *in vitro* followed by their analysis in *E.coli* cells producing high levels of FPP and GGPP.

3.3 Methods

3.3a Reagents

Radiolabeled substrates [1-^{3}H]-FPP (specific activity, 18.2 Ci/mmol) and [1-^{3}H]-GGPP (specific activity, 20.0 Ci/mmol) were purchased from PerkinElmer, and non-labeled FPP and GGPP were purchased from Sigma. All other chemicals were purchased from VWR unless otherwise noted.

3.3b Protein sequences for amino acid comparison

The following proteins were used for sequence comparison to LOS (AC# KT388101): LSS, *B. braunii* L race SS (AC# KT388100); BSS, *B. braunii* B race SS (AC# AH009227); AtSS, *Arabidopsis thaliana* SS (AC# D29017.1); YSS, yeast (*Saccharomyces cerevisiae*) SS (AC# M63979.1); HSS, human SS (AC# Q6IAX1).

3.3c Structure comparison

The LOS protein 3-D model was generated using the Iterative Threading ASSEmbly Refinement (I-TASSER) bioinformatics program[139-141]. Among five models of LOS predicted by I-TASSER, the model with a high confidence score (C-score = 0.19) was used for structural analysis in this study. The molecular graphics and structural analyses of the predicted LOS structure, the HSS crystal structure (PDB ID; 3WEF), and the CrtM crystal structure (PDB ID; 3AE0) were done using the USCF Chimera package[146].

3.3d Protein expression and purification

The expression constructs, LOS^{1-391}:pET28a designed to encode for 6xHis-LOS^{1-391} was made using NheI and HindIII restriction sites and LOS^{1-391}:pET28c SUMO designed to encode for 6xHis-SUMO-TEV-LOS^{1-391} was made using NdeI and HindIII restriction sites. Similarly, the double truncated LOS constructs, LOS^{33-381}:pET28a designed to encode for 6xHis-LOS^{33-381} and LOS^{33-381}:pET28c SUMO designed to encode for 6xHis-SUMO-TEV-LOS^{33-381} were made using NdeI and HindIII restriction sites. The desired expression construct was transformed into *E.coli* BL21(DE3), and an overnight culture from a single colony was used to inoculate (1.7% v/v) large volume of TB medium. The culture was grown at 37°C to OD_{600} = 0.8, and protein expression induced by adding 1 mM isopropyl β-D-thiogalactoside. The induced cultures were then grown for an additional 6 hours at 25°C, cells harvested by centrifugation, and the pellets were stored at -80°C for future use. Purification of 6xHis tagged proteins was conducted using Ni-NTA agarose resin (QIAGEN) following manufacture's recommendations with some modifications. All steps of purification were done either in ice or in a 4°C cold room. In a typical purification, the 750 mL culture pellet was suspended in 75 mL of lysis buffer, incubated for 30 min using a rocking table at cold room, and the cells were lysed by four successive probe sonication at 70% maximum power for 20 sec. DNase I and RNase A (Sigma) at a concentration of 1.2% w/v were added to the lysed cells, incubated for 15 min using a rocking table at cold room, and the samples were centrifuged at 16,000 x *g* for 30 min at 4°C. The supernatant was incubated with 4 mL of pre-equlibrated Ni-NTA resin for 60 min at 4°C using a rocking

table, the sample loaded into a gravity column, and the flow-through was discarded. The Ni-NTA column was washed with 80 mL of wash buffer, and the protein bound to nickel resin were eluted with 20 mL of elution buffer. The eluted protein sample were dialyzed against storage buffer (300 mM NaCl, 20 mM Tris-HCl, pH=7.5, 5 mM DTT and 2 mM $MgCl_2$), concentrated using Amicon Ultra centrifugal filter (30kDa cutoff; EMD Millipore) to the concentration of ~10mg/mL. The protein sample was loaded into a gel filtration column (HiLoad 16/600 Superdex 200pg; GE Helathcare) in ATKA prime plus FPLC system, and the sample was run using storage buffer at a flow rate of 0.5 mL/min and detection at 210 nm. The fractions corresponding to individual peak were collected, concentrated, flash frozen with liquid nitrogen and then stored at -80°C for future use.

The sodium phosphate based buffer used for purification are lysis buffer (50 mM sodium phosphate buffer (pH=7.8), 300 mM NaCl, 10 mM imidazole, 1x general protease inhibitor cocktail (Sigma), 1 mM $MgCl_2$, and 1% glycerol (v/v)), wash buffer (50 mM sodium phosphate buffer (pH=7.8), 300 mM NaCl, 40 mM Imidazole, 1 mM $MgCl_2$, and 1% glycerol (v/v)) and the elution buffer (50 mM sodium phosphate buffer (pH=7.8), 300 mM NaCl, 400 mM imidazole, 1 mM $MgCl_2$, and 1% glycerol (v/v)). Similarly, the MOPS based buffer used for purification are lysis buffer (50 mM MOPS (pH=7.8), 300 mM NaCl, 10 mM imidazole, 1x general protease inhibitor cocktail (Sigma), 20 mM $MgCl_2$, 2.5 mM DTT and 1% glycerol (v/v)), wash buffer (50 mM MOPS (pH=7.8), 300 mM NaCl, 40 mM imidazole, 20 mM $MgCl_2$, 2.5 mM DTT and 1% glycerol (v/v)), elution buffer (50 mM MOPS (pH=7.8), 300 mM NaCl, 400 mM imidazole, 20 mM $MgCl_2$, 2.5 mM DTT and 1% glycerol (v/v)) and the running buffer

(50 mM MOPS (pH=7.8), 300 mM NaCl, 20 mM $MgCl_2$ and 5 mM DTT and 1% glycerol (v/v)). For purification involving detergent, Triton X-100 at a final concentration of 0.1% was added to the lysis buffer.

The protein samples for enzyme assays were purified using MOPS based buffer using the protocol described above and did not utilize any detergents. The wash buffer contained higher concentration of imidazole (60mM) to remove non-specifically bounds protein to the nickel resin. Furthermore, the eluted protein after dialysis against running buffer was concentrated to 1-2 mg/mL, an equal volume glycerol added, and the protein sample stored at -20°C for shorter term or at -80°C for future use. For protein samples that were subjected to cleavage of 6xHis-SUMO tag, 6xHis-TEV protease was added to 20mL of eluted proteins (0.15mg of TEV/mg of eluted protein) followed by incubation at cold room for 17.5hrs without shaking. The protein samples after cleavage reaction were concentrated to ~3mL volume and buffer exchanged with 40mL of MOPS based running buffer using Amicon Ultra centrifugal filter (30kDa cutoff; EMD Millipore). The protein solution was further concentrated to ~12mL final volume followed by incubation with 1.5 mL of pre-equlibrated Ni-NTA resin at 4°C using a rocking table. After 60 min incubation, the sample was loaded into a gravity column and the flow-through was collected. The flow-through, protein samples without 6xHis-SUMO tag, was concentrated to desired protein concentration and analyzed by gel filtration chromatography as described above.

3.3e Site-directed mutagenesis of LOS

The LOS mutants were generated using LOS^{1-391}:pET28a construct. The desired point mutations were created using primers containing the altered nucleotide sequence(s). Conditions for PCR-mediated site directed mutagenesis reaction using Pfu Turbo DNA Polymerase (Agilent Technologies) were as follows: 98°C initial denaturation for 2 min, 17 cycles of denaturation, annealing, and extension (98°C for 30 sec, 55°C for 30 sec, 72°C for 8 min, respectively), and a final extension at 72°C for 10 min. All mutants were verified by DNA sequencing.

3.3f Radioactive *in vitro* enzyme assays

The enzyme assays were conducted using the protocol described previously[59]. In brief, a 50 µL reaction was initiated by adding 1 µg of purified protein to a reaction buffer containing 50 mM MOPS, pH 6.8, 2.5 mM β-mercaptoethanol, 20 mM $MgCl_2$, 2 mM NADPH, and 10 µM of ^{3}H-prenyl-PP (0.125 µCi; specific activity = 0.25 Ci/mmol). For the mixed substrate assay 10 µM of both ^{3}H-FPP and ^{3}H-GGPP was used (0.125 µCi for each substrate; specific activity = 0.25 Ci/mmol). The specific activity for each substrate was adjusted by adding non-labeled FPP and/or GGPP. The reactions were incubated for 60 min at 37°C and terminated by adding 60µL of *n*-hexane, followed by brief vortexing and centrifugation. Thirty microliters of the supernatant was then analyzed on Silica gel 60 TLC plates using *n*-hexane as the mobile phase. ^{3}H incorporation into reaction products was determined by scraping the spots corresponding to authentic standards of lycopaoctaene (R_f = 0.09), $C_{35}H_{58}$ (R_f = 0.12) and squalene (R_f = 0.17) followed by an analysis on a liquid scintillation counter.

3.3g *Ec*DXS and *Ec*FPS cloning

The cDNA sequences of *Ec*DXS (AC# NP_414954) and *Ec*FPS (a.k.a. ispA; AC# NP_414955) were obtained from the NCBI nucleotide database[147]. Genomic DNA from *E. coli* K-12 strain MG1655 and gene specific primers were used to amplify the PCR product using Phusion DNA polymerase followed by cloning into the pGEM-T vector (Promega). Gene specific primers were as follows: For *Ec*DXS, forward primer 5'-ATGAGTTTTGATATTGCCAAATACCCG-3' and reverse primer 5'-TTATGCCAGCCAGGCCTTGATTTTG-3'; for *Ec*FPS, forward primer 5'-ATGGACTTTCCGCAGCAACTC-3' and reverse primer 5'-TTATTTATTACGCTGGATGATGTAGTCCGC-3'.

3.3h *Ec*DXS, *Ec*FPS, and *At*GGPPS11 expression constructs

The DNA template for *At*GGPPS11 was obtained from our previous study[59] and the DNA sequence encoding the 56 N-terminal chloroplast targeting signal amino acids was deleted to yield *At*GGPPS11$^{57\text{-}371}$. The expression constructs were made using the following restriction sites and vectors: *Ec*DXS in NdeI and XhoI of pET22b, *Ec*FPS in BamHI and SalI of pACYDuet-1, and *At*GGPPS11$^{57\text{-}371}$ in NdeI and XhoI of pACYDuet-1. For experiments involving expression of *Ec*FPS and *At*GGPPS11$^{57\text{-}371}$, the two cDNAs were expressed using a single expression construct, *Ec*FPS + *At*GGPPS11$^{57\text{-}371}$ in pACYDuet-1, by cloning *Ec*FPS into the second multiple cloning site of AtGGPPS11$^{57\text{-}371}$:pACYDuet-1 using the BamHI and SalI restriction sites.

3.3i LOS *in vivo* mutant analysis in *E.coli*

The expression construct(s) described above were either transformed individually or in combination into *E.coli* BL21 (DE3) cells and the positive transformants were selected using standard procedures. The *E.coli* line harboring the desired construct(s) was grown in 60 mL of TB medium at 37°C to OD_{600} = 0.8, and gene expression was induced by adding 1mM isopropyl β-D-thiogalactoside. After induction, cells were grown for 6 hrs at 25°C, 50 mL of cells were harvested by centrifugation, the pellet snap frozen with liquid nitrogen, and the pellet lyophilized. Freeze-dried cells were transferred to a glass vial and extracted with *n*-hexane by vortexing vigorously at room temperature. The organic extracts were centrifuged at 1,000 x *g*, the supernatant dried, resuspended in 200 µL *n*-hexane, and a 5 µL aliquot analyzed by GC-MS.
GC-MS analyses were conducted using 5% Phenyl BR-5ms capillary column (30 m × 0.25 mm, film thickness: 0.25 µm) in electron ionization (70 eV) mode and utilized helium at a flow rate of 2.58mL/min as a carrier gas. The initial oven temperature was held at 220°C for 1 min followed by increase in temperature at a rate of 5°C/min to 280°C, ramped again to 300°C at a rate of 2°C/min, and held for 20 min. The temperature for injection port, interface and ion source were set at 280ºC, 250ºC, and 200ºC, respectively.

3.3j GC-MS analysis

A Bruker 436-GC-SCION SQ Premium GC-MS system was used to evaluate hydrocarbon production. GC-MS analyses were conducted using a 5% Phenyl BR-5ms capillary column (30 m × 0.25 mm, film thickness: 0.25 µm) in electron ionization (70

eV) mode using helium as a carrier gas at a flow rate of 2.58 mL/min. The initial oven temperature was held at 220°C for 1 min followed by an increase in temperature at a rate of 5°C/min to 280°C, ramped again to 300°C at a rate of 2°C/min, and held for 20 min. The temperature for the injection port, the interface, and ion source were set at 280ºC, 250ºC, and 200ºC, respectively. Quantitation for hydrocarbon production in *E. coli* was done based on the calibration curve generated with commercially available squalene standard (Sigma).

CHAPTER IV

CONCLUSIONS AND FUTURE DIRECTIONS

Among the three races (A, B and L) of *Botryococcus braunii*, race L was the last one to be isolated and characterized as a chemical race[41, 42]. Several strains of race L were first reported to be isolated from freshwater lakes of Ivory Coast (strains Kossou and Yamoussoukro) and Thailand (strain Songkla Nakarin)[42]. Lycopadiene was identified as the sole hydrocarbon produced by this alga accounting for up to 8% of the cellular dry weight, and lycopadiene was suggested as a precursor for the formation of lycopane fossil deposits found in oil shales and coal deposits[42]. Although lycopadiene was first isolated almost 30 years ago, nothing was known about the hydrocarbon biosynthesis until recently.

In this study, we have elucidated and characterized the biosynthetic pathway for the tetraterpenoid hydrocarbon lycopadiene (Figure 37). We have discovered a novel enzyme, lycopaoctaene synthase (LOS), that initiates the first committed step of the lycopadiene pathway (Figure 37). LOS is a squalene synthase (SS)-like enzyme and catalyzes a two-step reaction to produce lycopaoctaene (Figure 35A). The first step involves the dimerization of two molecules of geranylgeranyl diphosphate (GGPP) to produce prelycopaoctaene diphosphate as the cyclopropyl intermediate, which then undergoes NADPH dependent reductive rearrangement reaction to create a 1-1' (head-to-head) linkage between the two GGPP molecules to produce lycopaoctaene in the second step (Figure 35A). The eight double bond containing lycopaoctaene molecule is

sequentially reduced to produce lycopahexaene, lycopapentaene, lycopatetraene and lycopatriene as the intermediates, and finally to lycopadiene as the end product of the pathway (Figure 37).

4.1 Identification of reductase enzyme(s) for the lycopadiene biosynthetic pathway

The enzyme(s) catalyzing the reduction of lycopaoctaene to lycopadiene is yet to be identified and could be similar to geranylgeranyl reductase (GGR); an enzyme that catalyzes the sequential reduction of three double bonds of GGPP to produce phytyl diphosphate (PPP; Figure 38). Our pulse-chase experiment with ^{14}C labeled GGPP show accumulation of a majority of the radioactivity into lycopapentaene and lycopadiene, indicating the reduction of lycopapentaene to lycopatetraene as the rate limiting step of the pathway (Figures 27B and 37). Furthermore, this data suggests involvement of at least two reductase enzymes in the reduction of lycopaoctaene to lycopadiene. In order to identify the reductase enzymes in the lycopadiene pathway, two *GGR*-like cDNAs (*GGR-1* and *GGR-2*) from race L were cloned and characterized (Figure 39). The reductase activity of two candidate enzymes were tested by overexpressing these cDNAs in an *E.coli* line producing high levels of lycopaoctaene (Figures 41 and 42). However, preliminary data showed that the GGR-like enzymes do not reduce lycopaoctaene to its higher saturated homologues, indicating lycopaoctaene may not be their putative substrate (Figures 41 and 42).

It is possible that GGR-1 and GGR-2 could be involved in the reduction of lycopadiene pathway intermediates other than lycopaoctaene. This hypothesis can be tested by conducting *in vitro* reductase activity assays with purified proteins and one of

the lycopadiene intermediates as the substrate. Before testing reductase activity of the candidate enzymes *in vitro*, it may be best to develop the tetraterpenoid reductase assay using radiolabeled lycopaoctaene as the substrate and a lysate from race L as the protein source. Although commercially unavailable, radiolabeled lycopaoctaene can be generated using ^{3}H-GGPP and the LOS enzyme. The radioactive assay is preferred over assays with non-labeled substrates due to the high sensitivity of the radioactive assay compared to GC-MS analysis. Since we hypothesize the reaction conditions for reduction of lycopaoctaene to lycopadiene should be similar to that of GGR, reaction conditions for a tetraterpenoid reductase assay should be first optimized using the GGR reductase assay conditions as a template[121].

It is possible that GGR-1 and GGR-2 may not be the putative reductase enzymes involved in the lycopadiene biosynthetic pathway. It should be noted that GGR-1 and GGR-2 characterized in this study were identified from the transcriptome database of race L using the *Arabidopsis thaliana* GGR[122] as a query. Since GGR enzymes are conserved among all photosynthetic organisms and archaea[121-125], DNA sequences of GGR and GGR-like genes from these organisms can be used for more rigorous bioinformatic searches for candidate genes in the transcriptome as well the genome of race L. It is also possible that the current transcriptome data of race L may not be sufficient to identidy candidate reductase genes since the transcriptome was generated based on the RNA-Seq data from six culture time points (days 0, 3, 7, 14, 21 and 28) over the 4-week culture cycle. The reasoning for this argument is based on the time course analysis of LOS enzyme activity, growth rate and total hydrocarbon

production over a 39-day culture period (Figure 26B), which showed a direct correlation between LOS enzyme activity and hydrocarbon accumulation in this alga (Figure 26B). Furthermore, a rapid increase in the LOS activity was observed after day 30 of the culture cycle (Figure 26B). We hypothesize that expression of putative reductases will be up-regulated as a result of increased LOS expression. Thus, the current RNA-Seq data may not include these candidate genes. Identification of biosynthetic reductase genes could be accomplished by generating new transcriptome of race L based on the RNA-Seq data from multiple time points over the 8-week culture cycle. The genes with a similar expression profile to that of LOS can be first filtered from the transcriptome, the enzyme activities of new set of genes can be predicted in order to select the potential reductase genes, and the candidate genes can be tested both *in vitro* and *in vivo* for their ability to reduce either lycopaoctaene, lycopahexaene, lycopapentaene, lycopatetraene and lycopatriene to their corresponding higher saturated hydrocarbons. A similar transcriptome mining approach has been utilized to identify genes involved in the biosynthesis of many plant natural products[148-150].

Another approach to identify reductase genes for lycopadiene biosynthesis is by expressing a *B. braunii* race L cDNA library in a yeast line producing high levels of lycopaoctaene. Although LOS can make lycopaoctaene in yeast, the amount of lycopaoctaene generated in this line is low (Figure 30C) and may not be ideal for screening a cDNA library. Thus, a yeast line accumulating large amounts of lycopaoctaene should be generated, and could be achieved by overexpressing isoprenoid genes from the mevalonate pathway. Yeast should be used as a screening platform

instead of *E.coli* since the putative reductase enzymes are expected to have a membrane domain as observed for LOS enzyme. The needed cDNA library for screening of reductase enzyme is expected to have at least 10^6 independent clones. As a result, identification of yeast line with the ability to produce reduced forms of lycopaoctaene will require extensive screening of cell populations from the expression library, and will be labor intensive. Thus a high-throughput single-cell resolution screening platform that has the ability to distinguish reduced forms of lycopaoctaene should be developed. A high-throughput droplet microfluidics system is an ideal screening platform for this kind of study as it allows encapsulation of a single cell (independent line) in a droplet, and can function as an independent photobioreactor[151]. Our recent Raman spectroscopy analysis of tetraterpenoid hydrocarbons from race L shows unique Raman signatures of each hydrocarbon molecule and allows us to identify and distinguish between lycopaoctaene, lycopahexaene, lycopapentaene, lycopatetraene, lycopatriene and lycopadiene molecules[152]. Thus, a high-throughput droplet microfluidics platform coupled with Raman spectroscopy could be utilized for screening yeast cells expressing a race L cDNA library, and subsequently identify reductase enzymes involved in the lycopadiene biosynthetic pathway. Raman spectroscopy integrated with a droplet microfluidics platform has been recently developed and was successfully utilized for characterization of microalgal lipids[153].

4.2 Substrate and product specificity of the LOS enzyme

The LOS enzyme utilizes C_{20} GGPP as a substrate to produce C_{40} lycopaoctaene, the first committed step of the lycopadiene biosynthetic pathway. LOS is

promiscuous in its ability to utilize prenyl diphosphate substrates, and can catalyze C_{15} FPP to produce C_{30} squalene and C_{20} PPP to produce C_{40} lycopadiene (Figure 32). Furthermore, LOS can make chimeric hydrocarbons of all combinations using mixtures of the three substrates FPP, GGPP and PPP (Figure 32). To gain structural insights into the mechanism of how LOS utilizes three prenyl diphosphate substrates to produce combinatorial hydrocarbons of varied chain lengths (C_{30}, C_{35} and C_{40}), we pursued crystallization of the LOS protein as detailed in chapter III. Although we were successful in generating large amounts of highly pure, soluble, and active LOS^{1-391} protein by removing the C-terminal transmembrane domain, we were unable to crystallize this protein even in the presence of substrate analogs farnesylthiol diphosphate (FSPP) and geranylgeranylthiol diphosphate (GGSPP). The double truncated construct of LOS (LOS^{33-381}) was also generated based on the comparison of the predicted LOS structure and the published human squalene synthase (HSS) crystal structure (Figure 45B)[94]. Although enzymatically active, we were unable to generate large amounts of highly pure LOS^{33-381} protein as previously seen for LOS^{1-391}. Future experiments should involve extensive optimization of purification conditions to increase the yield as well the purity of LOS^{33-381}, and should be followed by screening of crystallization conditions with substrate analogs FSPP and GGSPP.

In order to identify the critical amino acids involved in conferring substrate and product specificity of LOS, we conducted rational mutagenesis experiments based on the sequence alignment of LOS with several SS proteins as well as comparison of the predicted LOS structure with the HSS/FSPP complex crystal structure (Figures 44 and

52)[94]. Our analysis of LOS mutants suggests that five residues (T65, M180, S276, A288 and V289), located on the hydrophobic floor of the two substrate binding sites, control the binding of GGPP in this enzyme (Figures 52 and 53). When these five residues of LOS were mutated to the corresponding residues found in HSS (V69, L183, Y276, F288 and C289), the resulting LOS mutant (M12) almost completely lost its ability to produce lycopaoctaene whereas SS activity was increased by 3-fold (Figure 53B). Among the five residues, S276 and A288 play a dominant role in defining the two substrate binding pockets of the LOS enzyme. When the double mutant M8 (S276Y and A288F) was designed to resemble the typical SS enzyme, binding of the longer GGPP substrate was limited but binding of the smaller substrate FPP was favored as seen by a drastic increase in SS activity (Figure 53B). Furthermore, the *in vitro* enzyme activity of select mutants (M3, M4, M8 and M12) were verified by expressing each mutant in *E.coli* engineered to produce high levels of FPP and GGPP (Figure 55). Intriguingly, when LOS was expressed along with three isoprenoid synthases (*Ec*DXS, *Ec*FPS and *At*GGPPS11) in *E.coli*, lycopaoctaene production was favored over $C_{35}H_{58}$ and squalene suggesting the substrate and product specificity of LOS enzyme *in vivo* could result by metabolic channeling process (Figure 54).

Altogether, characterization of the LOS mutants *in vitro* and *in vivo* along with the structural comparison of LOS with the HSS/FSPP complex and a dehydrosqualene synthase (CrtM)/GGSPP complex suggests that the small residues S276 and A288 created larger substrate binding pockets in LOS, allowing for binding of GGPP and therefore its novel activity for lycopaoctaene production (Figures 53, 55 and 56).

However, the larger substrate binding sites of LOS did not inhibit FPP binding, and thus result in its substrate promiscuity. Considering the crystal structure of WT LOS will be obtained, future experiments should include crystallization of LOS and its mutants (M3, M4, M8 and M12) with FSPP, GGSPP and mixture of both substrate analogs as it will provide new structural insights into the substrate binding and product specificity of this enzyme. In addition, directed evolution experiments with LOS should also be conducted in an *E.coli* line accumulating high levels of FPP and GGPP substrates. Such experiments will not only identify additional residues involved in substrate binding but may also result in LOS variants with higher/improved activity for either squalene, $C_{35}H_{58}$ or lycopaoctaene production. Altogether the studies in chapter II extended our understanding of the neofunctionalization of the LOS enzyme from the progenitor SS enzyme.

REFERENCES

[1] Moody, J. W., McGinty, C. M., and Quinn, J. C. (2014) Global evaluation of biofuel potential from microalgae, *Proc. Natl. Acad. Sci. U S A 111*, 8691-8696.

[2] Chisti, Y. (2013) Constraints to commercialization of algal fuels, *J. Biotechnol. 167*, 201-214.

[3] Kröger, M., and Müller-Langer, F. (2014) Review on possible algal-biofuel production processes, *Biofuels 3*, 333-349.

[4] Hillen, L. W., Pollard, G., Wake, L. V., and White, N. (1982) Hydrocracking of the oils of *Botryococcus braunii* to transport fuels, *Biotechnol. Bioeng. 24*, 193-205.

[5] Banerjee, A., Sharma, R., Chisti, Y., and Banerjee, U. C. (2002) *Botryococcus braunii*: a renewable source of hydrocarbons and other chemicals, *Crit. Rev. Biotechnol. 22*, 245-279.

[6] Chisti, Y. (2010) Fuels from microalgae, *Biofuels 1*, 233-235.

[7] Haik, Y., Selim, M. Y. E., and Abdulrehman, T. (2011) Combustion of algae oil methyl ester in an indirect injection diesel engine, *Energy 36*, 1827-1835.

[8] Johnson, M., and Wen, Z. (2009) Production of biodiesel fuel from the microalga *Schizochytrium limacinum* by direct transesterification of algal biomass, *Energy Fuels 23*, 5179-5183.

[9] Shurin, J. B., Burkart, M. D., Mayfield, S. P., and Smith, V. H. (2016) Recent progress and future challenges in algal biofuel production, *F1000Res 5*.

[10] Chisti, Y. (2007) Biodiesel from microalgae, *Biotechnol. Adv. 25*, 294-306.

[11] Cordell, D., Drangert, J. O., and White, S. (2009) The story of phosphorus: Global food security and food for thought, *Global Environ. Chang. 19*, 292-305.

[12] Gilbert, N. (2009) Environment: The disappearing nutrient, *Nature 461*, 716-718.

[13] Carvalho, A. P., Meireles, L. A., and Malcata, F. X. (2006) Microalgal reactors: A review of enclosed system designs and performances, *Biotechnol. Progr. 22*, 1490-1506.

[14] Zamboni, A., Murphy, R. J., Woods, J., Bezzo, F., and Shah, N. (2011) Biofuels carbon footprints: Whole-systems optimisation for GHG emissions reduction, *Bioresource Technol 102*, 7457-7465.

[15] Gimpel, J. A., Specht, E. A., Georgianna, D. R., and Mayfield, S. P. (2013) Advances in microalgae engineering and synthetic biology applications for biofuel production, *Curr. Opin. Chem. Biol. 17*, 489-495.

[16] Radakovits, R., Jinkerson, R. E., Fuerstenberg, S. I., Tae, H., Settlage, R. E., Boore, J. L., and Posewitz, M. C. (2012) Draft genome sequence and genetic transformation of the oleaginous alga *Nannochloropis gaditana*, *Nat. Commun. 3*.

[17] Courchesne, N. M. D., Parisien, A., Wang, B., and Lan, C. Q. (2009) Enhancement of lipid production using biochemical, genetic and transcription factor engineering approaches, *J. Biotechnol. 141*, 31-41.

[18] Illman, A. M., Scragg, A. H., and Shales, S. W. (2000) Increase in *Chlorella* strains calorific values when grown in low nitrogen medium, *Enzyme Microb. Tech. 27*, 631-635.

[19] Rodolfi, L., Zittelli, G. C., Bassi, N., Padovani, G., Biondi, N., Bonini, G., and Tredici, M. R. (2009) Microalgae for oil: Strain Selection, induction of lipid synthesis and outdoor mass cultivation in a low-cost photobioreactor, *Biotechnol. Bioeng. 102*, 100-112.

[20] Khozin-Goldberg, I., and Cohen, Z. (2011) Unraveling algal lipid metabolism: Recent advances in gene identification, *Biochimie 93*, 91-100.

[21] Stephenson, P. G., Moore, C. M., Terry, M. J., Zubkov, M. V., and Bibby, T. S. (2011) Improving photosynthesis for algal biofuels: toward a green revolution, *Trends Biotechnol. 29*, 615-623.

[22] Giordano, M., Beardall, J., and Raven, J. A. (2005) CO2 concentrating mechanisms in algae: Mechanisms, environmental modulation, and evolution, *Annu. Rev. Plant Biol. 56*, 99-131.

[23] Meyer, M., and Griffiths, H. (2013) Origins and diversity of eukaryotic CO2-concentrating mechanisms: lessons for the future, *J. Exp. Bot 64*, 769-786.

[24] Badger, M. R., and Price, G. D. (2003) CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution, *J. Exp. Bot. 54*, 609-622.

[25] Savile, C. K., and Lalonde, J. J. (2011) Biotechnology for the acceleration of carbon dioxide capture and sequestration, *Curr Opin Biotech 22*, 818-823.

[26] Long, S. P., Zhu, X. G., Naidu, S. L., and Ort, D. R. (2006) Can improvement in photosynthesis increase crop yields?, *Plant Cell Environ. 29*, 315-330.

[27] Chen, M., and Blankenship, R. E. (2011) Expanding the solar spectrum used by photosynthesis, *Trends Plant Sci. 16*, 427-431.

[28] Zehr, J. P. (2011) Nitrogen fixation by marine cyanobacteria, *Trends Microbiol. 19*, 162-173.

[29] Chaurasia, A. K., and Apte, S. K. (2011) Improved eco-friendly recombinant *Anabaena sp*. strain PCC7120 with enhanced nitrogen biofertilizer potential, *Appl Environ Microb 77*, 395-399.

[30] Christenson, L., and Sims, R. (2011) Production and harvesting of microalgae for wastewater treatment, biofuels, and bioproducts, *Biotechnol. Adv. 29*, 686-702.

[31] Craggs, R. J., Heubeck, S., Lundquist, T. J., and Benemann, J. R. (2011) Algal biofuels from wastewater treatment high rate algal ponds, *Water Sci. Technol. 63*, 660-665.

[32] Yoshida, M., Tanabe, Y., Yonezawa, N., and Watanabe, M. M. (2012) Energy innovation potential of oleaginous microalgae, *Biofuels 3*, 761-781.

[33] Atadashi, I. M., Aroua, M. K., and Aziz, A. A. (2010) High quality biodiesel and its diesel engine application: A review, *Renew. Sust. Ener. Rev. 14*, 1999-2008.

[34] Lee, S. Y., Kim, H. M., and Cheon, S. (2015) Metabolic engineering for the production of hydrocarbon fuels, *Curr. Opin. Biotechnol. 33*, 15-22.

[35] Rezanka, T., Zahradnik, J., and Podojil, M. (1982) Hydrocarbons in green and blue-green-algae, *Folia Microbiol. 27*, 450-454.

[36] Metzger, P., and Largeau, C. (2005) *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids, *Appl. Microbiol. Biotechnol. 66*, 486-496.

[37] Weiss, T. L., Roth, R., Goodson, C., Vitha, S., Black, I., Azadi, P., Rusch, J., Holzenburg, A., Devarenne, T. P., and Goodenough, U. (2012) Colony organization in the green alga *Botryococcus braunii* (Race B) is specified by a complex extracellular matrix, *Eukaryot. Cell 11*, 1424-1440.

[38] Weiss, T. L., Chun, H. J., Okada, S., Vitha, S., Holzenburg, A., Laane, J., and Devarenne, T. P. (2010) Raman spectroscopy analysis of botryococcene hydrocarbons from the green microalga *Botryococcus braunii*, *J. Biol. Chem. 285*, 32458-32466.

[39] Achitouv, E., Metzger, P., Rager, M. N., and Largeau, C. (2004) C31-C34 methylated squalenes from a Bolivian strain of *Botryococcus braunii*, *Phytochemistry 65*, 3159-3165.

[40] Huang, Z., and Poulter, C. D. (1989) Tetramethylsqualene, a triterpene from *Botryococcus-braunii* var Showa, *Phytochemistry 28*, 1467-1470.

[41] Metzger, P., Berkaloff, C., Casadevall, E., and Coute, A. (1985) Alkadiene-Producing and Botryococcene-Producing Races of Wild Strains of Botryococcus-Braunii, *Phytochemistry 24*, 2305-2312.

[42] Metzger, P., and Casadevall, E. (1987) Lycopadiene, a tetraterpenoid hydrocarbon from new strains of the green-alga *Botryococcus braunii*, *Tetrahedron Lett. 28*, 3931-3934.

[43] Metzger, P., Pouet, Y., and Summons, R. (1997) Chemtotaxonomic evidence for the similarity between *Botryococcus braunii* L race and *Botryococcus neglectus*, *Phytochemistry 44*, 1071-1075.

[44] Largeau, C., Casadevall, E., Berkaloff, C., and Dhamelincourt, P. (1980) Sites of accumulation and composition of hydrocarbons in *Botryococcus-braunii*, *Phytochemistry 19*, 1043-1051.

[45] Metzger, P., Pouet, Y., Bischoff, R., and Casadevall, E. (1993) An aliphatic polyaldehyde from *Botryococcus-braunii* (A Race), *Phytochemistry 32*, 875-883.

[46] Metzger, P., Rager, M. N., and Largeau, C. (2007) Polyacetals based on polymethylsqualene diols, precursors of algaenan in *Botryococcus braunii* race B, *Org. Geochem. 38*, 566-581.

[47] Metzger, P., Allard, B., Casadevall, E., Berkaloff, C., and Coute, A. (1990) Structure and Chemistry of a New Chemical Race of Botryococcus-Braunii (Chlorophyceae) That Produces Lycopadiene, a Tetraterpenoid Hydrocarbon, *J Phycol 26*, 258-266.

[48] Niehaus, T. D., Kinison, S., Okada, S., Yeo, Y. S., Bell, S. A., Cui, P., Devarenne, T. P., and Chappell, J. (2012) Functional identification of triterpene methyltransferases from *Botryococcus braunii* race B, *J. Biol. Chem. 287*, 8163-8173.

[49] Pan, J. J., Solbiati, J. O., Ramamoorthy, G., Hillerich, B. S., Seidel, R. D., Cronan, J. E., Almo, S. C., and Poulter, C. D. (2015) Biosynthesis of squalene from farnesyl diphosphate in bacteria: three steps catalyzed by three enzymes, *ACS Cent. Sci. 1*, 77-82.

[50] Glikson, M., Lindsay, K., and Saxby, J. (1989) *Botryococcus*- A planktonic green alga, the source of petroleum through the ages: Transmission electron microscopical studies of oil shales and petroleum source rocks, *Org. Geochem. 14*, 14.

[51] Lichtfouse, E., Derenne, S., Mariotti, A., and Largeau, C. (1994) Possible algal origin of long chain odd n-alkanes in immature sediments as revealed by distributions and carbon isotope ratios, *Org. Geochem. 22*, 1023-1027.

[52] Adam, P., Schaeffer, P., and Albrecht, P. (2006) C40 monoaromatic lycopane derivatives as indicators of the contribution of the alga *Botryococcus braunii* race L to the organic matter of Messel oil shale (Eocene, Germany), *Org. Geochem. 37*, 584-596.

[53] Mckirdy, D. M., Cox, R. E., Volkman, J. K., and Howell, V. J. (1986) Botryococcane in a new class of Australian non-marine crude oils, *Nature 320*, 57-59.

[54] Moldowan, J. M., and Seifert, W. K. (1980) First discovery of botryococcane in petroleum, *J. C. S. Chem. Commun. 19*, 912-914.

[55] Wolf, F. R., Nemethy, E. K., Blanding, J. H., and Bassham, J. A. (1985) Biosynthesis of Unusual Acyclic Isoprenoids in the Alga Botryococcus-Braunii, *Phytochemistry 24*, 733-737.

[56] Hou, L., Park, H., Okada, S., and Ohama, T. (2014) Release of single cells from the colonial oil-producing alga *Botryococcus braunii* by chemical treatments, *Protoplasma 251*, 191-199.

[57] Okada, S., Devarenne, T. P., and Chappell, J. (2000) Molecular characterization of squalene synthase from the green microalga *Botryococcus braunii*, race B, *Arch. Biochem. Biophys. 373*, 307-317.

[58] Niehaus, T. D., Okada, S., Devarenne, T. P., Watt, D. S., Sviripa, V., and Chappell, J. (2011) Identification of unique mechanisms for triterpene biosynthesis in *Botryococcus braunii*, *Proc. Natl. Acad. Sci. U S A 108*, 12260-12265.

[59] Thapa, H. R., Naik, M. T., Okada, S., Takada, K., Molnar, I., Xu, Y., and Devarenne, T. P. (2016) A squalene synthase-like enzyme initiates production of tetraterpenoid hydrocarbons in *Botryococcus braunii* Race L, *Nat. Commun. 7*, 11198.

[60] Jiang, Z. D., Kempinski, C., Bush, C. J., Nybo, S. E., and Chappell, J. (2016) Engineering triterpene and methylated triterpene production in plants provides biochemical and physiological insights into terpene metabolism, *Plant Physiol. 170*, 702-716.

[61] Zhao, L. S., Chang, W. C., Xiao, Y. L., Liu, H. W., and Liu, P. H. (2013) Methylerythritol phosphate pathway of isoprenoid biosynthesis, *Annu. Rev. Biochem. 82*, 497-+.

[62] Vranova, E., Coman, D., and Gruissem, W. (2012) Structure and dynamics of the isoprenoid pathway network, *Mol. Plant 5*, 318-333.

[63] Vranova, E., Coman, D., and Gruissem, W. (2013) Network analysis of the MVA and MEP pathways for isoprenoid synthesis, *Annu. Rev. Plant Biol. 64*, 665-700.

[64] Barkovich, R., and Liao, J. C. (2001) Metabolic engineering of isoprenoids, *Metab. Eng. 3*, 27-39.

[65] Kirby, J., and Keasling, J. D. (2009) Biosynthesis of plant isoprenoids: perspectives for microbial engineering, *Annu. Rev. Plant Biol. 60*, 335-355.

[66] Eisenreich, W., Rohdich, F., and Bacher, A. (2001) Deoxyxylulose phosphate pathway to terpenoids, *Trends. Plant. Sci. 6*, 78-84.

[67] Sato, Y., Ito, Y., Okada, S., Murakami, M., and Abe, H. (2003) Biosynthesis of the triterpenoids, botryococcenes and tetramethylsqualene in the B race of *Botryococcus braunii* via the non-mevalonate pathway, *Tetrahedron Lett. 44*, 7035-7037.

[68] Xu, W., Ma, X., and Wang, Y. (2016) Production of squalene by microbes: an update, *World J. Microb. Biot. 32*.

[69] Chun, H. J., Weiss, T. L., Devarenne, T. P., and Laane, J. (2013) Vibrational spectra and DFT calculations of squalene, *J. Mol. Struct. 1032*, 203-206.

[70] Tsujimoto, M. (1919) A highly unsaturated hydrocarbon in shark liver oil - Correction, *J. Ind. Eng. Chem-Us 11*, 798-798.

[71] Tsujimoto, M. (1920) Squalene a highly unsaturated hydrocarbon in shark liver oil, *J. Ind. Eng. Chem-Us 12*, 63-72.

[72] Reddy, L. H., and Couvreur, P. (2009) Squalene: A natural triterpene for use in disease management and therapy, *AdV. Drug Deliver. Rev. 61*, 1412-1426.

[73] Valachovic, M., and Hapala, I. (2017) Biosynthetic approaches to squalene production: The case of Yeast, *Methods Mol. Biol. 1494*, 95-106.

[74] Rude, M. A., and Schirmer, A. (2009) New microbial fuels: a biotech perspective, *Curr. Opin. in Microbiol. 12*, 274-281.

[75] Tracy, N. I., Crunkleton, D. W., and Price, G. L. (2011) Catalytic cracking of squalene to gasoline-range molecules, *Biomass Bioenerg. 35*, 1060-1065.

[76] Jennings, S. M., Tsay, Y. H., Fisch, T. M., and Robinson, G. W. (1991) Molecular-cloning and characterization of the yeast gene for squalene synthetase, *P Natl Acad Sci USA 88*, 6038-6042.

[77] Nakashima, T., Inoue, T., Oka, A., Nishino, T., Osumi, T., and Hata, S. (1995) Cloning, expression, and characterization of cDNAs encoding *Arabidopsis-thaliana* squalene synthase, *P Natl Acad Sci USA 92*, 2328-2332.

[78] Ohtake, K., Saito, N., Shibuya, S., Kobayashi, W., Amano, R., Hirai, T., Sasaki, S., Nakano, C., and Hoshino, T. (2014) Biochemical characterization of the water-soluble squalene synthase from *Methylococcus capsulatus* and the functional analyses of its two DXXD(E)D motifs and the highly conserved aromatic amino acid residues, *Febs J. 281*, 5479-5497.

[79] Robinson GW, T. Y., Kienzle BK, Smith-Monroy CA, Bishop RW. (1993) Conservation between human and fungal squalene synthetases: similarities in structure, function, and regulation, *Mol. Cell. Biol. 13*, 12.

[80] Poulter, C. D. (1990) Biosynthesis of non-head-to-tail terpenes - formation of 1'-1 and 1'-3 linkages, *Acc. Chem. Res. 23*, 70-77.

[81] Gu, P., Ishii, Y., Spencer, T. A., and Shechter, I. (1998) Function-structure studies and identification of three enzyme domains involved in the catalytic activity in rat hepatic squalene synthase, *J. Biol. Chem. 273*, 12515-12525.

[82] Stamellos, K. D., Shackelford, J. E., Shechter, I., Jiang, G., Conrad, D., Keller, G. A., and Krisans, S. K. (1993) Subcellular localization of squalene synthase in rat hepatic cells. Biochemical and immunochemical evidence, *J. Biol. Chem. 268*, 12825-12836.

[83] Shechter, I., Klinger, E., Rucker, M. L., Engstrom, R. G., Spirito, J. A., Islam, M. A., Boettcher, B. R., and Weinstein, D. B. (1992) Solubilization, purification, and characterization of a truncated form of rat hepatic squalene synthetase, *J. Biol. Chem. 267*, 8628-8635.

[84] Tansey, T. R., and Shechter, I. (2000) Structure and regulation of mammalian squalene synthase, *Bba-Mol. Cell Biol. L. 1529*, 49-62.

[85] Peng, B. Y., Plan, M. R., Chrysanthopoulos, P., Hodson, M. P., Nielsen, L. K., and Vickers, C. E. (2017) A squalene synthase protein degradation method for improved sesquiterpene production in *Saccharomyces cerevisiae*, *Metab. Eng. 39*, 209-219.

[86] Do, R., Kiss, R. S., Gaudet, D., and Engert, J. C. (2009) Squalene synthase: a critical enzyme in the cholesterol biosynthesis pathway, *Clin. Genet. 75*, 19-29.

[87] Hartmann, M. A. (1998) Plant sterols and the membrane environment, *Trends Plant. Sci. 3*, 170-175.

[88] Devarenne, T. P., Ghosh, A., and Chappell, J. (2002) Regulation of squalene synthase, a key enzyme of sterol biosynthesis, in tobacco, *Plant Physiol. 129*, 1095-1106.

[89] Chappell, J., Vonlanken, C., Vogeli, U., and Bhatt, P. (1989) Sterol and sesquiterpenoid biosynthesis during a growth-cycle of tobacco cell-suspension cultures, *Plant Cell Rep. 8*, 48-52.

[90] Vogeli, U., and Chappell, J. (1988) Induction of sesquiterpene cyclase and suppression of squalene synthetase activities in plant-cell cultures treated with fungal elicitor, *Plant Physiol. 88*, 1291-1296.

[91] Memon, R. A., Shechter, I., Moser, A. H., Shigenaga, J. K., Grunfeld, C., and Feingold, K. R. (1997) Endotoxin, tumor necrosis factor, and interleukin-1 decrease hepatic squalene synthase activity, protein, and mRNA levels in Syrian hamsters, *J. Lipid Res. 38*, 1620-1629.

[92] Pandit, J., Danley, D. E., Schulte, G. K., Mazzalupo, S., Pauly, T. A., Hayward, C. M., Hamanaka, E. S., Thompson, J. F., and Harwood, H. J., Jr. (2000) Crystal structure of human squalene synthase. A key enzyme in cholesterol biosynthesis, *J. Biol. Chem. 275*, 30610-30617.

[93] Liu, C. I., Jeng, W. Y., Chang, W. J., Ko, T. P., and Wang, A. H. (2012) Binding modes of zaragozic acid A to human squalene synthase and *Staphylococcal* dehydrosqualene synthase, *J. Biol. Chem 287*, 18750-18757.

[94] Liu, C. I., Jeng, W. Y., Chang, W. J., Shih, M. F., Ko, T. P., and Wang, A. H. J. (2014) Structural insights into the catalytic mechanism of human squalene synthase, *Acta Crystallogr. D 70*, 231-241.

[95] Mookhtiar, K. A., Kalinowski, S. S., Zhang, D. L., and Poulter, C. D. (1994) Yeast squalene synthase - a mechanism for addition of substrates and activation by NADPH, *J. Biol. Chem. 269*, 11201-11207.

[96] Lin, F. Y., Liu, C. I., Liu, Y. L., Zhang, Y., Wang, K., Jeng, W. Y., Ko, T. P., Cao, R., Wang, A. H., and Oldfield, E. (2010) Mechanism of action and inhibition of dehydrosqualene synthase, *Proc. Natl .Acad. Sci. U S A 107*, 21337-21342.

[97] Jarstfer, M. B., Blagg, B. S. J., Rogers, D. H., and Poulter, C. D. (1996) Biosynthesis of squalene. Evidence for a tertiary cyclopropylcarbinyl cationic intermediate in the rearrangement of presqualene diphosphate to squalene, *J. Am. Chem. Soc. 118*, 13089-13090.

[98] Templier, J., Largeau, C., and Casadevall, E. (1984) A mechanism of non-isoprenoid hydrocarbon biosynthesis in *Botryococcus-braunii*, *Phytochemistry 23*, 1017-1028.

[99] Templier, J., Largeau, C., and Casadevall, E. (1987) Effect of various inhibitors on biosynthesis of non-isoprenoid hydrocarbons in *Botryococcus-braunii*, *Phytochemistry 26*, 377-383.

[100] Templier, J., Largeau, C., and Casadevall, E. (1991) Nonspecific elongation decarboxylation in biosynthesis of cis-alkadienes and trans-alkadienes by *Botryococcus-braunii*, *Phytochemistry 30*, 175-183.

[101] Cornejo-Corona, I., Thapa, H., Devarenne, T., and Lozoya-Gloria, E. (2015) The biofuel potential of the green colonial microalga *Botryococcus braunii*, In *Microalgae and other phototrophic bacteria* (Torres-Bustillos, L., Ed.), pp 41-58, Nova Science Publishers, New York.

[102] Dennis, M., and Kolattukudy, P. E. (1992) A Cobalt-porphyrin enzyme converts a fatty aldehyde to a hydrocarbon and CO *P Natl Acad Sci USA 89*, 5306-5310.

[103] Dennis, M. W., and Kolattukudy, P. E. (1991) Alkane biosynthesis by decarbonylation of aldehyde catalyzed by a microsomal preparation from *Botryococcus-braunii*, *Arch. Biochem. Biophys. 287*, 268-275.

[104] Wang, X., and Kolattukudy, P. E. (1995) Solubilization and purification of aldehyde-generating fatty acyl-CoA reductase from green-alga *Botryococcus-braunii*, *Febs Lett. 370*, 15-18.

[105] Okada, S., Devarenne, T. P., Murakami, M., Abe, H., and Chappell, J. (2004) Characterization of botryococcene synthase enzyme activity, a squalene synthase-like activity from the green microalga *Botryococcus braunii*, Race B, *Arch. Biochem. Biophys. 422*, 110-118.

[106] Khan, N. E., Nybo, E. S., Chappell, J., and Curtis, W. R. (2015) Triterpene hydrocarbon production engineered into a metabolically versatile host—*Rhodobacter capsulatus*, *Biotechnol. Bioeng. 112*, 1523-1532.

[107] Rohmer, M. (1999) The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants, *Nat. Prod. Rep. 16*, 565-574.

[108] Rohmer, M. (2010) *Hopanoids*, Springer, Berin Heidelberg.

[109] Zhang, Z., Metzger, P., and Sachs, J. P. (2007) Biomarker evidence for the co-occurrence of three races (A, B and L) of *Botryococcus braunii* in El Junco Lake, Gala´pagos, *Org. Geochem. 38*, 20.

[110] Sato, T., Yoshida, S., Hoshino, H., Tanno, M., Nakajima, M., and Hoshino, T. (2011) Sesquarterpenes (C35 terpenes) biosynthesized via the cyclization of a linear C35 isoprenoid by a tetraprenyl-beta-curcumene synthase and a tetraprenyl-beta-curcumene cyclase: identification of a new terpene cyclase, *J. Am. Chem. Soc. 133*, 9734-9737.

[111] Thai, L., Rush, J. S., Maul, J. E., Devarenne, T., Rodgers, D. L., Chappell, J., and Waechter, C. J. (1999) Farnesol is utilized for isoprenoid biosynthesis in plant cells via farnesyl pyrophosphate formed by successive monophosphorylation reactions, *P Natl Acad Sci USA 96*, 13080-13085.

[112] Inoue, H., Korenaga, T., Sagami, H., Koyama, T., and Ogura, K. (1994) Phosphorylation of farnesol by a cell-free system from *Botryococcus braunii*, *Biochem. Biophys. Res. Commun. 200*, 1036-1041.

[113] Derenne, S., Largeau, C., Casadevall, E., and Berkaloff, C. (1989) Occurrence of a resistant biopolymer in the L race of *Botryococcus braunii*, *Phytochemistry 28*, 1137-1142.

[114] Derenne, S., Largeau, C., Casadevall, E., and Sellier, N. (1990) Direct relationship between the resistant biopolymer and the tetraterpenic hydrocarbon in the lycopadiene race of *Botryococcus braunii*, *Phytochemistry 29*, 2187-2192.

[115] Beck, G., Coman, D., Herren, E., Ruiz-Sola, M. A., Rodriguez-Concepcion, M., Gruissem, W., and Vranova, E. (2013) Characterization of the GGPP synthase gene family in *Arabidopsis thaliana*, *Plant Mol. Biol. 82*, 393-416.

[116] Ignea, C., Trikka, F. A., Nikolaidis, A. K., Georgantea, P., Ioannou, E., Loupassaki, S., Kefalas, P., Kanellis, A. K., Roussis, V., Makris, A. M., and Kampranis, S. C. (2015) Efficient diterpene production in yeast by engineering Erg20p into a geranylgeranyl diphosphate synthase, *Metab. Eng. 27*, 65-75.

[117] Qureshi, A. A., Barnes, F. J., Semmler, E. J., and Porter, J. W. (1973) Biosynthesis of prelycopersene pyrophosphate and lycopersene by squalene synthetase, *J. Biol. Chem. 248*, 2755-2767.

[118] Baxter, A., Fitzgerald, B. J., Hutson, J. L., McCarthyS, A. D., Motteram, J. M., Ross, B. C., Sapra, M., Snowden, M. A., Watson, N. S., Williamsll, R. J., and Wright, C. (1992) Squalestatin 1, a potent inhibitor of squalene synthase, which lowers serum cholesterol in vivo, *J. Biol. Chem. 267*, 11705-11708.

[119] Bergstrom, J. D., Kurtz, M. M., Rew, D. J., Amend, A. M., Karkas, J. D., Bostedor, R. G., Bansal, V. S., Dufresne, C., Vanmiddlesworth, F. L., Hensens, O. D., Liesch, J. M., Zink, D. L., Wilson, K. E., Onishi, J., Milligan, J. A., Bills, G., Kaplan, L., Omstead, M. N., Jenkins, R. G., Huang, L., Meinz, M. S., Quinn, L., Burg, R. W., Kong, Y. L., Mochales, S., Mojena, M., Martin, I., Pelaez, F., Diez, M. T., and Alberts, A. W. (1993) Zaragozic Acids - a family of fungal metabolites that are picomolar competitive inhibitors of squalene synthase, *P Natl Acad Sci USA 90*, 80-84.

[120] Lindsey, S., and Harwood, H. J. (1995) Inhibition of mammalian squalene synthetase-activity by zaragozic acid a is a result of competitive-inhibition followed by mechanism-based irreversible inactivation, *J. Biol. Chem. 270*, 9083-9096.

[121] Keller, Y., Bouvier, F., d'Harlingue, A., and Camara, B. (1998) Metabolic compartmentation of plastid prenyllipid biosynthesis--evidence for the involvement of a multifunctional geranylgeranyl reductase, *Eur. J. Biochem. 251*, 413-417.

[122] Tanaka, R., Oster, U., Kruse, E., Rudiger, W., and Grimm, B. (1999) Reduced activity of geranylgeranyl reductase leads to loss of chlorophyll and tocopherol

and to partially geranylgeranylated chlorophyll in transgenic tobacco plants expressing antisense RNA for geranylgeranyl reductase, *Plant Physiol. 120*, 695-704.

[123] Addlesee, H. A., and Hunter, C. N. (1999) Physical mapping and functional assignment of the geranylgeranyl-bacteriochlorophyll reductase gene, bchP, of Rhodobacter sphaeroides, *J Bacteriol 181*, 7248-7255.

[124] Murakami, M., Shibuya, K., Nakayama, T., Nishino, T., Yoshimura, T., and Hemmi, H. (2007) Geranylgeranyl reductase involved in the biosynthesis of archaeal membrane lipids in the hyperthermophilic archaeon *Archaeoglobus fulgidus*, *Febs J. 274*, 805-814.

[125] Kung, Y., McAndrew, R. P., Xie, X. K., Liu, C. C., Pereira, J. H., Adams, P. D., and Keasling, J. D. (2014) Constructing tailored isoprenoid products by structure-guided modification of geranylgeranyl reductase, *Structure 22*, 1028-1036.

[126] Reiling, K. K., Yoshikuni, Y., Martin, V. J. J., Newman, J., Bohlmann, J., and Keasling, J. D. (2004) Mono and diterpene production in *Escherichia coli*, *Biotechnol. Bioeng. 87*, 200-212.

[127] Umeno, D., and Arnold, F. H. (2004) Evolution of a pathway to novel long-chain carotenoids, *J Bacteriol 186*, 1531-1536.

[128] Tobias, A. V., and Arnold, F. H. (2006) Biosynthesis of novel carotenoid families based on unnatural carbon backbones: a model for diversification of natural product pathways, *Biochim. Biophys. Acta 1761*, 235-246.

[129] Ortiz de Montellano, P. R., Wei, J. S., Vinson, W. A., Castillo, R., and Boparai, A. S. (1977) Substrate selectivity of squalene synthetase, *Biochemistry 16*, 2680-2685.

[130] Koyama, T., Ogura, K., and Seto, S. (1980) Substrate specificity of squalene synthetase, *Biochim. Biophys. Acta 617*, 218-224.

[131] Vining, L. C. (1992) Roles of secondary metabolites from microbes., *Ciba Found. Symp. 171*, 184-194.

[132] Metzger, P., Allard, B., Casadevall, E., Berkaloff, C., and Coute, A. (1990) Structure and chemistry of a new chemical race of *Botryococcus braunii* (Chlorophyceae) that produces lycopadiene, a tetraterpenoid hydrocarbon, *J. Phycol. 26*, 258-266.

[133] Metzger, P., David, M., and Casadevall, E. (1987) Biosynthesis of Triterpenoid Hydrocarbons in the B-Race of the Green-Alga Botryococcus-Braunii - Sites of Production and Nature of the Methylating Agent, *Phytochemistry 26*, 129-134.

[134] Beller, H. R., Lee, T. S., and Katz, L. (2015) Natural products as biofuels and bio-based chemicals: fatty acids and isoprenoids, *Nat. Prod. Rep. 32*, 1508-1526.

[135] Gupta, P., and Phulara, S. C. (2015) Metabolic engineering for isoprenoid-based biofuel production, *J. Appl. Microbiol. 119*, 605-619.

[136] Behar, F., Derenne, S., and Largeau, C. (1995) Closed pyrolyses of the isoprenoid algaenan of *Botryococcus-braunii*, L race - Geochemical implications for derived kerogens, *Geochim. Cosmochim. Acta 59*, 2983-2997.

[137] Beytia, E., Qureshi, A. A., and Porter, J. W. (1973) Squalene synthetase mechanism of reaction, *J. Biol. Chem. 248*, 1856-1867.

[138] Thompson, J. F., Danley, D. E., Mazzalupo, S., Milos, P. M., Lira, M. E., and Harwood, H. J., Jr. (1998) Truncation of human squalene synthase yields active, crystallizable protein, *Arch. Biochem. Biophys. 350*, 283-290.

[139] Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction, *Nat. Protoc. 5*, 725-738.

[140] Roy, A., Yang, J. Y., and Zhang, Y. (2012) COFACTOR: an accurate comparative algorithm for structure-based protein function annotation, *Nucleic Acids Res. 40*, W471-W477.

[141] Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction, *Bmc Bioinformatics 9*.

[142] Carson, M., Johnson, D. H., McDonald, H., Brouillette, C., and DeLucas, L. J. (2007) His-tag impact on structure, *Acta Crystallogr. D 63*, 295-301.

[143] Panavas, T., Sanders, C., and Butt, T. R. (2009) SUMO fusion technology for enhanced protein production in prokaryotic and eukaryotic expression systems, *Methods Mol. Biol. 497*, 303-317.

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

[145] Jorgensen, K., Rasmussen, A. V., Morant, M., Nielsen, A. H., Bjarnholt, N., Zagrobelny, M., Bak, S., and Moller, B. L. (2005) Metabolon formation and metabolic channeling in the biosynthesis of plant natural products, *Curr. Opin. Plant Biol. 8*, 280-291.

[146] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF chimera - A visualization system for exploratory research and analysis, *J. Comput. Chem. 25*, 1605-1612.

[147] Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., ColladoVides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome sequence of *Escherichia coli* K-12, *Science 277*, 1453-&.

[148] Klein, A. P., and Sattely, E. S. (2015) Two cytochromes P450 catalyze S-heterocyclizations in cabbage phytoalexin biosynthesis, *Nat. Chem. Biol. 11*, 837-839.

[149] Lau, W., and Sattely, E. S. (2015) Six enzymes from mayapple that complete the biosynthetic pathway to the etoposide aglycone, *Science 349*, 1224-1228.

[150] Rajniak, J., Barco, B., Clay, N. K., and Sattely, E. S. (2015) A new cyanogenic metabolite in *Arabidopsis* required for inducible pathogen defence, *Nature 525*, 376-379.

[151] Kim, H. S., Guzman, A. R., Thapa, H. R., Devarenne, T. P., and Han, A. (2016) A droplet microfluidics platform for rapid microalgal growth and oil production analysis, *Biotechnol. Bioeng. 113*, 1691-1701.

[152] Chun, H. J., Waqued, S., Thapa, H. R., Han, A., Yakovlev, V. V., Laane, J., and Devarenne, T. P. (2017) Raman spectra and DFT calculations for tetraterpene hydrocarbons from the L race of the green microalga *Botryococcus braunii*, *J. Mol. Struct. 1129*, 216-221.

[153] Kim, H. S., Waqued, S. C., Nodurft, D. T., Devarenne, T. P., Yakovlev, V. V., and Han, A. (2017) Raman spectroscopy compatible PDMS droplet microfluidic culture and analysis platform towards on-chip lipidomics, *Analyst 142*, 1054-1060.