

NEW STRUCTURAL INSIGHTS INTO THE EXTRACELLULAR MATRIX AND
HYDROCARBON DIVERSITY OF *Botryococcus Braunii* RACE B

A Dissertation

by

MEHMET TATLI

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,	Jaan Laane
	Vladislav M. Panin
	Paul A. Lindahl
Head of Department,	Gregory D. Reinhart

May 2018

Major Subject: Biochemistry

Copyright 2018 Mehmet Tatli

ABSTRACT

Botryococcus braunii is a colonial green microalga that can produce large amounts of liquid isoprenoid hydrocarbons known as botryococcenes, which can be easily converted into conventional combustion engine fuels. *B. braunii* colony cells are held together by a complex extracellular matrix (ECM). ECM serves as a storage unit for liquid hydrocarbons, and contains a retaining wall and a polysaccharide fibrillar sheath. Analysis of “shells” revealed a single protein. Here we use peptide mass fingerprinting and bioinformatics to identify this protein called polysaccharide associated protein (PSAP). PSAP does not show similarity to any protein in databases, and contains several hydroxyproline domains and a predicted sugar binding domain. Staining studies confirm PSAP as a glycoprotein, and mass spectrometry analysis identified ten N-linked glycosylation sites comprising seven different glycans containing mostly mannose and N-acetylglucosamine with fucose and arabinose. Additionally, four hydroxyproline residues have short O-linked glycans of mainly arabinose and galactose, with 6-deoxyhexose. PSAP secretion and localization to shell material is confirmed using western blot analysis and microscopy. These studies indicate PSAP contains unique glycans and suggest its involvement in ECM polysaccharide fiber biosynthesis.

Moreover, *B. braunii* is capable of producing many botryococcene isomers ranging in carbon number from 30 to 37, as well as several methylated squalenes. Although studies have reported the isolation of isomers based on the C₃₁, C₃₂, C₃₃, and C₃₄ botryococcene

structures, there has not been a report of cyclic C₃₃ botryococenes or an isomer based on the trimethylsqualene structure until this study. Three cyclic C₃₃ botryococenes and one new trimethylsqualene isomer were isolated from the B race, Showa (Berkeley) strain of *B. braunii*.

In addition, a detailed examination of the molecular structure for ten botryococenes including the newly discovered cyclic C₃₃ botryococenes and five methylsqualenes isolated from race B was carried out using Raman spectroscopy and density functional theory (DFT) calculations in an effort to distinguish between these structurally similar molecules by spectroscopic approaches. A comparison of the experimental Raman spectroscopy and DFT calculations indicates several spectral regions such as those for $\nu(\text{C}=\text{C})$ stretching, CH₂/CH₃ bending, and ring bending can be used to distinguish between these molecules. Altogether, these studies bring more insights into understanding the hydrocarbon diversity and ECM of *B. braunii* race B.

DEDICATION

I would primarily dedicate this dissertation to my mother. I would also like to dedicate this dissertation to all my family for always supporting me and sending good vibes from thousands of miles away. I cannot imagine how difficult my life would have been without feeling your support with all the decisions I have made in my life.

ACKNOWLEDGEMENTS

I would like to thank my Ph.D. thesis advisor Dr. Timothy Devarenne for his guidance and support that he has provided me throughout my Ph.D. career. I am grateful for his mentorship, patience, and the chance he has given me to grow as a scientist in his lab.

I would like to thank my committee, Dr. Jaan Laane, Dr. Vlad Panin, Dr. Paul Lindahl and Dr. Andreas Holzenburg for their time, guidance, and support throughout the course of this research.

My sincere thanks to Dr. Shigeru Okada and Dr. Mandar Naik for their support and suggestions on my projects. Thanks to my colleagues and current and past members of Devarenne lab – in particular Daniel Browne whom I had many valuable discussions and good times over the years with, Dr. Hem Thapa, Dongyin Su, and Incheol Yeo for their valuable suggestions. I would also like to thank the NSF REU Summer students Laura Krings and Patrick Sweet, and Texas A&M Energy Institute summer student Alyssa Dubrow that worked with me and contributed to my work and allowed me improve my mentorship skills.

Last, I would like to thank all my loving family and close friends for their great support.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Timothy P. Devarenne [advisor] of the Department of Biochemistry and Biophysics, Dr. Jaan Laane of the Department of Chemistry, Dr. Vladislav M. Panin of the Department of Biochemistry and Biophysics, and Dr. Paul A. Lindahl of the Departments of Chemistry and Biochemistry and Biophysics.

Most of the work for this dissertation was carried out by the student, under the supervision of Dr. Timothy P. Devarenne of the Department of Biochemistry and Biophysics. The NMR analysis of the hydrocarbons in section II was done with the help of the following collaborators: Dr. Mandar Naik of the Biomolecular NMR facility in the Department of Biochemistry and Biophysics at Texas A&M University (currently at the Department of Molecular Pharmacology, Physiology, and Biotechnology at Brown University) and Dr. Shigeru Okada of the School of Agricultural and Life Sciences at the University of Tokyo. Collaborators for section III are as follows: The DFT calculations for the hydrocarbons were done by Dr. Hye-Jin Chun and Dr. Jaan Laane of the Department of Chemistry at Texas A&M University, BCARS experiments were done by Dr. Charles H. Camp Jr. and Dr. Marcus Cicerone of the Biosystems and Biomaterials Division at the National Institute of Standards and Technology in Gaithersburg, Maryland, and in vitro Raman spectroscopy experiments of the purified hydrocarbons were

completed with the help of Dr. Wei-Chuan Shih and Jingting Li of the Department of Electrical and Computer Engineering at University of Houston. Collaborators and their contributions to section IV are as follows: Dr. Mayumi Ishihara of the Complex Carbohydrate Research Center at University of Georgia did the glycomics studies for PSAP and Dr. Christian Heiss and Dr. Parastoo Azadi of the Complex Carbohydrate Research Center at University of Georgia provided help planning experiments for the glycomics studies for PSAP. Dr. Lawrence J. Dangott of the Protein Chemistry Laboratory in the Department of Biochemistry and Biophysics and Daniel R. Browne of the Department of Biochemistry and Biophysics at Texas A&M University provided help with the bioinformatics to find full length cDNA for PSAP.

Funding sources

This work was made possible in part by NSF-EFRI-PSBR under grant #1240478 to Dr. Timothy P. Devarenne; by the Robert A. Welch Foundation under grant #A-0396 to Dr. Jaan Laane; by National Science Foundation under grant #CHE-0541587 to the Texas A&M Department of Chemistry Medusa computer system; by a grant from the Japan Science and Technology (JST) agency program to Dr. Shigeru Okada.; by the U.S. Department of Energy under grant #DE-SC0015662 and by NIH under grant #1S10OD018530 and grant # P41GM103490-14 to Dr. Parastoo Azadi.

TABLE OF CONTENTS

	Page
ABSTRACT	ii
DEDICATION	iv
ACKNOWLEDGEMENTS	v
CONTRIBUTORS AND FUNDING SOURCES.....	vi
TABLE OF CONTENTS	viii
LIST OF FIGURES	xi
LIST OF TABLES	xv
 1. INTRODUCTION AND LITERATURE REVIEW	 1
1.1 Alternative energy demand and biofuels	1
1.2 Microalgae.....	2
1.2.1 Microalgae-based biofuels	3
1.2.2 Limitations of microalgae-based biofuels	4
1.2.3 Potential and future of microalgae for biofuel production	6
1.2.4 Microalgae-based hydrocarbons.....	9
1.3 <i>Botryococcus braunii</i>	10
1.3.1 Ultrastructure of <i>B. braunii</i>	14
1.4 ECMs in various organisms	15
1.4.1 ECMs in plants	15
1.4.2 ECMs in green microalgae	16
1.4.3 ECMs and EPSs in <i>B. braunii</i> and contrast to other organisms	18
1.5 Isoprenoid biosynthesis	20
1.6 Hydrocarbon biosynthesis in <i>B.braunii</i>	23
1.6.1 Hydrocarbon biosynthesis in race A	25
1.6.2 Hydrocarbon biosynthesis in race L.....	25
1.6.3 Hydrocarbon biosynthesis in race B.....	28
1.7 Raman Spectroscopy	31
1.7.1 Raman spectroscopy studies in microalgae.....	34
1.7.2 Raman spectroscopy studies in <i>B. braunii</i>	35

2.	ISOLATION AND CHARACTERIZATION OF CYCLIC C ₃₃ BOTRYOCOCCENES AND A TRIMETHYLSQUALENE ISOMER FROM <i>BOTRYOCOCCUS BRAUNII</i> RACE B	37
2.1	Background and rationale.....	37
2.2	Results and discussion.....	39
2.2.1	Purification of <i>B. braunii</i> hydrocarbons.....	39
2.2.2	Identification of purified <i>B. braunii</i> hydrocarbons.....	40
2.2.3	NMR and structure analysis	46
2.3	Experimental Section	65
2.3.1	General experimental procedures	65
2.3.2	Culturing and harvesting <i>Botryococcus braunii</i> race B	66
2.3.3	Extraction and purification of race B hydrocarbons	66
2.3.4	GC-MS analysis	70
3.	RAMAN SPECTRA AND DFT CALCULATIONS FOR BOTRYOCOCCENE AND METHYLSQUALENE HYDROCARBONS FROM THE B RACE OF THE GREEN MICROALGA <i>BOTRYOCOCCUS BRAUNII</i>	71
3.1	Background and rationale.....	71
3.2	Results and discussion.....	74
3.2.1	Density function theory (DFT) computations for bond frequencies of B race hydrocarbons.....	74
3.2.2	Comparison of experimental and calculated full Raman spectra for B race hydrocarbons	82
3.2.3	Raman spectral regions that can distinguish between hydrocarbons	83
3.2.4	Distinguishing between botryococcene isomers using broadband CARS	86
3.3	Experimental Methods	90
3.3.1	Culturing of <i>B. braunii</i>	90
3.3.2	Hydrocarbon purification	91
3.3.3	Raman spectroscopy	92
3.3.4	Theoretical computations	93
4.	POLYSACCHARIDE ASSOCIATED PROTEIN (PSAP) FROM THE GREENMICROALGA <i>BOTRYOCOCCUS BRAUNII</i> IS A UNIQUE EXTRA- CELLULAR MATRIX HYDROXYPROLINE-RICH GLYCOPROTEIN	95
4.1	Background and rationale.....	95
4.2	Results	98
4.2.1	Identification of cDNA and protein sequences for polysaccharide associated protein	98
4.2.2	PSAP is a glycoprotein.....	102

4.2.3 Identification of PSAP N-linked glycans	105
4.2.4 Identification of PSAP hydroxyproline O-linked glycans	112
4.2.5 Estimation of PSAP molecular mass based on glycosylations.....	113
4.2.6 PSAP contains a functional secretion signal peptide	116
4.2.7 Localization of PSAP to shell material	116
4.3 Discussion	127
4.3.1 PSAP is a unique hydroxyproline-rich glycoprotein (HRGP)	127
4.3.2 PSAP has unique N-linked glycans.....	129
4.3.3 Potential functional roles for PSAP.....	130
4.4 Experimental procedures.....	133
4.4.1 Identifying the PSAP full length cDNA.....	133
4.4.2 In-gel glycoprotein staining.....	134
4.4.3 PSAP N-linked glycan site mapping.....	134
4.4.4 Free N-linked glycan preparation for linkage analysis	136
4.4.5 N-linked glycan linkage analysis	137
4.4.6 N-linked oligosaccharide profiling.....	138
4.4.7 Hydroxyproline O-linked glycan analysis.....	138
4.4.8 Functional analysis of the PSAP secretion signal peptide	139
4.4.9 Lectin blot.....	141
4.4.10 PSAP peptide antibody generation and western blot analysis	141
4.4.11 PSAP localization to shell material	142
5. CONCLUSIONS AND FUTURE DIRECTIONS	145
5.1 Hydrocarbon diversity of <i>B. braunii</i> race B.....	145
5.2 <i>In vitro</i> Raman spectroscopy analysis of race B hydrocarbons	146
5.3 Polysaccharide Associated Protein (PSAP) from the ECM of Green Microalga <i>Botryococcus braunii</i>	150
REFERENCES.....	152
APPENDIX A	184

LIST OF FIGURES

	Page
Figure 1 Model and image of <i>Botryococcus braunii</i>	12
Figure 2 Types of hydrocarbons produced by each race of <i>Botryococcus braunii</i>	13
Figure 3 Production of various isoprenoids via MEP pathway.....	22
Figure 4 Biosynthesis of <i>B. braunii</i> race B hydrocarbons	24
Figure 5 Two possible routes to C27 alkadiene production in race A of <i>B. braunii</i>	26
Figure 6 The proposed pathway for lycopadiene biosynthesis in <i>B. braunii</i> race L.....	27
Figure 7 Botryococcene isomers from the B race of <i>B. braunii</i>	30
Figure 8 Structures of <i>B. braunii</i> race B hydrocarbons.....	38
Figure 9 Chromatograph of <i>B. braunii</i> race B hydrocarbons from C18 HPLC	41
Figure 10 GC-MS total ion count profiles of each C18 HPLC fraction (A-H) shown in Figure 9	42
Figure 11 Structures for known botryococcenes and methylsqualenes isolated in this study	44
Figure 12 GC-MS total ion count profiles and structures for 5 (A), 6 (B), 7, (C) and 8 (D)	45
Figure 13 The EI mass spectra from GC-MS and structures for 5 (A), 6 (B), 7 (C) and 8 (D)	47
Figure 14 The mass spectra from chemical ionization GC-MS for 5 (A), 6 (B), 7 (C), and 8 (D)	48
Figure 15 High resolution mass spectra from chemical ionization MS for 5 (A), 6 (B), 7 (C), and 8 (D)	49
Figure 16 Multiplicity edited ¹³ C-HSQC spectra for 5 (A), 6 (B), 7 (C) and 8 (D).....	55

Figure 17 Selected correlations observed in 2D NMR experiments for 5	57
Figure 18 Correction of C ₃₃ botryococcene, molecule 3-2, structure reported in Okada, S., Murakami, M., and Yamaguchi, K. (1997) Phytochemistry 8, 198- 203	58
Figure 19 ¹³ C-HMBC spectrum acquired on 5.....	59
Figure 20 ¹ H-COSY spectrum acquired on 5	60
Figure 21 ¹³ C-TOCSY-HSQC spectrum acquired on 5	61
Figure 22 ¹³ C-H2BC spectrum acquired on 5	62
Figure 23 Proposed mechanism for ring formation in 5.	64
Figure 24 ECD spectra for molecules 5 (A), 6 (B), 7 (C), and 8 (D).....	69
Figure 25 Hydrocarbons purified from the B race of <i>B. braunii</i> for use in Raman spectroscopy for this study	73
Figure 26 Experimental and calculated spectra for botryococcenes	77
Figure 27 Experimental and calculated spectra for methylsqualenes	78
Figure 28 The experimental and calculated ν(C=C) stretching Raman spectral regions for botryococcenes and methylsqualenes	79
Figure 29 The experimental and calculated Raman spectral regions containing δ(CH ₃) deformation, ρ(CH ₃) rock, ω(CH ₂) wag, t(CH ₂) twist, ωi(CH) wag (in-plane), and ν(C-C) stretch for botryococcenes and methylsqualenes	80
Figure 30 The experimental and calculated cyclohexane b(Ring) bending Raman spectral regions for botryococcenes	81
Figure 31 Spectral differentiation of C ₃₀ botryococcene, C ₃₁ isoshowacene, C ₃₂ botryococcene, and C ₃₂ braunicene by BCARS	89
Figure 32 Analysis of PSAP molecular mass by SDS-PAGE	99
Figure 33 Identification of <i>B. braunii</i> race B transcriptome contigs matching PSAP MS peptides	100

Figure 34 Nucleotide sequence of the three <i>B. braunii</i> race B contigs that matched MS peptides from PSAP	101
Figure 35 Identification of the PSAP cDNA and protein sequences using bioinformatics tools	103
Figure 36 Nucleotide and protein sequence of PSAP	104
Figure 37 In-gel glycoprotein staining of PSAP	106
Figure 38 PSAP N-linked on O-linked glycan analysis	107
Figure 39 PSAP N-linked glycan analysis by NSI-MSn.....	110
Figure 40 Structures of the N-linked PSAP glycans	111
Figure 41 Analysis of PSAP Hydroxyproline O-linked glycans by LC-MS	115
Figure 42 Functional analysis of the PSAP signal peptide	117
Figure 43 Electron micrographs of <i>B. braunii</i> race B shell material	119
Figure 44 Lectin blot and western blot detection of PSAP from shell material.....	120
Figure 45 Detection of PSAP in shell material by fluorescence microscopy using PSAP glycan specific lectins and the α -PSAP antibody (α -PSAP).....	122
Figure 46 Detection of PSAP in shell material by fluorescence microscopy using lectins specific for PSAP glycans	123
Figure 47 Light microscopy detection of PSAP in shell material using lectins specific for PSAP glycans and DAB	124
Figure 48 Detection of PSAP in shell material by fluorescence microscopy using the PSAP specific antibody (α -PSAP)	126
Figure 49 Identification of disordered regions in the PSAP protein sequence	132
Figure A-1 Overlay of 1D ^1H NMR spectra for C ₃₃ -2 trimethylsqualene isomer 8 (top, magenta), cyclic C ₃₃ -3 botryococcene 7 (green), cyclic C ₃₃ -2 botryococcene 6 (red), and cyclic C ₃₃ -1 botryococcene 5 (bottom, blue) is shown here.....	184
Figure A-2 1D ^1H NMR spectra for cyclic C ₃₃ -1 botryococcene 5.....	185

Figure A-3 1D ^1H NMR spectra for cyclic C_{33} -2 botryococcene 6.....	186
Figure A-4 1D ^1H NMR spectra for cyclic C_{33} -3 botryococcene 7.....	187
Figure A-5 1D ^1H NMR spectra for C_{33} -2 trimethylsqualene isomer 8.....	188
Figure A-6 Overlay of 1D ^{13}C NMR spectra for C_{33} -2 trimethylsqualene isomer 8 (top, magenta), cyclic C_{33} -3 botryococcene 7 (green), cyclic C_{33} -2 botryococcene 6 (red), and cyclic C_{33} -1 botryococcene 5 (bottom, blue) is shown here.....	189
Figure A-7 1D ^{13}C NMR spectra for cyclic C_{33} -1 botryococcene 5	190
Figure A-8 1D ^{13}C NMR spectra for cyclic C_{33} -2 botryococcene 6	191
Figure A-9 1D ^{13}C NMR spectra for cyclic C_{33} -3 botryococcene 7	192
Figure A-10 1D ^{13}C NMR spectra for C_{33} -2 trimethylsqualene isomer 8.....	193

LIST OF TABLES

	Page
Table 1 ^{13}C NMR assignments for hydrocarbons from race B. δ denotes ^{13}C chemical shifts determined for carbons of each hydrocarbon molecule.	43
Table 2 ^{13}C and ^1H NMR Spectroscopic Data (^1H 500 MHz, ^{13}C 125 MHz, CDCl_3) for 5.....	50
Table 3 ^{13}C and ^1H NMR Spectroscopic Data (^1H 500 MHz, ^{13}C 125 MHz, CDCl_3) for 6, 7 ^a , and 8.....	51
Table 4 ^{13}C assignments for previously identified hydrocarbons used for comparison to 5-7 and 8. δ denotes ^{13}C chemical shifts determined for each hydrocarbon molecule.....	52
Table 5 ^{13}C and ^1H NMR Spectroscopic Data for 5.....	53
Table 6 ^{13}C and ^1H NMR Spectroscopic Data for 6.....	54
Table 7 Vibrations of <i>B. braunii</i> race B botryococenes.	75
Table 8 Vibrations of <i>B. braunii</i> race B squalenes.....	76
Table 9 Linkage analysis of the <i>N</i> -linked glycans on PSAP.	109

1. INTRODUCTION AND LITERATURE REVIEW

1.1 Alternative energy demand and biofuels

The growing energy demand across the world because of the rapidly growing population, fast consumption of the current energy supplies, and most importantly the increasing emission of greenhouse gases in the last century have driven researchers, governments, and energy organizations to seek for alternative energy sources in order to supply the large energy demand not only in a quantity manner but also in a clean energy manner¹. Even though there has been a long standing use of the fossil feedstocks from petroleum and coal to generate commonly used products such as waxes, lubricants, fuel, fertilizers, asphalt, etc. in industry, these processes are not considered to be sustainable or environment friendly^{2, 3}. One of the major problems is the large amount of CO₂ released upon burning fossil fuels, which triggers global warming upon accumulation of high concentrations of CO₂ in the atmosphere¹. Some of the sustainable energy sources that have been considered as an alternative to fossil fuels include wind, solar, biofuel, and so on. Each one of these renewable energy sources possesses its own advantages and disadvantages in terms of politics, economics, and applications⁴. However, biofuel production has been considered one of the promising sustainable alternative energy candidates as it is derived from a biological system, which can provide continuous supply of feedstocks used in biofuel production¹. There are several types of biofuels based on the sources they are obtained from. ‘First generation’ biofuels, i.e. bio-ethanol, biodiesel, and biogas are obtained from food crops like oils seeds, corn, sugar cane, etc. and viewed as

ground breaking in transition from fossil based fuels to green fuels^{5, 6}. Though these biofuels provide some advantages in terms of CO₂ cycle, large scale production, and are commercialized widely, there still have been strong debates about their use as fuels. Some of the issues with first generation biofuels include using large amounts of land and energy input to obtain these biofuels as well as other economic barriers⁷. But, the major problem with first generation biofuels is the use of food crops to produce biofuels, which in return causes an increase in food prices⁵. Because of the above-mentioned limitations of first generation biofuels, focus has switched to ‘second generation’ biofuels. These biofuels are derived from lignocellulosic biomass, which can be obtained from non-food part of the plants grown mainly for the biofuel production purposes, and industrial wastes⁶. Second generation biofuels help reduce CO₂ release into the atmosphere, overcome the fuel versus food issue by using non-edible part of the plants, generate primary metabolites such as cellulose and starch for biofuel production, and provide secondary metabolites that can be used to produce valuable pharmaceuticals and flavors^{5, 6, 8}. However cost and complexity of the processing technology and pretreatment steps to release cellulosic material are considered as drawbacks of second generation biofuels⁶. Altogether, the first and second generation biofuels share the common problem of using a large amount of land space and fail to meet the high sustainable energy demand⁷.

1.2 Microalgae

Biofuel production has long been thought to have the potential to minimize or replace fossil fuel demand¹. However, because of the above-mentioned limitations of first

and second generation biofuels, ‘third generation’ biofuels have arisen as a more efficient alternative biofuel production platform⁹. Third generation biofuels are produced using lipid/oil feedstocks from photosynthetic microalgae^{10, 11}. Microalgae have been considered to be a future sustainable alternative energy source and have the potential to replace petroleum-based engine fuels^{7, 11}. Microalgae-based biofuel production has gained lots of attention because of a variety of reasons, such as the enormous carbon dioxide fixation potential of algae¹², the best oil production per acre compared to other oil producing plants^{11, 13, 14}, growing in many different climates where other plants cannot grow¹⁵, and not competing with food crops^{11, 16}. Microalgae have been grown either in open pounds¹⁷ or in photobioreactors¹⁸ and cultivated under four different conditions including, heterotrophic, photoautotrophic, photoheterotrophic, and mixotrophic cultivation, which can determine the type and content of lipids produced by different types of microalgae¹⁰. Though open pound systems are cheap and easy to scale up, there is a risk of contamination in addition to some other issues that reduce the efficiency^{11, 19}. On the other hand, photobioreactors, in particular tubular photobioreactors seem to be the most efficient way to produce biomass from microalgae for biofuel production purposes⁷.

1.2.1 Microalgae-based biofuels

Microalgae can provide feedstocks such as triglycerides (TAGs) lipids and hydrocarbons to produce various types of fuels including biodiesel, bioethanol, biohydrogen, gasoline, and kerosene^{7, 11, 12, 15}. Triglycerides (TAGs) lipids produced by microalgae are used in biodiesel production via the transesterification process and have

received lots of attention over the years^{9, 11}. Currently, most biodiesel production comes from plant and animal oils, but recently there has been a big effort to produce biodiesel from microalgae using the same biodiesel production platform^{7, 9, 11}. Biogas including biomethane, biohydrogen, and CO₂ can be generated from waste microalgal biomass by either biochemical (anaerobic) or thermochemical (gasification) processes^{9, 11}. A great advantage of biogas production from microalgae is that even biomass from wastewater treatment or blooms can be used in this process⁹. Moreover, petroleum equivalent hydrocarbons can be produced by various microalgae in large amounts, up to 86% of dry algal weight, and these hydrocarbons can be converted into engine fuels using existing petroleum refinery platforms^{15, 20}. Algae-based hydrocarbons will be discussed more in later sections.

1.2.2 Limitations of microalgae-based biofuels

Despite the great potential of microalgae like tremendous growth rate, large oil production, less use of terrestrial land, and no competition with food crops, there are still several challenges for large scale algae-based biofuel production and commercialization^{1, 21}. Some of the requirements for continuous microalgae-based biofuel production are supply of sunlight, CO₂, water, and minerals such as potassium (K), phosphorous (P), and nitrogen (N)^{12, 22}. Larger amounts of lipids are produced when extra CO₂ is fed to microalgae, but the issue here is that the extra CO₂ is purchased which corresponds to almost half of the biofuel production cost²¹. One way to reduce this cost is to inexpensively provide the extra CO₂ for microalgae, and great efforts have been made to transfer the CO₂

released by power-plants into microalgae growth platforms²¹. This approach might solve the cost issue of CO₂, but there has not been enough transportation of CO₂ from these plants to algae growth platforms to meet the demand for large scale biofuel production yet²¹.

Another challenge with microalgae-based biofuel production is the supply of P and N²¹. One common source of these minerals is fertilizers, but there are not sufficient supplies to fulfill the demand of large scale algae cultivation. There have been efforts made to produce more fertilizers, but this type of production requires the use of fossil fuels and causes release of large amounts of CO₂ into the atmosphere which is not desirable in terms of clean energy production^{1, 21, 23, 24}. On the other hand, there is also the challenge of oil extraction from algal biomass. There are several methods to extract oils from microalgae such as extraction with hexane or an expeller press, but these methods are expensive as they require a certain energy input²⁵. Another constraint for algal biofuel production is water supply, and it has been shown that fresh water supplies are not enough to meet the large scale biofuel production demand²¹. Though seawater is abundant and can overcome the issue of water supply, the evaporation occurring in growing cultures requires fresh water addition to compensate the evaporation²¹. Last, the current technologies used to grow and harvest algae, extract oils and provide efficient conversion of raw materials to large amounts of biofuels from microalgae are not at a desirable level^{1, 21}. On top of these, there is still the problem of overall oil production cost. Algal-based biofuels need to be economically feasible and produced sustainably so that algal biofuels can compete with fossil fuels, which are currently cheaper^{9, 21}.

1.2.3 Potential and future of microalgae for biofuel production

Despite all of the above-mentioned constraints, microalgae-based biofuels still outcompete all other sources of biofuel production and have the potential to meet the large scale sustainable energy demand if the challenges can be overcome^{7, 12, 21, 26}. For example, plant-based biofuels including first and second generation biofuels cannot meet the desired scale of biofuel production due to the mentioned reasons such as the food versus fuel debate and the low efficiency of plant-based biofuel production because of the limited land space and expensive pretreatment processes^{7, 11}. Thus, microalgae still stand as the most powerful biofuel production platform and have the potential to replace fossil fuels²². A powerful proposed approach to tackle the challenges with large scale microalgae biofuel production is genetic/metabolic engineering^{11, 21, 27, 28}. Generating microalgae strains that can produce high yields of oil and biomass, are easy to harvest, and have a high CO₂ fixation potential might be one of the best ways of making microalgae industrially efficient and a candidate to replace fossil fuels in the future. This can only be possible with genetic modifications or genetic/metabolic engineering by targeting specific pathways/genes in microalgae, in particular pathways related to photosynthesis, nutrient uptake, and metabolite production. Considerable effort has been made toward that purpose including transformation of algae and overexpression and knock-out of genes of interest^{21, 29-33}. Even though most of the genetic manipulation studies have been done in *Chlamydomonas reinhardtii*, successful transformation of several other green microalgae has been reported including *Chlorella vulgaris*, *Nannochloropsis* sp., *Volvox carteri*, *Dunaliella salina*, and *Chlorella sorokiniana*, and a red microalga *Porphyridium* sp. In addition, gene

manipulations in several of these microalgae has been reported^{34, 35}. Furthermore, the advancement in genome sequencing technologies in recent years led to the release of genomic sequences for several microalgae including *Ostreococcus tauri*, *C. reinhardtii*, *D. salina*, *Nannochloropsis* sp, and *Botryococcus braunii*³⁵⁻³⁹. These advancements in genome sequencing have brought more insights and opportunities for genetic engineering, and made it possible to target genes and pathways for generating microalgae strains that have improved traits in terms of biofuel production¹. The genetic modifications have been mostly focused on either photosynthetic efficiency or lipid biosynthesis in microalgae³⁵. Over the years, several genes have been targeted to increase photosynthetic efficiency in several microalgae species including *C. reinhardtii*, and great progress has been made toward that purpose^{21, 40, 41}. For example, photosystems and several enzymes in the Calvin cycle including Rubisco have been modified and engineered to have superior carbon fixation powers for overexpression in a desired microalgae to improve photosynthetic efficiency^{21, 41}. Similarly, several studies have shown that genetic modifications targeting light harvesting complexes can increase light capture/absorption and leads to improved photosynthesis in microalgae^{27, 41, 42}. In addition, other organisms like cyanobacteria have been shown to have mechanisms that allow for CO₂ delivery into the cells from the surrounding environment, and key steps of these mechanisms in cyanobacteria can be modified via genetic engineering to improve photosynthetic efficiency and biomass production^{21, 43}.

Another significant target of genetic/metabolic engineering in microalgae and cyanobacteria has been the pathways related to lipid biosynthesis since the products of

these pathways can be used for biofuel production^{11, 21, 22}. The oils produced by microalgae range from membrane lipids to TAGs and hydrocarbons^{15, 21}. The focus has shifted towards TAG production as the amount of TAGs produced can be as high as 60% of algal dry weight when algae are grown under desirable nitrogen, temperature, and salinity conditions^{21, 44}. Moreover, TAG biosynthesis pathways can be targeted by genetic/metabolic engineering to generate the desirable types of lipids^{1, 33}. For example, TAGs with shorter chain length fatty acids are more favorable when they are used in biodiesel production as they are easier to process and increase the quality of biodiesel⁴⁵.

On the other hand, the challenge with the supply of nitrogen and phosphorous can be overcome by genetically modifying microalgae to fix more N and by providing phosphorous from wastewater²¹.

More advancements in genetic/metabolic engineering tools along with genome sequencing technology may significantly contribute to make improvements in microalgae-based biofuel production and obtain the desired biofuel products at lower costs³³. Moreover, developing new photobioreactors at a low cost and high efficiency will be another significant aspect in sustainable biofuel production from microalgae¹¹. However, the key factor to make microalgae-based biofuel production sustainable will be identifying algal strains that have the combination of high growth rate, oil content and biomass production, and produce valuable byproducts such as pharmaceuticals at low costs^{1, 21}. Luckily, there have been a great number of microalgae strains that produce many types of oils and by sequencing more of the genomes of these strains many significant pathways

and key genes of interest can be revealed and targeted by genetic engineering and breeding strategies in order to increase biofuel production from microalgae¹.

As mentioned above, there are several products produced by microalgae for biofuel purposes and significant effort has been made on understanding and modifying TAG biosynthetic pathways³³. Biodiesel production from microalgae certainly is a great alternative to fossil fuels, but growing microalgae for solely large scale biodiesel production purposes might not seem possible because of economical unfeasibilities⁴⁴.

1.2.4 Microalgae-based hydrocarbons

Hydrocarbons produced by microalgae are another potential feedstock for biofuel production^{46, 47}. Hydrocarbons are favored over TAGs for several reasons. For example, they possess a higher energy density, they are more stable which gives them the advantage of long storage, and finally they are compatible with the existing petrochemical refinery and hydrocracking technology^{20, 46-48}. Moreover, algal hydrocarbons such as isoprenoids can exist in many different structures such as cyclic/acyclic, short or long chain alkanes/alkenes with various properties^{49, 50}. The pathways producing these molecules can be tailored by genetic engineering to generate valuable biofuels or precursors that can be used as engine fuels such as gasoline, kerosene, and diesel⁴⁸. For example, biofuel production can focus on a specific biofuel that can be used for a certain application such as jet fuels⁹. There have been a great amount of research focusing on producing biofuels or biofuel precursors in fast growing organisms such as *E.coli* and *Saccharomyces cerevisiae*⁵¹⁻⁵³. These organisms have been engineered to produce both gasoline precursors

which are short chain (C₅-C₁₂) hydrocarbons and long chain (C₁₀-C₂₃) hydrocarbons that are precursors of kerosene and diesel⁵². Though this approach is advantageous, only small scale studies have been conducted with these organisms⁵² and there exists a big economical constraint before such platforms can be used for large scale biofuel production as they depend on external carbon source for growth⁵². Therefore, photosynthetic microalgae might serve as a better platform for biofuel production as they can produce the needed carbon for growth by fixing CO₂. Microalgae strains that can produce high yields of isoprenoid-like hydrocarbons such as *B. braunii* and microalgae that produce low yield of hydrocarbons but suitable for genetic manipulations like *C. reinhardtii* and *Pseudochoricystis ellipsoidea* can be good alternatives for microalgae based biofuel production^{15, 33, 54, 55}.

Among many hydrocarbon producing green microalgae such as *Pseudochoricystis ellipsoidea*, *Dunaliella* spp, *Scenedesmus acutus*, *S. acuminatus*, *S. obliquus*, *Chlorella kessleri*, *C. vulgaris*, and *Chlorella* sp.^{47, 56}, *B. braunii* has gained lots of attention as it produces large amounts of liquid petroleum like hydrocarbons up to 86% of its dry weight^{15, 57}. Extensive details about *B. braunii* will be discussed later in this report.

1.3 *Botryococcus braunii*

B. braunii is a green colonial microalga that can grow in different conditions from brackish water to fresh water, in various climates such as alpine, temperate or tropical zones, and on all continents except Antarctica⁵⁷. Individual cells of *B. braunii* are held together by an extracellular matrix (ECM) which will be discussed later in detail (Figure

1)⁵⁸. Race A, race B, and race L are the three chemical races of *B. braunii* and are classified based on the hydrocarbons they produce⁵⁷. Race A produces hydrocarbons up to 61 % of its dry weight in the form of C₂₃-C₃₃ alkadienes and alkatrienes derived from fatty acids; race B, which is the focus of this study, produces the triterpene C₃₀-C₃₇ hydrocarbons called botryococcenes in large quantities and C₃₁-C₃₄ methylsqualenes in small quantities with a total production up to 86% of dry algal weight, but typically 30-50% of the dry weight; and race L produces the tetraterpene C₄₀ hydrocarbon known as lycopadiene as a major product up to 8 % of its dry weight (Figure 2)^{15, 57, 59}.

B. braunii has received great attention throughout the years and is considered as a potential organism for sustainable large-scale biofuel production for several reasons. First of all, *B. braunii* produces large amounts of liquid hydrocarbons that are mostly secreted into the ECM and are easily extractable^{59, 60}. Second, the hydrocarbons can easily be processed to obtain high quality engine fuels via basic physical processes such as hydrocracking and distillation with a 67 % gasoline, 15 % aviation turbine fuel, and 3 % residual oil production using the existing petroleum refineries²⁰. Last, all three races of *B. braunii* have been reported as contributors to the formation of existing oil shales and coal deposits, which points out the geological importance of *B. braunii*⁶¹⁻⁶⁵.

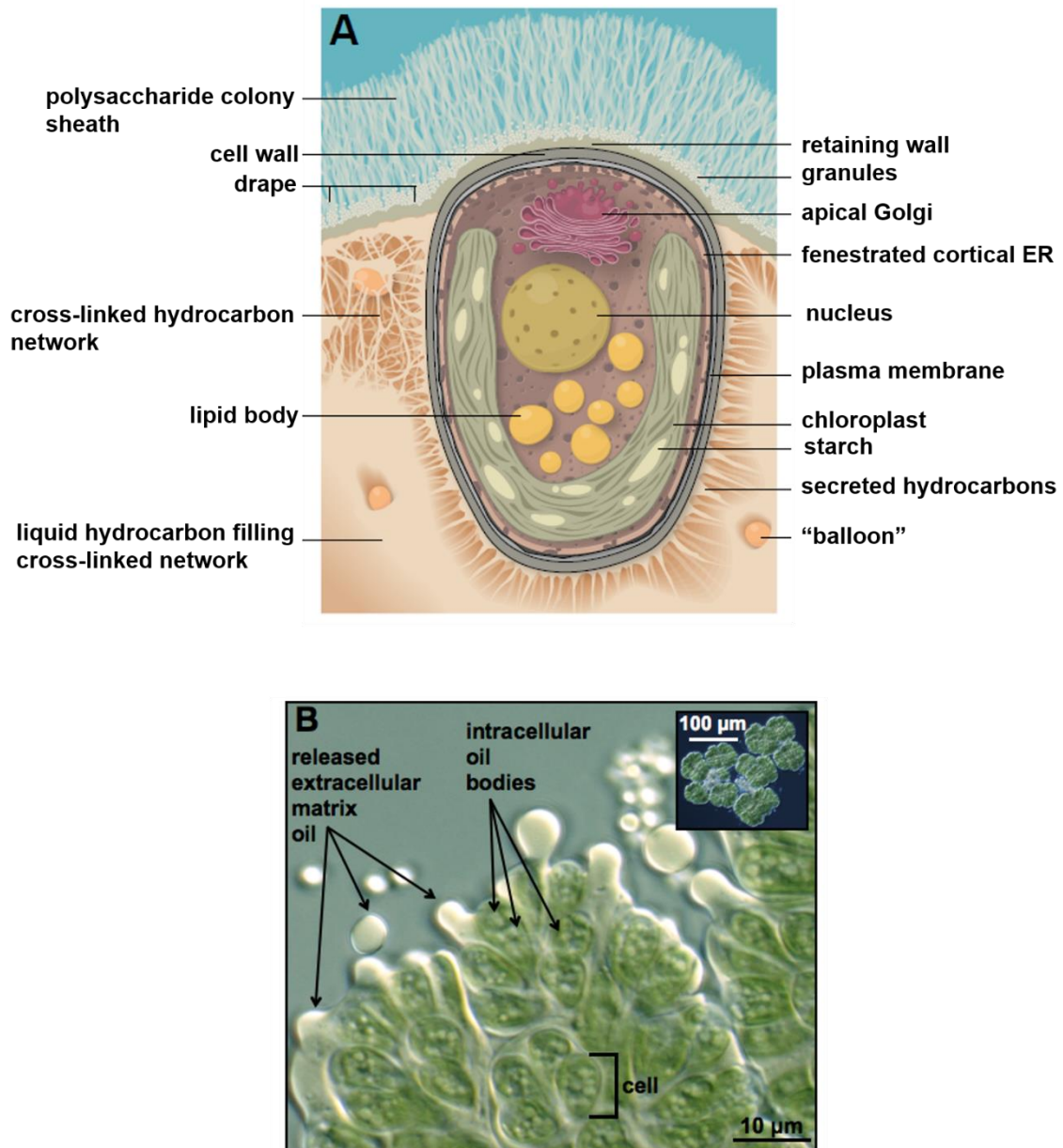


Figure 1 Model and image of *Botryococcus braunii*. (A) Model of *B. braunii* colony focused on a single cell (Figure adapted from Weiss TL et al., 2012 *Eukaryotic cell.* 11:1424-1440). (B) Transmitted microscopic image *B. braunii* colony when pressure was applied to release oils from extracellular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

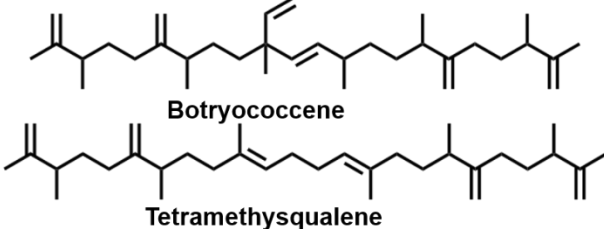
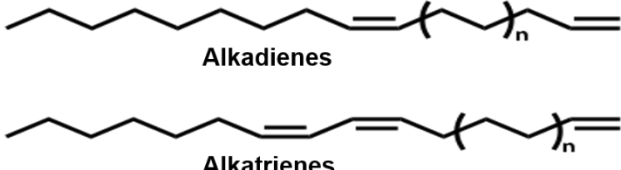
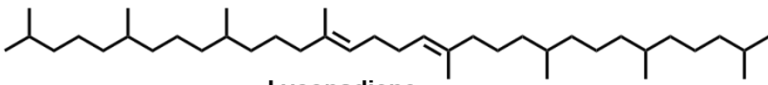
B Race:	 <p>Botryococcene</p> <p>Tetramethysqualene</p>
A Race:	 <p>Alkadienes</p> <p>Alkatrienes</p>
L Race:	 <p>Lycopadiene</p>

Figure 2 Types of hydrocarbons produced by each race of *Botryococcus braunii*.

However, there are still several challenges that need to be overcome before biofuel production from *B. braunii* becomes economically feasible. One of the major disadvantages is that *B. braunii* is a slow growing organism with a cell doubling time of 6 days⁶⁶. The slow doubling time might be attributed to spending most of the cellular energy to produce the large amount of hydrocarbons⁶⁶. This issue with *B. braunii* doubling time can be overcome via genetic modification strategies. Unfortunately, attempts to transform *B. braunii* in recent years have not been successful due to the complex ECM⁶⁷. Though studies have shown that single cells of *B. braunii* can be isolated which leaves cells ECM free and might allow for biolistic transformation, isolated single cells failed to survive for more than a week⁶⁷. Nonetheless, it certainly is promising that single cells of *B. braunii* can be isolated, and with the discovery of methods/chemicals that increases the survival rate of single cells, biolistic transformation of *B. braunii* might be possible in the near future. Moreover, key enzymes involved in hydrocarbon biosynthesis, the squalene synthase like enzymes SSL-1, SSL-2, and SSL-3 from race B and a lycopaoctane synthase enzyme from race L have been identified and characterized. Overexpression of these enzymes in various organisms including *E.coli*, *S. cerevisiae*, and tobacco to produce isoprenoid hydrocarbons have been successful⁶⁸⁻⁷¹.

1.3.1 Ultrastructure of *B. braunii*

B. braunii cells are clustered and held together via an extracellular matrix (ECM) which is a storage location for the hydrocarbons secreted from the cells⁶⁰. The ECM is formed upon cross-linking of hydrocarbons of *B. braunii* via ether bridges, and thus the

structure of the ECM is different for each race of *B. braunii*⁷²⁻⁷⁴. Approximately 95 % of the hydrocarbons produced by this alga are exported into the ECM while the remaining 5% are stored in the form of intracellular oil bodies (Figure 1A and B)⁶⁰. A recent histochemical and biochemical study about the ultrastructure of *B. braunii* has brought new insights into the colony organization and ECM structure of *B. braunii*⁵⁸. It has been shown that colonies of *B. braunii* are surrounded by a retaining wall that sequesters the hydrocarbons into the ECM⁵⁸. A polysaccharide fibril sheath emanates outward from the retaining wall and completely surrounds the colony⁵⁸. Another unique feature of *B. braunii* reported in this study is that *B. braunii* possesses a fenestrated endoplasmic reticulum, which surrounds the interior of the cell, and its fenestrae possibly allow for the transition of the hydrocarbons into the ECM (Figure 1A)⁵⁸. A detailed review of ECMs from several organisms and their comparison to ECM from *B. braunii* is discussed below in section 1.4.

1.4 ECMs in various organisms

1.4.1 ECMs in plants

The main component of the ECM in plants is the cell wall, which is mostly made of cellulose formed by β -1, 4 glucan chains, and is one of the most abundant polymers on earth⁷⁵. In addition, there are also several other components that form the plant cell wall including matrix polysaccharides, the pectic polysaccharides⁷⁶ and the hemicellulosic polysaccharides⁷⁷, many proteins and glycoproteins including various enzymes and structural proteins⁷⁸. One of the most abundant proteins in the cell wall of plants is extensin, which is a hydroxyproline rich protein essential for cell wall assembly involved

in cell extension⁷⁹. Another major component of the plant cell wall is lignin which is a polymer derived from aromatic alcohols and gives mechanical strength to plants by linking various ECM polysaccharides⁸⁰. Cellulose biosynthesis takes place at the plasma membrane in the form of rosettes while ECM polysaccharides and glycoproteins are synthesized in the Golgi where the products are kept in the lumen before they transit to the cell wall via vesicles⁷⁵. There are several functional roles for these ECM macromolecules such as providing a supporting framework, which helps hold the cells and tissues together⁸¹. Moreover, the cell wall prevents cells from swelling and bursting upon osmotic pressure changes, which constantly directs the flow of water into the cell by providing rigidity to the cell⁸². The plant cell wall and its components are very well characterized structurally, but how these components are assembled to form the ECM and its function still remains unknown⁷⁵.

1.4.2 ECMs in green microalgae

Evolutionary studies indicate that all species of the plant kingdom arose from one common ancestor, charophycean algae, and when plants were defined as photosynthetic eukaryotes, this included green, brown, and red microalgae⁸³. Thus, it was thought that these groups shared a common complex polysaccharide-rich cell wall⁸⁴. However, more recent studies reported that cell walls of these groups developed a distinct composition as they evolved⁸⁴. There is limited knowledge about why and how these cell walls evolved differently especially considering there are around 36,000 algae that show a great diversity of cell wall components⁸⁴. Green algae are surrounded by a complex ECM as well as their

land plant offspring, and by investigating and elucidating the cell wall components from different green algae species, more information towards understanding the evolution of the ECM in green algae as well as potential applications of cell wall components in food and pharmaceutical industries can be gained⁸⁵.

Besides the cell wall components, the ECM in green microalgae can consist of other material beyond the cell wall such as glycoproteins and hydrocarbons^{57, 58, 86, 87}, however, the most common component is glycoproteins that are often cross-linked with each other⁸⁷. One of the best studied green microalgal ECMs is that of *Volvox carteri*, which consists solely of glycoproteins, can hold over 2,000 cells differentiated mainly into somatic cells with a few reproductive gonidia cells, and is highly enriched in 4-hydroxyproline (Hyp) rich glycoproteins (HRGPs) that are cross-linked for stabilization⁸⁷⁻⁹⁰. The use of HRGPs is a common feature in the ECMs and cell walls of both algae and land plants, and they are commonly *O*-glycosylated at Ser, Thr, and Hyp residues with arabinose and galactose^{79, 91-93}. Moreover, green microalgae are known to release polysaccharide material which is a part of their ECM into their media as they grow⁹⁴. This material is called exopolysaccharides (EPSs) and is produced by different organisms including bacteria, fungi, marine microalgae, and photosynthetic plants⁹⁵. EPSs produced by these organisms differ in terms of structure and function, which is a great advantage as new functions and applications of EPSs can be explored. It has been reported that EPSs in bacteria play a role in pathogenesis, symbiosis, biofilm formation, and protection from predators⁹⁶. There are several other applications of the exopolysaccharides from bacteria including antitumor and antiviral activities, and anticoagulant features^{97, 98}. On the other

hand, several applications of EPSs in cyanobacteria have been reported including detoxification of heavy metals and radionuclides from contaminated water, and water purification by filtering out solid matter⁹⁹. Similarly, algal exopolysaccharides were found to differ in structural and functional properties which have been used for many industrial purposes including pharmaceutical, cosmetic, and food industries⁹⁴. All these interesting properties of exopolysaccharides have driven researchers to investigate new exopolysaccharides from green algae that can be used towards novel industrial and biotechnological applications. Moreover, obtaining these kinds of valuable byproducts from microalgae grown for biofuel purposes could help in economics of biofuel production from microalgae.

1.4.3 ECMs and EPSs in *B. braunii* and contrast to other organisms

As mentioned above, the individual cells of *B. braunii* within the colonies are held together through a polymer network ECM, which is formed upon cross-linking of hydrocarbons of *B. braunii* via ether bridges and serves as a storage location for hydrocarbons that are secreted by the individual cells^{58, 60}. The ECM biopolymer is composed of long chain polyacetal hydrocarbons and is present in all the three races of *B. braunii*, with slight differences in each race. In race A, the polyacetals are cross linked to each other to form the polymer⁷²; in race B, tetramethylsqualene is cross linked to the polyacetals and forms the polymer⁷³; and in race L the biopolymer is formed upon cross linking of lycopadiene to the polyacetals¹⁰⁰. The main function of these biopolymers is to provide a storage location for the hydrocarbons secreted by the *B. braunii* cells¹⁵. In

addition, these biopolymers possibly are the reason for coorongite formation¹⁵. Bearing these polymer networks gives *B. braunii* a striking difference compared to other microalgae such as *Volvox carteri*.

It has also been shown that colonies of *B. braunii* are surrounded by a retaining wall that sequester the hydrocarbons into the polymer network between the cells and also aid in the formation of the *B. braunii* colonies⁵⁸. Along with the retaining wall, there exist many granules that appear to be proteinaceous and a polysaccharide fibril sheath that extends toward the media and surrounds the entire colony (Figure 1A)⁵⁸. In addition, the cell wall of *B. braunii* is composed of cellulosic polysaccharides while the retaining wall and fibrils are not⁵⁸. When *B. braunii* cells divide, the apical parts of individual cells in the colony are shed into the media and accumulate in the form of “shells.” Transmission electron microscopy analysis of the shells indicate that the shells consist of the retaining wall, proteinaceous granules, and the polysaccharide fibrils⁵⁸. Periodic acid-Schiff reagent (PAS) staining and GC/MS analysis of the shells revealed that the major constituent of the shells are carbohydrates (97.9 %), and a Bradford assay indicated that 1.25 % of the shells was found to be protein⁵⁸. Therefore, this indicates that *B. braunii* also releases exopolysaccharides or EPSs like other microalgae.

Carbohydrate analysis has shown that the dominant sugars in the EPSs of *B. braunii* are mostly arabinose and galactose and some deoxyhexoses⁵⁸. The linkage analysis has shown that *B. braunii* EPSs contains 2-3 and 2-4 glycosidic linkages in general, which indicates that the structure of the *B. braunii* EPSs are distinct from the EPSs of other organisms since they do not contain any of the very common 1,4 or 1,6

linkages⁵⁸. Unlike the commonly known structural carbohydrates, EPSs from *B. braunii* do not contain uronic acids or *N*-acetyl glycosyl residues and has unusual structures⁵⁸. It is suggested that the retaining wall of *B. braunii* is formed by deoxyhexoses since they are common in bacteria while arabinose and galactose form the uniquely structured fibrils^{58, 101}.

Several studies reported that production of EPSs in *B. braunii* serves various functions including protecting the cells so that *B. braunii* can resist environmental changes such as salinity, temperature, light intensity, pH, and dessication^{102, 103}. Some of the other possible functions of the *B. braunii* fibrils include (1) to play a mediator role between the cells and growth environment; (2) involvement in symbiotic relationships with the bacteria in the environment; and (3) to likely act as a protecting barrier against pathogens⁵⁸. Section IV of this study will discuss the characterization of a polysaccharide associated ECM protein, PSAP, which has been found to be the only protein localized in the shell material released by *B. braunii* race B.

1.5 Isoprenoid biosynthesis

Isoprenoids (a.k.a. terpenoids) are one of the biggest and most structurally diverse natural product families present in all life forms with production of more than 55,000 different terpenoids, particularly in plants¹⁰⁴. Isoprenoids can be used as hormones, pigments, and electron carriers; can possess antimicrobial, anti-inflammatory and anticancer properties and also play a role in signaling, defense mechanism, photosynthesis, and many other cellular activities¹⁰⁴⁻¹⁰⁶. Besides the great functional diversity of the

isoprenoid family, another significant feature of this family is their structural diversity. Some terpenoids can either directly be fuel like molecules or valuable biofuel precursors⁴⁸⁻⁵⁰. Isopentenyl diphosphate (IPP) and its isomer dimethylallyl diphosphate (DMAPP) are the common precursors used to produce these isoprenoids (Figure 3)^{106, 107}. Combining IPP and DMAPP will produce the C₁₀ intermediate geranyl diphosphate (GPP). Successive additions of IPP can form C₁₅ farnesyl diphosphate (FPP), C₂₀ GGPP, and additional molecules up to hundreds of carbons long¹⁰⁷. Use of these intermediates by several enzymes for reactions such as cyclization, condensation, rearrangements, and oxidation leads to the formation of different isoprenoid families, such as monoterpenes (C₁₀), sesquiterpenes (C₁₅), diterpenes (C₂₀), triterpenes (C₃₀), and tetraterpenes (C₄₀)^{108, 109}.

There are two reported pathways for the biosynthesis of IPP^{49, 110}. The first is the mevalonate or MVA pathway which is found in all mammals, fungi, plants, archaea, and some gram-positive bacteria, which is localized in the cytosol^{106, 111}. The other pathway is the methylerythritol phosphate (MEP) pathway which was later characterized and found in the majority of prokaryotes and the plastid of green algae and higher plants^{49, 106, 107}. The MEP pathway starts with the reaction of glyceraldehyde-3-phosphate and pyruvate catalyzed by the enzyme 1-deoxy-D-xylulose-5-phosphate (DXS) which produces an intermediate called 1-deoxy-D-xylulose-5-phosphate (DXP) (Figure 3)¹⁰⁷. This step has been proposed as rate limiting step of the MEP pathway, and is followed by the conversion of DXP to MEP catalyzed by DXP reductoisomerase or DXR (Figure 3)¹⁰⁷.

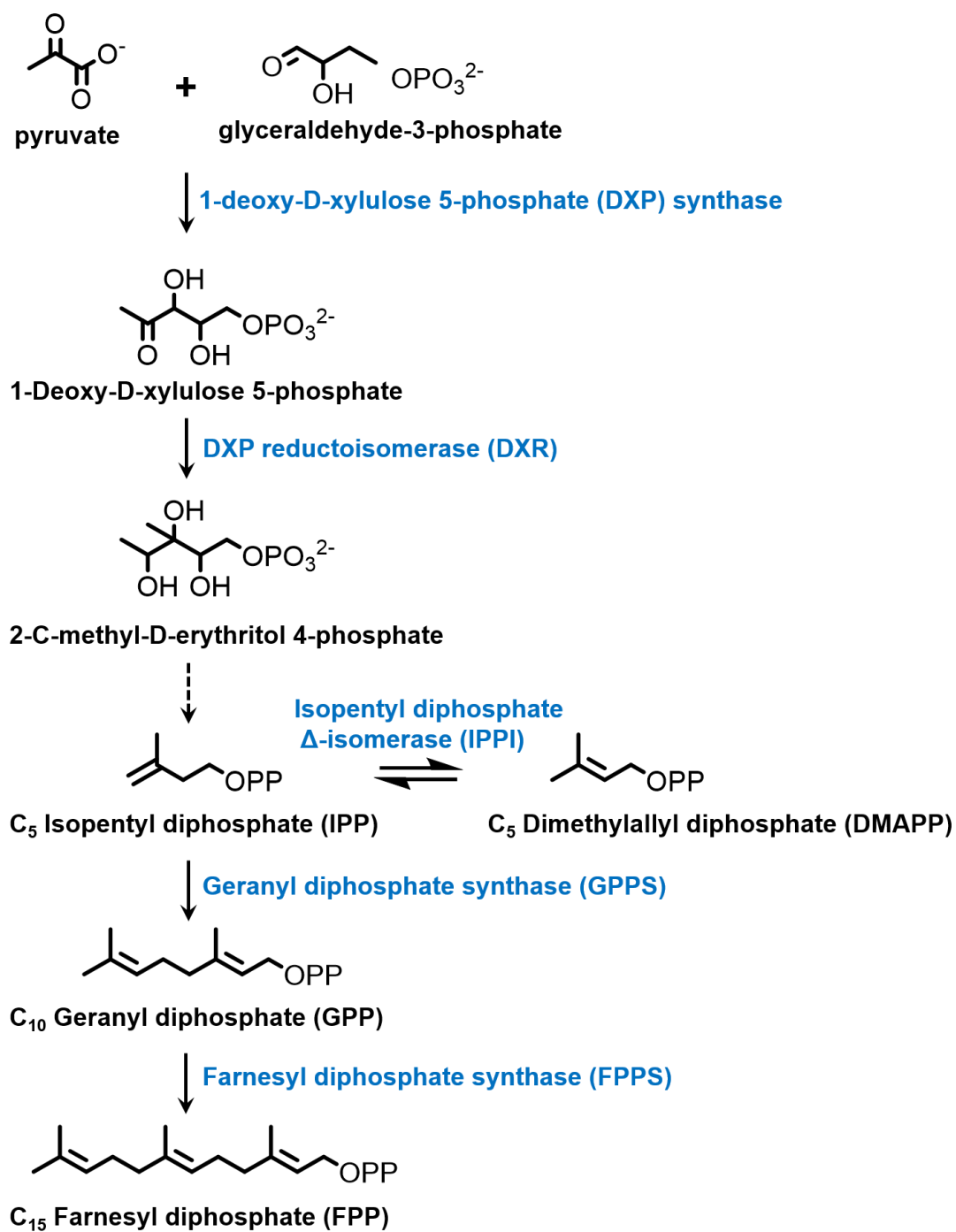


Figure 3 Production of various isoprenoids via MEP pathway.

After several subsequent enzymatic reactions, IPP and DMAPP are produced (Figure 3)¹⁰⁷. These molecules are then used to produce higher number isoprenoids including C₁₀ geranyl diphosphate (GPP) and C₁₅ farnesyl diphosphate (FPP) in the later steps of the MEP pathway (Figure 3)¹⁰⁷.

One of the other products produced from IPP and DMAPP is the C₃₀ triterpene squalene, an intermediate product in sterol metabolism, and common to humans, yeast, and plants¹¹². Squalene also serves as a precursor in the biosynthesis of many other cyclic triterpenes with a variety of structures¹¹³. The biosynthesis of C₃₀ squalene is catalyzed by an enzyme called squalene synthase or SS¹¹². The catalysis mechanism of SS follows a two-step reaction; the first step involves condensation of two molecules of C₁₅ farnesyl diphosphate (FPP) to form the stable cyclopropyl intermediate presqualene diphosphate (PSPP), and in the second step PSPP undergoes reductive arrangement to form a 1-1' linkage between two FPP molecules to yield C₃₀ squalene (Figure 4)^{114, 115}.

1.6 Hydrocarbon biosynthesis in *B.braunii*

B. braunii produces hydrocarbons that differ between the A, B, and L chemical races⁵⁷. While the B and L races produce isoprenoid hydrocarbons, botryococcenes and methyl squalenes in race B and lycopadiene by race L, race A produces the non-isoprenoid hydrocarbons alkadienes and alkatrienes, which are derived from fatty acids⁵⁷.

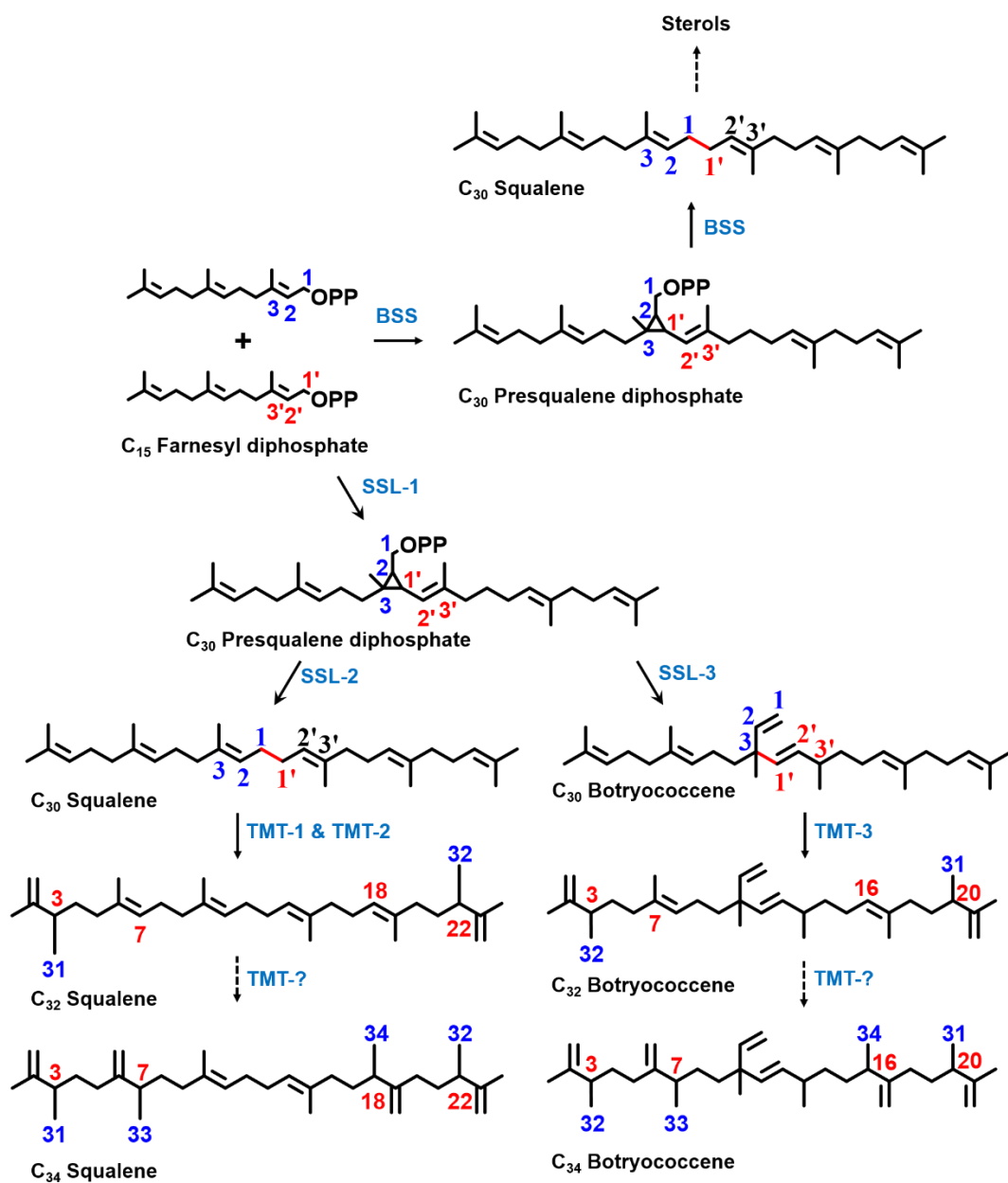


Figure 4 Biosynthesis of *B. braunii* race B hydrocarbons. In C₃₂ - C₃₄ squalenes and C₃₂ - C₃₄ botryococcenes red carbon numbers indicate methylation steps and blue carbon numbers indicate added methyl group.

1.6.1 Hydrocarbon biosynthesis in race A

In race A, alkadienes and alkatrienes are derived by initially elongating fatty acids, generally oleic acid (18:1 *cis*- Δ^9), to form very long chain fatty acids (VLCFA)^{116, 117}. The VLCFAs are then converted to alkadienes and alkatrienes from one of two proposed pathways; one pathway is the conversion of the VLCFA to an alkene through a decarboxylation reaction^{116, 117} and the other one is the conversion of the VLCFA to an aldehyde via reduction and then to an alkene via decarbonylation (Figure 5)¹¹⁸⁻¹²⁰. Currently, none of the enzymes involved in either pathway have been identified and characterized for race A.

1.6.2 Hydrocarbon biosynthesis in race L

In race L, the complete biosynthetic pathway of lycopadiene synthesis has been proposed (Figure 6)^{71, 121}. Based on this proposed pathway, a squalene synthase like enzyme called lycopaoctaene synthase (LOS) has been found to catalyze the condensation of two GGPP molecules to C₄₀ lycopaoctaene in a similar mechanism of squalene synthase that produces squalene^{71, 112}. This would be followed by subsequent reductions of lycopaoctane by yet to be known reductase(s) to produce lycopahexaene, lycopapentaene, lycopatetraene, lycopatriene, and finally the major product lycopadiene (Figure 6)⁷¹.

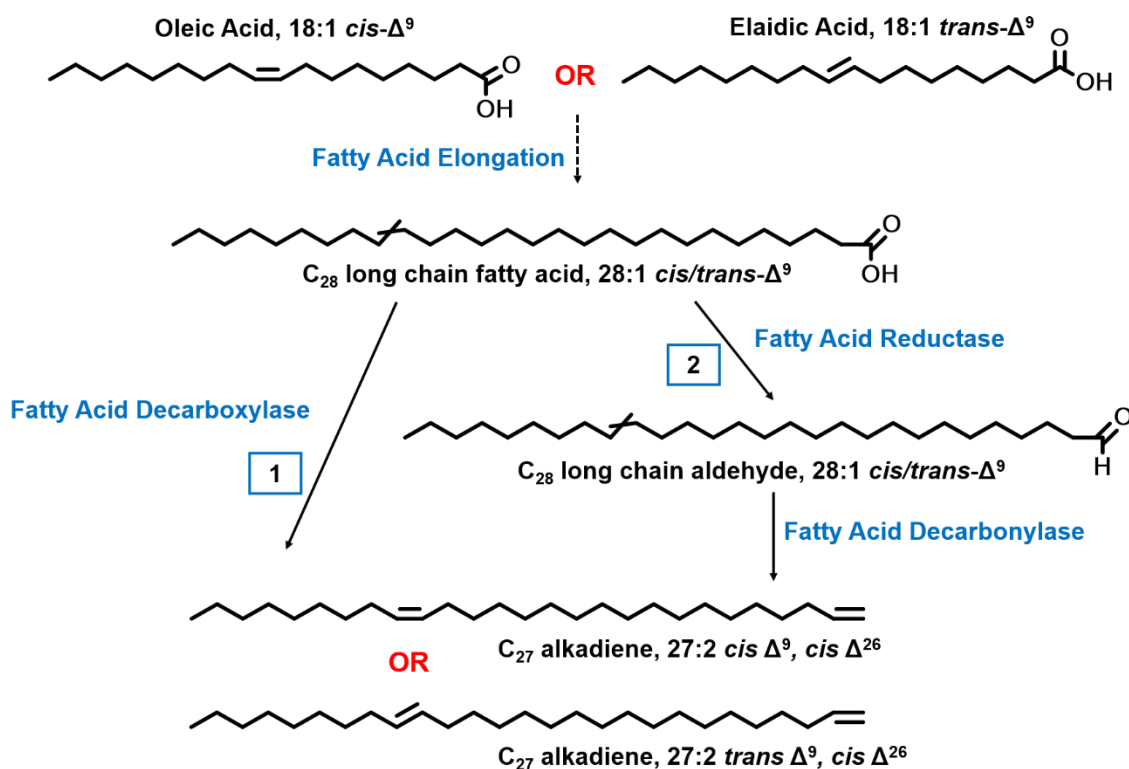


Figure 5 Two possible routes to C_{27} alkadiene production in race A of *B. braunii*. Elongation of either oleic acid or elaidic acid to generate C_{28} fatty acid. In the first route, the elongation product, C_{28} fatty acid, is decarboxylated by fatty acyl decarboxylase to produce C_{27} alkadiene. In the second route, C_{28} fatty acid, is first reduced by fatty acyl reductase to form aldehyde followed by decarbonylation by fatty acid decarbonylase to produce C_{27} alkadiene.

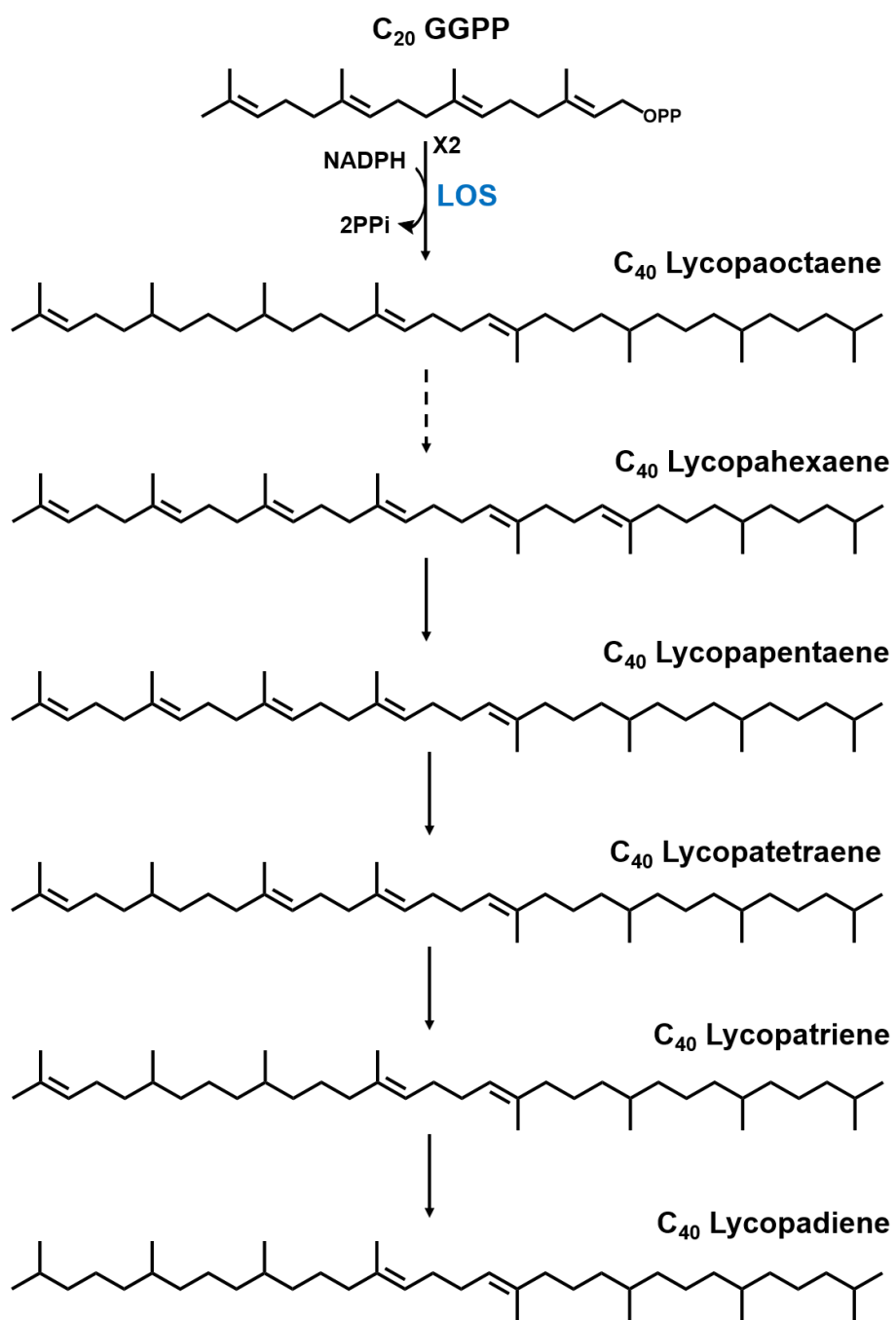


Figure 6 Proposed pathway for lycopadiene biosynthesis in *B. braunii* race L.

The pathway begins with the condensation of two GGPP molecules to produce lycopaoctaene via LOS enzyme followed by subsequent reductions of lycopaoctaene catalyzed by a yet unknown reductase(s) to produce lycopahexaene, lycopapentaene, lycopatetraene, lycopatriene, and lycopadiene.

1.6.3 Hydrocarbon biosynthesis in race B

Race B, the focus of this study, produces the triterpenes known as botryococcenes, mostly ranging from C₃₀ botryococcene to C₃₄ botryococcene, with C₃₄ as the major accumulating product. The B race also produces methylsqualenes, ranging from C₃₁ monomethylsqualene to C₃₄ tetramethylsqualene as minor components⁵⁷. Botryococcene production can be up to 86% of algal dry weight, but is typically in the range of 30-50%⁵⁷. The botryococcenes and methylsqualenes in race B are produced via the MEP pathway^{57, 122}. Because of the structural similarities between C₃₀ squalene and C₃₀ botryococcene (Figure 4), was proposed in the past that biosynthesis of C₃₀ botryococcene could take place by a reaction similar to that of squalene; formation of PSPP in the first step but alternative rearrangement of the cyclopropyl ring of PSPP to create a 1-3' linkage between the two FPP molecules (Figure 4)¹¹⁵. Searching the race B transcriptome for the squalene synthase like genes revealed a botryococcene squalene synthase (BSS) and three squalene synthase like genes; squalene synthase like-1 (SSL-1), squalene synthase like-2 (SSL-2), and squalene synthase like-3 (SSL-3)^{68, 114, 123}. It has been experimentally shown that BSS is a classical squalene synthase¹¹⁵. As BSS is the classical squalene synthase, it is proposed that squalene produced by BSS in race B plays a role in housekeeping functions such as sterol metabolism^{68, 114, 123}. On the other hand, an alternative mechanism to produce C₃₀ botryococcene as well as C₃₀ squalene in race B has also been found to follow a two-step reaction, however, in this case the two reactions are carried out by two separate enzymes⁶⁸. In the first reaction, two FPP molecules are condensed by the SSL-1 enzyme to form PSPP, and in the second step, PSPP is converted to C₃₀ botryococcene via the SSL-3

enzyme or to C₃₀ squalene by SSL-2 enzyme (Figure 4)^{68, 114, 123}. Squalene produced by SSL-1 and SSL-2 has been proposed to be involved in the biosynthesis of the ECM polymer network^{68, 69}. Moreover, production of the higher methylated hydrocarbons, C₃₁ to C₃₄ botryococcenes as well as C₃₁ monomethylsqualene to C₃₄ tetramethylsqualene is carried out by successive methylations of C₃₀ botryococcene and C₃₀ squalene, respectively, via triterpene methyl transferases (TMTs)⁶⁹. Two TMTs, TMT-1 and TMT-2, methylate C₃₀ squalene at C-3 and/or C-22 to generate C₃₁ monomethylsqualene and/or C₃₂ dimethylsqualene, while a single TMT, TMT-3, catalyzes two successive methylations of C₃₀ botryococcene at C-3 and C-20 to produce C₃₁ botryococcene and C₃₂ botryococcene (Figure 4)⁶⁹. The TMTs responsible for methylation of C₃₂ botryococcene at C-7 and C-16 to produce C₃₃ and C₃₄ botryococcene and methylations of C₃₂ dimethylsqualene at C-7 and C-18 to generate C₃₃ trimethylsqualene and C₃₄ tetramethylsqualene have yet to be identified (Figure 4).

1.6.3.1 Botryococcene isomers

There are several main acyclic botryococcenes produced by race B: C₃₀, C₃₁, C₃₂, C₃₃, and C₃₄ botryococcene (Figure 7). It has been reported that there are approximately twenty different isomers based on the C₃₁-C₃₄ structures (Figure 7). For example, there are four isomers based on the main C₃₁ molecule; two acyclic isomers, showacene **2** with a methyl addition on C-20 and isoshowacene **2-1**, which is the acyclic isomer of **2** and has the added methyl on C-3¹²⁴, two cyclic isomers, wolficene **2-2** and isowolficene **2-3** which

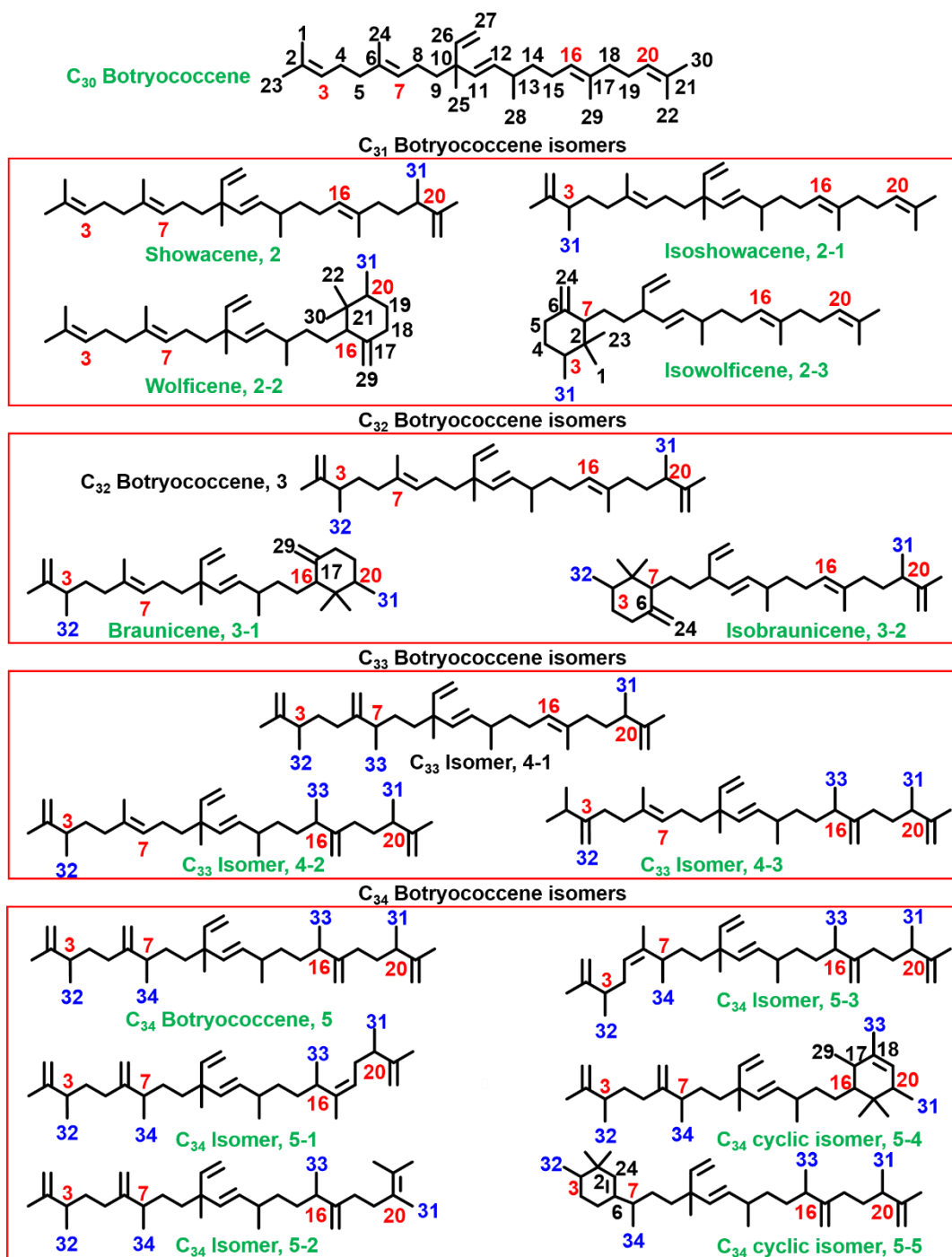


Figure 7 Botryococcene isomers from the B race of *B. braunii*. Isomers of C₃₀, C₃₁, C₃₂, C₃₃, and C₃₄ botryococcenes with methylation site carbon number in red and the methyl carbon number in blue.

both have a methylenecyclohexane group in opposite ends of the molecule (Figure 7)¹²⁵. The C₃₂ molecule **3** has two additional cyclic isomers; braunicene **3-1** has a methylenecyclohexane moiety linked to C-16 and isobraunicene **3-2** has the same ring structure on the opposite end of the molecule attached to C-7 (Figure 7)^{125, 126}. Three acyclic isomers of C₃₃ botryococcene have been reported; isomer **4-1** with a methyl addition at C-7, isomer **4-2** with a methyl at C-16, and isomer **4-3** with a methyl addition at C-16 and exomethylene group at C-32^{127, 128}. There are five isomers of the C₃₄ molecule **5**; isomers **5-1**, **5-2**, and **5-3** are the acyclic isomers which have several rotations on either end of the molecule mainly around C-20, C-17, and C-6, respectively, and isomers **5-4** and **5-5** are the cyclic C₃₄ isomers that contain a cyclohexyl moiety either attached to C-16 (molecule **5-4**) or C-6 (molecule **5-5**) at the ends of molecule 4 (Figure 7)^{127, 129}. Moreover, in a previous study, methylated squalenes (C₃₁, C₃₂, C₃₃, and C₃₄) from a Bolivian strain of *B. braunii* were purified and characterized¹³⁰. In addition to these structurally identified botryococcene isomers, there might be several other new botryococcene and squalene isomers that are produced by race B of *B. braunii*. Several new botryococcene and squalene isomers of race B that are isolated and characterized in this study will be discussed in section II.

1.7 Raman Spectroscopy

Raman spectroscopy is a technique that provides vibrational, rotational, and low frequency signals when laser light interacts with a molecule and generates structural fingerprints to identify the analyzed molecules¹³¹. Upon interaction of a photon with the

molecule, the molecule becomes excited and moves from its ground state to a virtual energy state¹³¹. If the frequency of the incident light is equal to the scattered light, this leads to an elastic scattering called Rayleigh scattering, and this does not carry any chemical information of the analyzed molecule¹³¹. However, if the frequency of the scattered light is different than the incident light, it causes an inelastic scattering by removing or adding energy from the molecule leading to Stokes or anti-Stokes scattering, respectively, and this inelastic scattering is also known as the Raman effect¹³¹. This phenomena was first discovered and introduced to public attention by an Indian physicist, Sir Chandrasekhara Venkata Raman, who received the 1930 Nobel Prize in Physics for this discovery¹³².

Raman spectroscopy can be applied to many disciplines including chemistry, physics, biology, biochemistry, and material science just to name a few. This versatility makes Raman spectroscopy a very powerful and indispensable method for scientists. However, there are also several issues that come along with this method. One challenge is the inefficient conversion of photons to Raman scattering, which is approximately 1 in 10^7 of the scattered photons while this conversion is about 1 in 10^4 for Rayleigh scattering¹³³. This inefficiency in Raman scattering leads to weak signaling and difficulties in interpreting spectra particularly when sample concentration is low and impurities are high in analyzed samples¹³¹. The other challenge of this method is the clear assignment of Raman vibrations to specific bonds within a molecule after the analysis¹³⁴. The first challenge has been overcome by the improvement in lasers and the techniques in Raman methodology for the enhancement of Raman signals^{134, 135}. Depending on the application,

there are several Raman techniques available today including surface-enhanced Raman scattering (SERS), resonance Raman spectroscopy, stimulated Raman spectroscopy (SRS), coherent anti-Stokes Raman spectroscopy (CARS), and many more. The second challenge has been overcome by the use of a quantum mechanics based method, called density functional theory (DFT), which calculates the electron density mapping of analyzed molecules and provides the theoretical spectra and assignment of Raman vibrations to molecule bonds¹³⁶. This method has made the data interpretation less complex and less prone to error.

Though Raman spectroscopy was mostly used in chemistry and physics for many years, there have been more biological applications such as *in vivo* applications in the last decade with the great improvements in lasers and techniques^{134, 135}. The features that make Raman spectroscopy a great choice for biological applications include, but are not limited to (1) water is not a Raman active molecule, thus water does not interfere with the spectra of molecules analyzed *in vivo* contrary to infrared spectroscopy; (2) there is no special requirement for sample preparation; (3) Raman spectroscopy is nondestructive and thus noninvasive, thus samples can be recovered after analysis; and (4) it is a label free method; (5) and Raman spectroscopy gives detailed molecular information¹³⁵. Therefore, the improvements in the Raman field are very critical for the use of Raman spectroscopy in life sciences. There has been a vast amount of use of Raman spectroscopy in the medical field, particularly in cancer research in the recent years. Raman spectroscopy provides detailed spectral features which are characteristic for specific molecules, and by targeting those molecules one can differentiate cancerous and non-cancerous cells¹³⁷, breast

tissues¹³⁸, and lung tissues¹³⁹. In addition, *in vivo* Raman spectroscopy has been incorporated into various other biological fields including molecular level analysis of imaging plant cell walls¹⁴⁰, lipid profiling in microalgae¹⁴¹⁻¹⁴⁴, and many others.

1.7.1 Raman spectroscopy studies in microalgae

Another growing use of Raman spectroscopy has been in the field of microalgae molecule analysis, which have received lots of attention due to the feedstocks they produce including proteins, lipids, carbohydrates¹⁴⁵. Each of these algal biomolecules can be identified by Raman spectroscopy as they have their own characteristic Raman signatures that can differentiate them from other molecules¹⁴⁵. In particular, lipids from microalgae have been the focus because of their potential in biofuel production, and thus Raman spectroscopy has been extensively used over the last decade to understand the types, molecular structures, and cellular localization of lipids¹⁴⁵. CARS and SRS have been preferred and considered as better Raman techniques to study lipid molecules in microalgae as they are more efficient, sensitive, and faster compared to conventional Raman spectroscopy¹⁴⁶. Some of the recent studies of lipids in microalgae include but are not limited to exploring; the TAG content of oleaginous the microalga *Monoraphidium neglectum* using CARS microscopy¹⁴⁷, the production of TAG in microalgae *Nannochloropsis oceanica* under the presence or absence of nitrogen at the single cell level¹⁴⁸, the determination of the degree of unsaturation in various microalgae TAGs including *B. braunii*, *N. oleoabundans*, and *C. reinhardtii* using laser-trapping Raman

spectroscopy focused on a single cell¹⁴¹, and measuring the total fatty acid content in *Chlorella vulgaris* using near-infrared Raman spectroscopy (NIR)¹⁴⁹.

1.7.2 Raman spectroscopy studies in *B. braunii*

Similar to other lipid producing microalgae, Raman spectroscopy analysis in *B. braunii* has also been growing in the recent years due to the potential of *B. braunii* for biofuel production. As mentioned above *B. braunii* produces different types of hydrocarbons including botryococcenes and methylated squalenes by race B, alkadienes and alkatrienes by race A, and lycopadiene by race L⁵⁷. Raman spectroscopy analysis of some of these hydrocarbons has been done over the years. One of the first Raman analyses was done to determine the distribution of the race A fatty acid derived hydrocarbons inside and outside of the cell⁶⁰. However, no significant spectral differences were obtained between the intra- and extra-cellular hydrocarbons via Raman spectroscopy while chemical analysis had already shown that intracellular oil was rich in C₂₇ and C₂₉ alkadienes and extracellular oil was rich in C₂₉ alkadiene and C₃₁ alkatriene⁶⁰.

Moreover, a recent DFT calculations and Raman spectroscopy analyses for six race L hydrocarbons, lycopaoctaene, lycopahexaene, lycopapentaene, lycopatetraene, lycopatriene, and lycopadiene, have shown that two spectral regions corresponding to the $\nu(\text{C}=\text{C})$ stretch between 1600 and 1700 cm⁻¹ and the CH₂/CH₃ bending region between 1250 and 1400 cm⁻¹ can be used to distinguish these molecules from each other¹⁵⁰.

In the recent years, several Raman spectroscopy studies were carried out on the race B botryococcenes to find the localization of some of the botryococcenes in *B. braunii*

colonies using their Raman signatures^{143, 151, 152}. The first attempt to identify specific Raman signatures (wavenumbers, cm^{-1}) for botryococcenes revealed the $\nu(\text{C}=\text{C})$ stretch vibration region ($1600\text{-}1700\text{ cm}^{-1}$) to be characteristic to botryococcenes and a further analysis for the purified individual botryococcenes C_{30} , C_{32} , C_{33} , and C_{34} (Figure 7) by Raman spectroscopy showed several specific Raman signatures, which was supported by DFT calculations, for these molecules in the $\nu(\text{C}=\text{C})$ region¹⁵¹. *In vivo* mapping attempts to localize botryococcenes proved the presence of these botryococcenes in the ECM and intracellular oil bodies, but the specific botryococcene makeup of the oil bodies was not determined as almost all botryococcenes possess the same specific signatures identified in the $\nu(\text{C}=\text{C})$ region^{144, 151, 152}. Thus, a more detailed analysis will require exploring specific Raman signatures for each purified botryococcene molecule outside the $\nu(\text{C}=\text{C})$ region. As mentioned above, race B is known to produce around 20 different isomers based on C_{31} , C_{32} , and C_{33} , and identification of a specific Raman signature for each botryococcene isomer is essential in order to map these botryococcenes in *B.braunii* cells. In section III of this study, the spectral differences of previously characterized botryococcene isomers in addition to several new C_{33} cyclic botryococcene isomers and methylated squalenes by *in vitro* Raman spectroscopy techniques will be discussed in detail.

2. ISOLATION AND CHARACTERIZATION OF CYCLIC C₃₃ BOTRYOCOCCENES AND A TRIMETHYLSQUALENE ISOMER FROM *BOTRYOCOCCUS BRAUNII* RACE B*

2.1 Background and rationale

Botryococcus braunii is a green colonial microalga that has received great attention from scientists for several reasons. First, it produces large amounts of liquid hydrocarbons which can be transformed into high quality combustion engine fuels via hydrocracking and distillation with a fuel distribution of 67% gasoline, 15% aviation turbine fuel, and 3% residual oil²⁰. Second, this species has been reported to be a major contributor to the formation of existing oil shale and coal deposits^{61, 62}.

B. braunii race B, the focus of this study, produces the triterpene isoprenoids known as botryococcenes, and production is typically in the range of 30-50% of algal dry weight, but it can also be up to 86% of dry weight^{57, 153}. The B race of *B. braunii* has been shown to make many isomers of botryococcenes, mostly ranging from C₃₀ botryococcene (**1**) to C₃₄ botryococcene (**2**), and squalenes ranging from C₃₀ squalene (**3**) to C₃₄ tetramethylsqualene (**3**) (Figure 8). For example, once produced **1** is successively methylated at positions C-3, C-7, C-16, and C-20 to generate C₃₁, C₃₂, C₃₃, and C₃₄

* Portions of this section have been reprinted with permission from: (1) Tatli, M., Naik, M. T., Okada, S., Dangott, L. J., and Devarenne, T. P. (2017) Isolation and characterization of cyclic C₃₃ botryococcenes and a trimethylsqualene isomer from *Botryococcus braunii* Race B. *Journal of natural products* 80: 953–958. doi: 10.1021/acs.jnatprod.1026b00934.

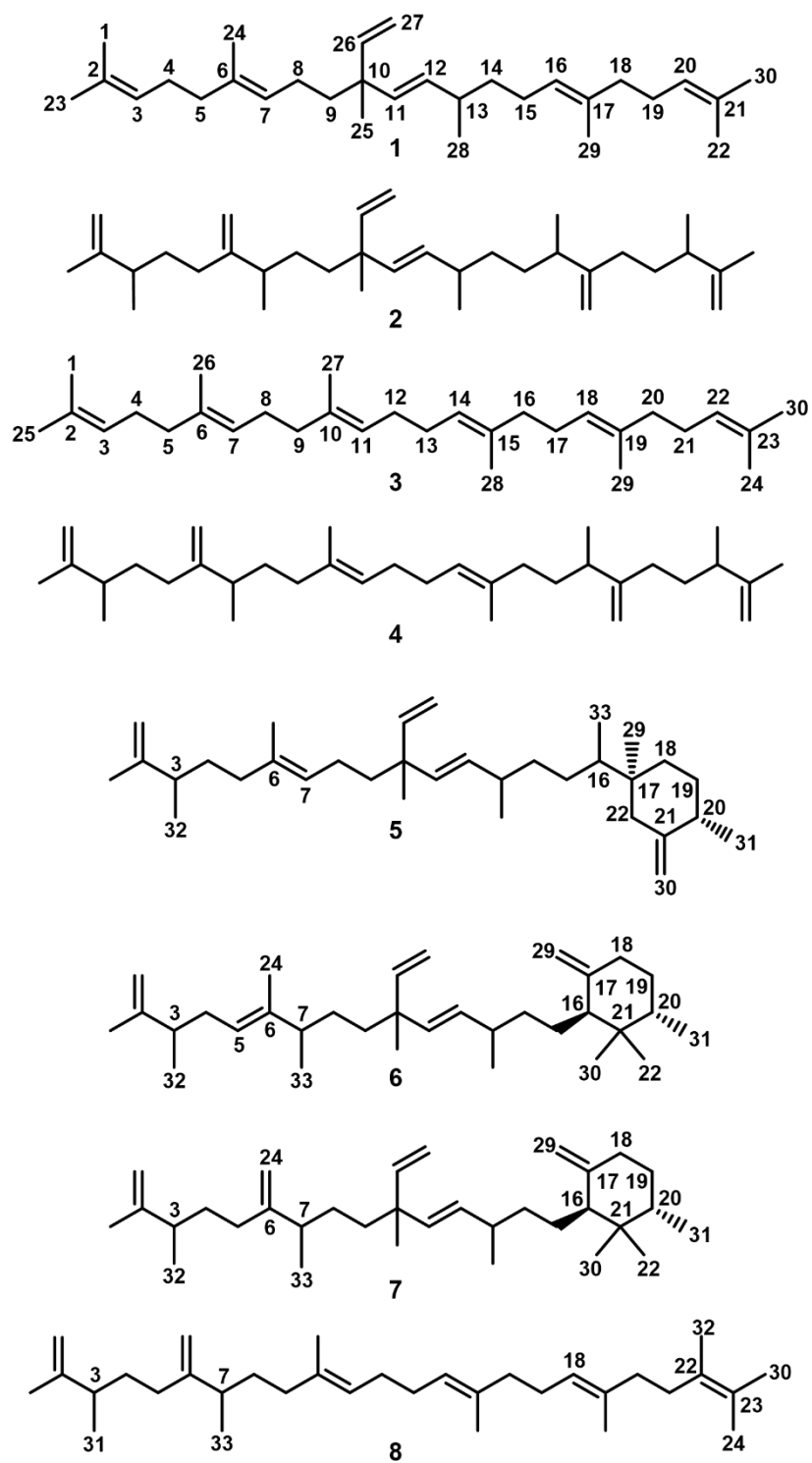


Figure 8 Structures of *B. braunii* race B hydrocarbons.

botryococcene (**2**) homologues in conventional strains up to C₃₆ and C₃₇ in some unique strains^{57, 154}. Approximately 20 different isomers based on the C₃₁ to C₃₄ (**2**) botryococcene structures exist. These isomers occur in acyclic and cyclic forms with the cyclic forms containing methylenecyclohexane rings^{15, 57}. In the B race, C₃₀ squalene (**3**) is also methylated at positions C-3, C-7, C-18, and C-22 to generate C₃₁ monomethylsqualene, C₃₂ dimethylsqualene, C₃₃ trimethylsqualene, and C₃₄ tetramethylsqualene (**4**) homologues^{130, 155}.

2.2 Results and discussion

2.2.1 Purification of *B. braunii* hydrocarbons

In the current study, we undertook an effort to identify new botryococcene and methylsqualene isomers from the Showa (a.k.a. Berkeley) strain of the B race of *B. braunii*. We would expect the presence of new B race hydrocarbons for several possible reasons: different growth conditions between labs may lead to the biosynthesis of alternative molecules, different strains of the B race may biosynthesize discrete hydrocarbons, or improvements in analytical techniques over the years may reveal new molecules not identified previously. From these studies three cyclic C₃₃ botryococcenes, cyclic C₃₃-1 botryococcene (**5**), cyclic C₃₃-2 botryococcene (**6**), and cyclic C₃₃-3 botryococcene (**7**), were purified and structurally characterized (Figure 8). A previous study used mass spectrometry to assign five botryococcene isomers to a single proposed structure similar to **6** and **7**¹⁵⁶. Here we use 2D NMR to resolve two of these structures as **6** and **7**. Furthermore, the structure of one new trimethylsqualene isomer, C₃₃-2

trimethylsqualene (**8**), is also described (Figure 8). *B. braunii* hydrocarbons were purified using a combination of reversed and normal phase HPLC. An initial separation of a total hydrocarbon extract was achieved using reversed phase HPLC, which generated eight peaks, labeled as A-H in Figure 9. Analysis by GC-MS revealed all peaks, except peaks A and H, contained hydrocarbon mixtures (Figure 10). Hydrocarbon molecules were not detected by GC-MS for peak G (not shown), and it was not further studied. Next, individual molecules from peaks D-F were purified to homogeneity using a normal phase HPLC silica column. The mixtures of molecules found in peaks B and C were not separable by silica column HPLC. Thus, a silver nitrate-coated silica-based HPLC column was used to purify these hydrocarbons to homogeneity as previously done to separate very similar botryococcene isomers¹²⁷. Final purity of all isolated molecules to at least $\geq 95\%$ was confirmed by GC-MS.

2.2.2 Identification of purified *B. braunii* hydrocarbons

Subsequent analysis of the purified hydrocarbons by GC-MS and 2D NMR spectroscopy identified the following previously identified compounds (Table 1 and Figure 11): From peak A, C₃₀ botryococcene (**1**) at $m/z = 410.4$ (NMR data not recorded); from peak B, C₃₂ botryococcene at $m/z = 438.5$, braunicene at $m/z = 438.5$, and showacene at $m/z = 424.4$; from peak C, wolficene at $m/z = 424.5$ and monomethylsqualene at $m/z = 424.5$; from peak D, dimethylsqualene at $m/z = 438.5$ and a C₃₄ botryococcene isomer at $m/z = 466.7$; from peak E, trimethylsqualene at $m/z = 452.5$; and from peak F,

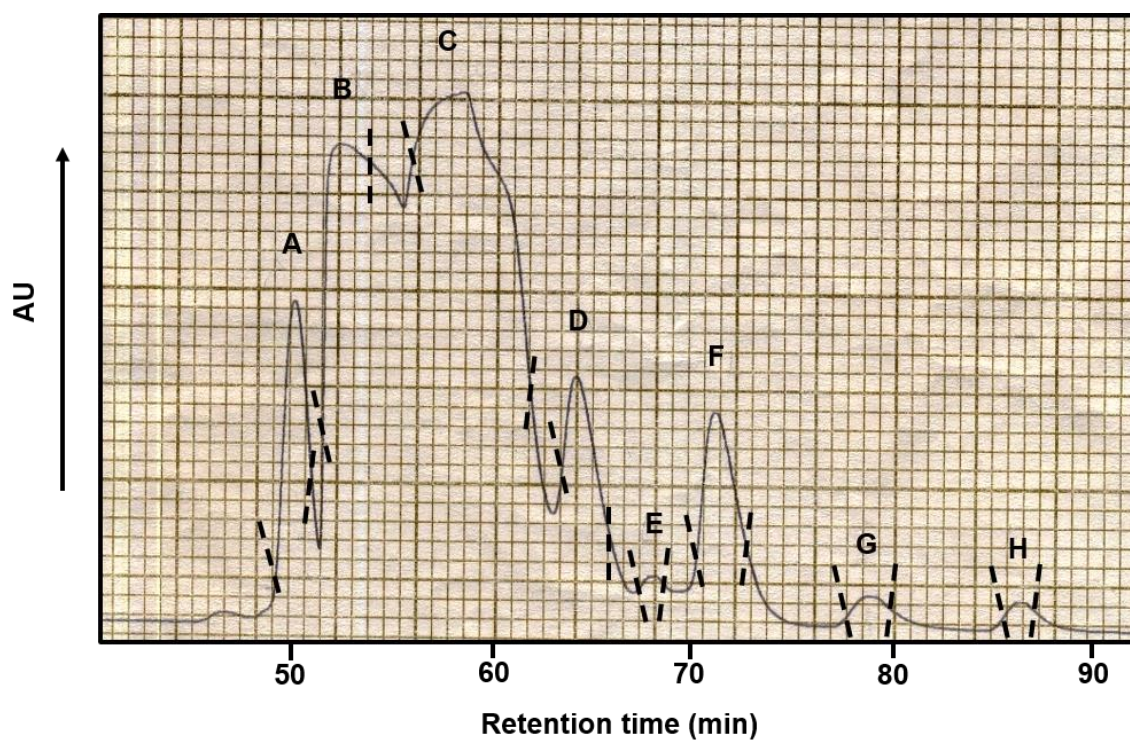


Figure 9 Chromatograph of *B. braunii* race B hydrocarbons from C₁₈ HPLC. The hydrocarbon peaks labeled A to H indicate the collected fractions, and dashed vertical lines indicate the portion of each peak that was collected.

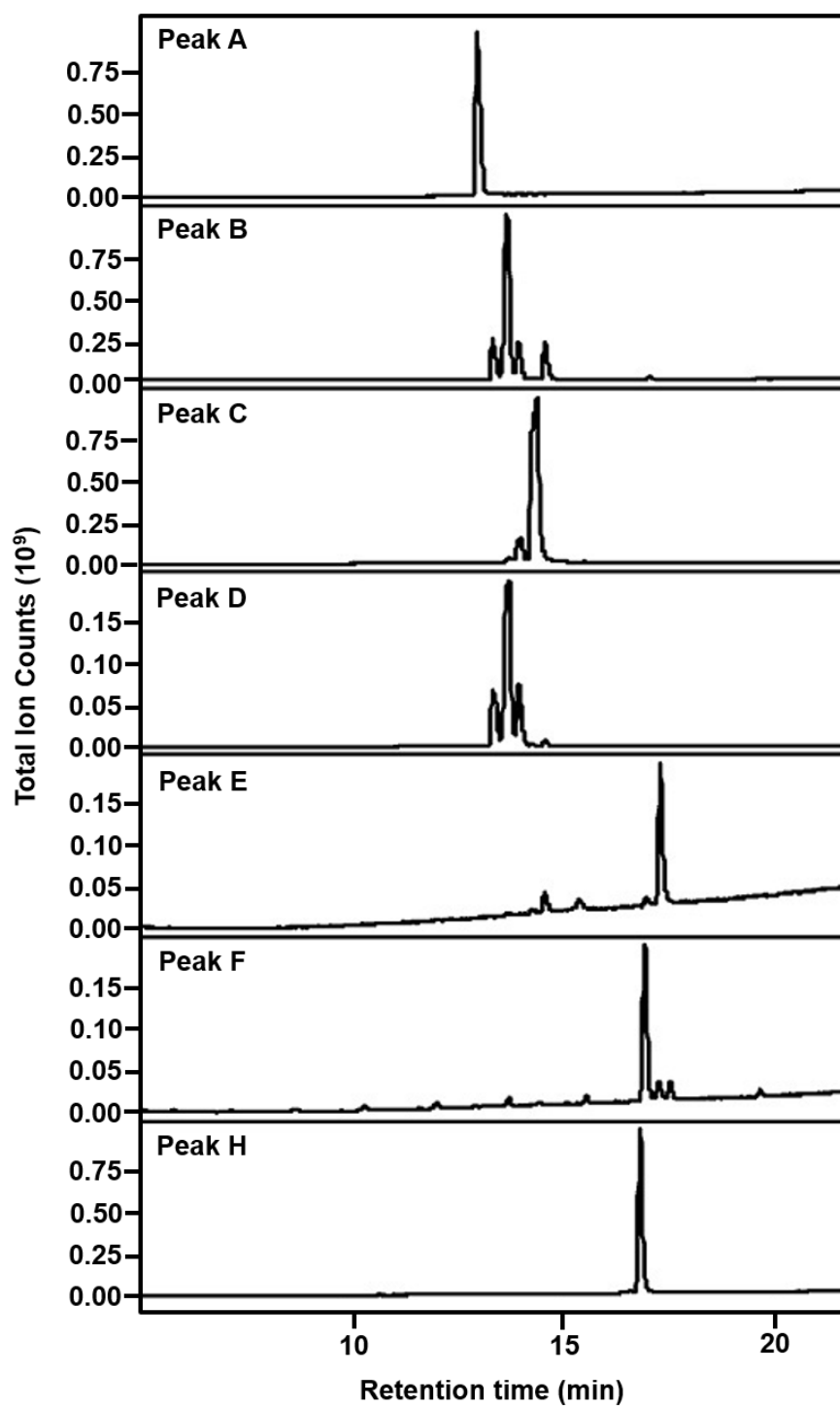


Figure 10 GC-MS total ion count profiles of each C₁₈ HPLC fraction (A-H) shown in Figure 9.

	C₃₂ Botry.¹²⁷	Braunicene¹²⁶	Showacene¹²⁴	Wolficene¹²⁵	Monomethyl Isqualene¹³⁰	Dimethyl squalene¹³⁰	C₃₄ Botry. isomer¹²⁸	Tetramethyl squalene¹³⁰
Position	δ_C, type	δ_C, type	δ_C, type	δ_C, type	δ_C, type	δ_C, type	δ_C, type	δ_C, type
1	109.4, CH ₂	109.5, CH ₂	17.9, CH ₃	17.9, CH ₂	109.6, CH ₂	109.6, CH ₂	109.3, CH ₂	109.6, CH ₂
2	150.2, C	ND ^a , C	ND, C	ND, C	150.1, C	ND, C	ND, C	ND, C
3	40.7, CH	41.0, CH	124.7, CH	124.6, CH	40.9, CH	40.9, CH	41.2, CH	41.2, CH
4	33.3, CH ₂	33.6, CH ₂	26.9, CH ₂	27.0, CH ₂	33.6, CH ₂	33.6, CH ₂	33.6, CH ₂	33.6, CH ₂
5	37.4, CH ₂	37.7, CH ₂	39.9, CH ₂	39.9, CH ₂	37.7, CH ₂	37.7, CH ₂	124.1, CH	31.9, CH ₂
6	135.0, C	ND, C	ND, C	ND, C	135.5, C	ND, C	ND, C	ND, C
7	124.5, CH	124.9, CH	125.0, CH	125.0, CH	124.5, CH	124.5, CH	34.7, CH	39.8, CH
8	23.1, CH ₂	23.4, CH ₂	23.3, CH ₂	23.4, CH ₂	26.9, CH ₂	26.9, CH ₂	29.5, CH ₂	34.2, CH ₂
9	41.4, CH ₂	41.6, CH ₂	41.5, CH ₂	41.6, CH ₂	40.0, CH ₂	40.0, CH ₂	39.7, CH ₂	37.7, CH ₂
10	41.8, C	ND, C	ND, C	ND, C	135.5, C	ND, C	ND, C	ND, C
11	135.8, CH	135.7, CH	136.0, CH	135.7, CH	124.5, CH	124.5, CH	136.0, CH	124.3, CH
12	133.8, CH	134.5, CH	134.0, CH	134.5, CH	28.5, CH ₂	28.5, CH ₂	134.0, CH	28.5, CH ₂
13	36.7, CH	37.6, CH	36.9, CH	37.6, CH	28.5, CH ₂	28.5, CH ₂	37.5, CH	28.5, CH ₂
14	37.5, CH ₂	36.1, CH ₂	37.7, CH ₂	36.0, CH ₂	124.5, CH	124.5, CH	35.2, CH ₂	124.3, CH
15	25.8, CH ₂	32.3, CH ₂	26.0, CH ₂	32.3, CH ₂	135.5, C	ND, C	33.6, CH ₂	ND, C
16	124.5, CH	56.0, CH	124.8, CH	56.0, CH	40.0, CH ₂	40.0, CH ₂	40.3, CH	37.8, CH ₂
17	135.0, C	ND, C	ND, C	ND, C	26.9, CH ₂	26.9, CH ₂	ND, C	33.6, CH ₂
18	37.4, CH ₂	31.2, CH ₂	37.6, CH ₂	31.2, CH ₂	124.5, CH	124.5, CH	31.8, CH ₂	39.8, CH
19	33.3, CH ₂	24.2, CH ₂	33.5, CH ₂	24.2, CH ₂	135.5, C	ND, C	33.0, CH ₂	ND, C
20	40.7, CH	35.0, CH	40.9, CH	35.0, CH	40.0, CH ₂	37.7, CH ₂	41.8, CH	31.9, CH ₂
21	150.2, C	ND, C	ND, C	ND, C	26.9, CH ₂	33.6, CH ₂	ND, C	33.6, CH ₂
22	109.4, CH ₂	27.1, CH ₃	109.6, CH ₂	27.1, CH ₃	124.5, CH	40.9, CH	109.7, CH ₂	41.2, CH
23	19.0, CH ₃	19.2, CH ₃	25.9, CH ₃	25.9, CH ₃	150.1, C	ND, C	19.1, CH ₃	ND, C
24	16.0, CH ₃	16.1, CH ₃	16.2, CH ₃	16.1, CH ₃	17.9, CH ₃	109.6, CH ₂	18.2, CH ₃	109.7, CH ₂
25	23.5, CH ₃	23.8, CH ₃	23.8, CH ₃	23.8, CH ₃	19.2, CH ₃	19.2, CH ₃	23.8, CH ₃	19.2, CH ₃
26	146.8, CH	147.1, CH	147.0, CH	147.1, CH	16.2, CH ₃	16.2, CH ₃	147.2, CH	107.5, CH ₂
27	111.1, CH ₂	111.3, CH ₂	111.3, CH ₂	111.3, CH ₂	16.2, CH ₃	16.2, CH ₃	111.2, CH ₂	16.2, CH ₃
28	21.1, CH ₃	21.2, CH ₃	21.3, CH ₃	21.2, CH ₃	16.2, CH ₃	16.2, CH ₃	21.4, CH ₃	16.2, CH ₃
29	16.0, CH ₃	109.4, CH ₂	16.2, CH ₃	109.3, CH ₂	16.2, CH ₃	16.2, CH ₃	107.4, CH ₂	107.5, CH ₂
30	19.0, CH ₃	21.9, CH ₃	19.2, CH ₃	21.9, CH ₃	25.9, CH ₃	19.2, CH ₃	19.9, CH ₃	19.2, CH ₃
31	19.7, CH ₃	16.9, CH ₃	19.9, CH ₃	16.9, CH ₃	19.9, CH ₃	19.9, CH ₃	19.6, CH ₃	19.9, CH ₃
32	19.7, CH ₃	19.9, CH ₃				19.9, CH ₃	20.0, CH ₃	20.4, CH ₃
33							20.5, CH ₃	20.4, CH ₃
34							19.8, CH ₃	19.9, CH ₃

a) ND: Not determined in this study, but has been shown in literature.

Table 1 ¹³C NMR assignments for hydrocarbons from race B. δ denotes ¹³C chemical shifts determined for carbons of each hydrocarbon molecule

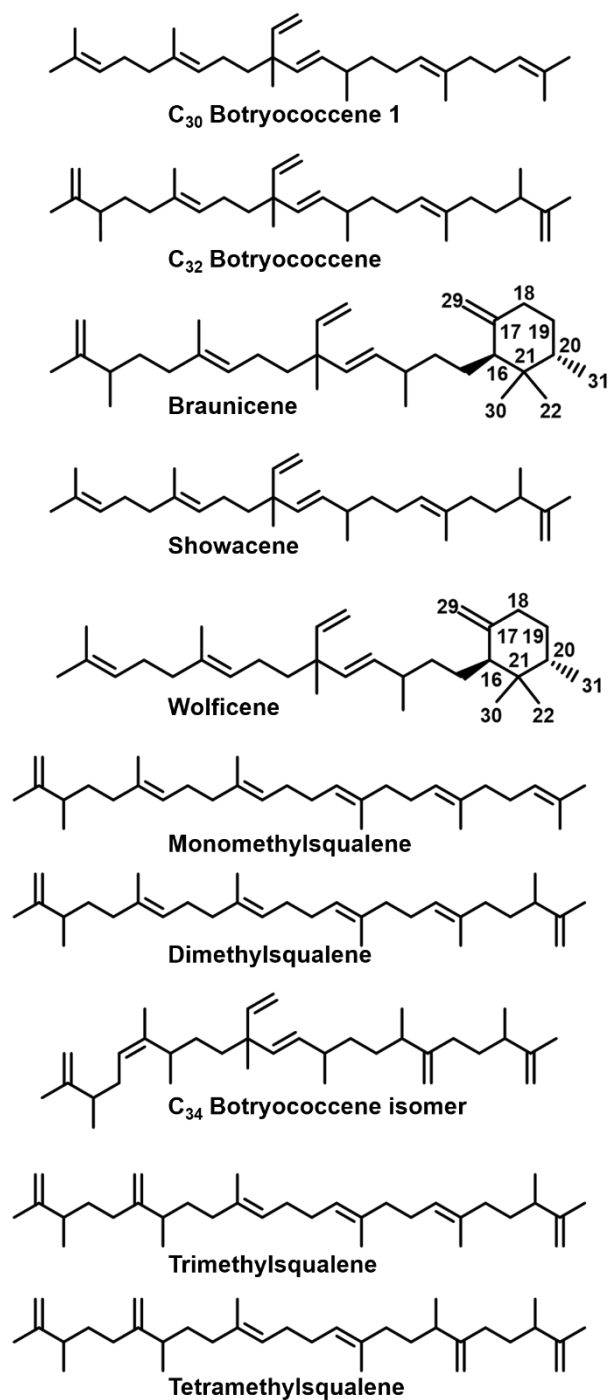


Figure 11 Structures for known botryococcenes and methylsqualenes isolated in this study.

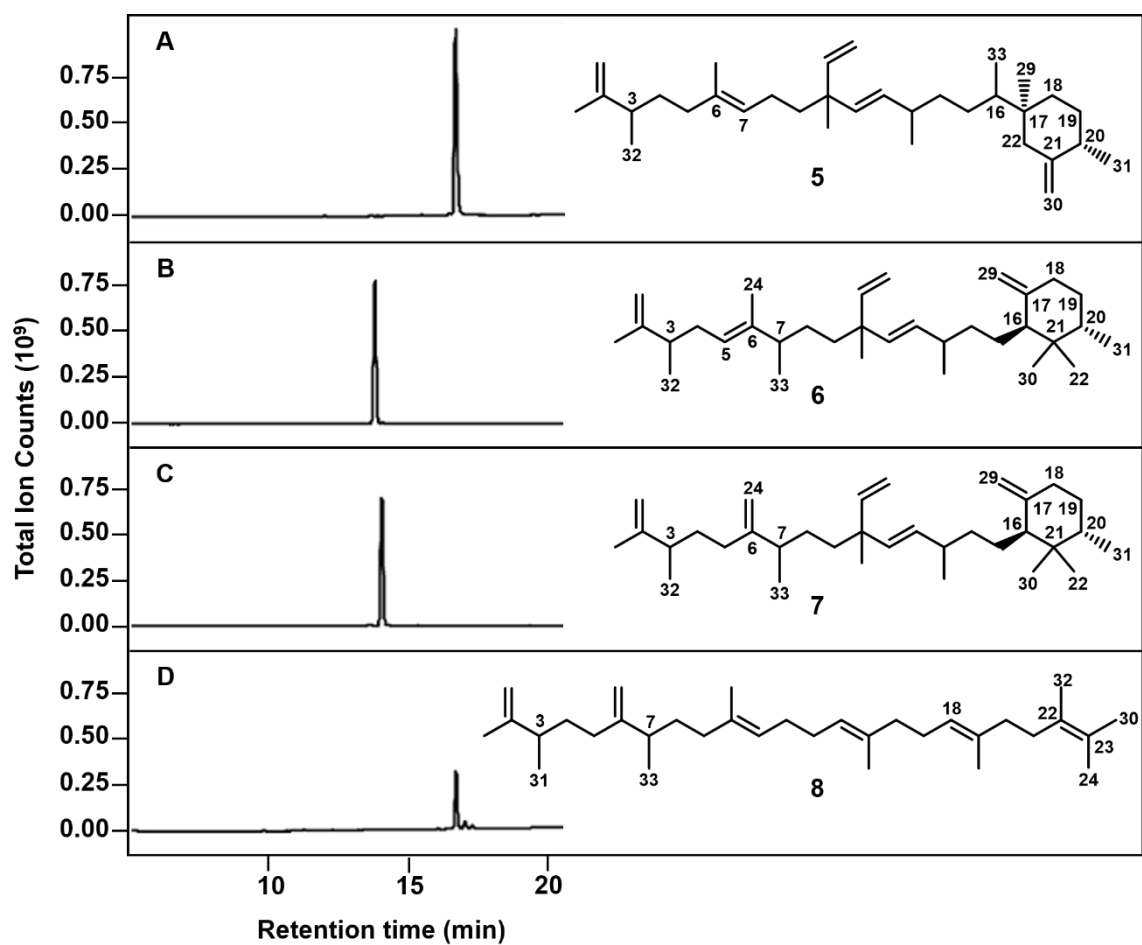


Figure 12 GC-MS total ion count profiles and structures for 5 (A), 6 (B), 7, (C) and 8 (D).

tetramethylsqualene at $m/z = 466.6$. Additionally, compounds **5-8** were identified during this process. Peak H contained the cyclic C_{33} botryococcene **5** (Figure 12A). Peak C contained the two additional cyclic C_{33} botryococcene isomers, **6** and **7** (Figure 12B and C). Peak F contained the trimethylsqualene isomer **8** (Figure 12D). Analysis by GC-MS in EI and CI mode confirmed parent ions for **5-8** at $m/z = 452.5$, 452.6 , 452.7 , and 452.5 , respectively, and the presence of several signature fragments for isoprenoids at $m/z = 69$, 81 , 95 , and 121 (Figures 13 and 14). The molecular formula of compounds **5-8** was determined as $C_{33}H_{56}$ from HRMS in CI mode with parent ions at $m/z = 452.4372$, 452.4373 , 452.4384 , 452.4372 (calculated as 452.4382) for **5-8**, respectively (Figure 15).

2.2.3 NMR and structure analysis

Each hydrocarbon was subjected to 1H and ^{13}C NMR analysis. Initially, COSY and TOCSY data were acquired to resolve severe signal overlap characteristics of the hydrocarbon molecules. We have relied on HSQC to deduce 1H and ^{13}C chemical shifts as reported in Tables 2 and 3 and ^{13}C chemical shifts in Tables 1, 4, 5, and 6. The molecules shown in Figure 11, were previously reported in the literature^{124-128, 130}, and the ^{13}C chemical shifts we observed matched those reported (C_{33} botryococcene in Table 4, all others in Table 1). Additionally, TOCSY-HSQC, H2BC, and HMBC were performed to elucidate the structures of molecules **5-8**. The HSQC spectra of these molecules (Figure 16) were obtained by reducing the spectral width in the indirect dimension to produce intentional ^{13}C resonance aliasing. The corresponding true ^{13}C chemical shifts (Tables 1-6) were calculated by adding the spectral width to the observed shifts.

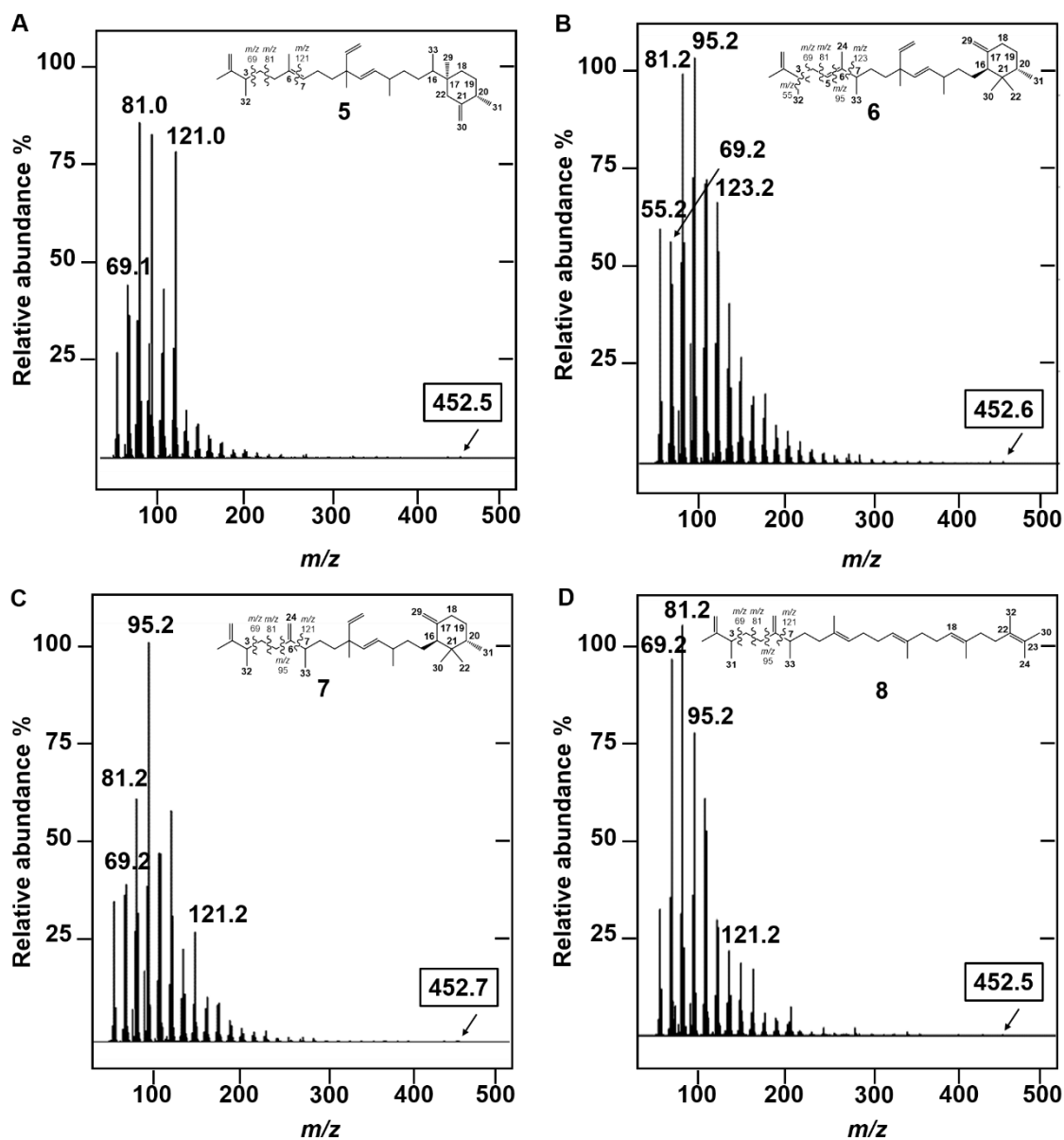


Figure 13 The EI mass spectra from GC-MS and structures for **5** (A), **6** (B), **7** (C) and **8** (D).

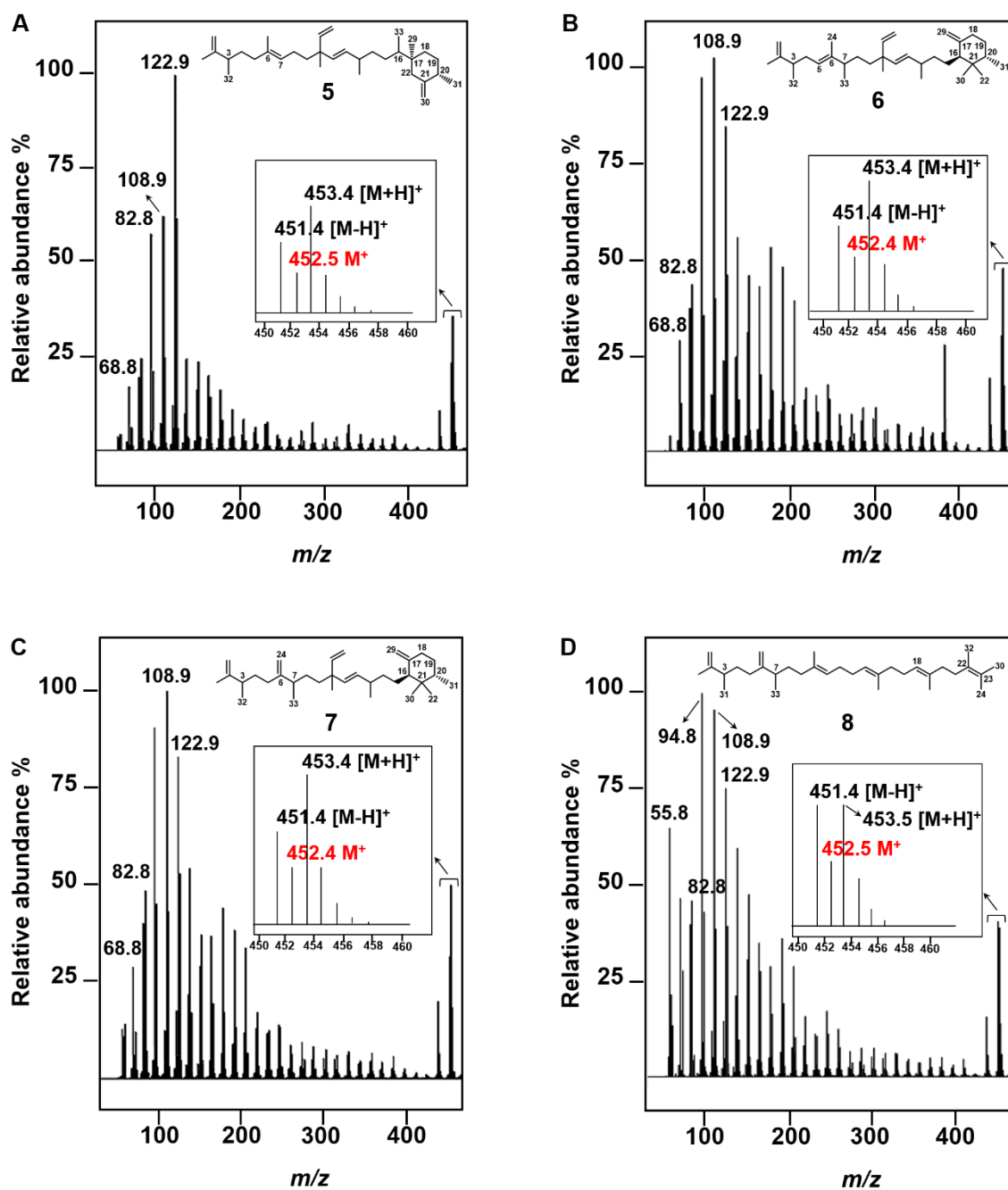


Figure 14 The mass spectra from chemical ionization GC-MS for **5** (A), **6** (B), **7** (C), and **8** (D). Spectra in boxes shows a blow up of the parent ion region indicating the parent ion (shown in red) and other ionic species.

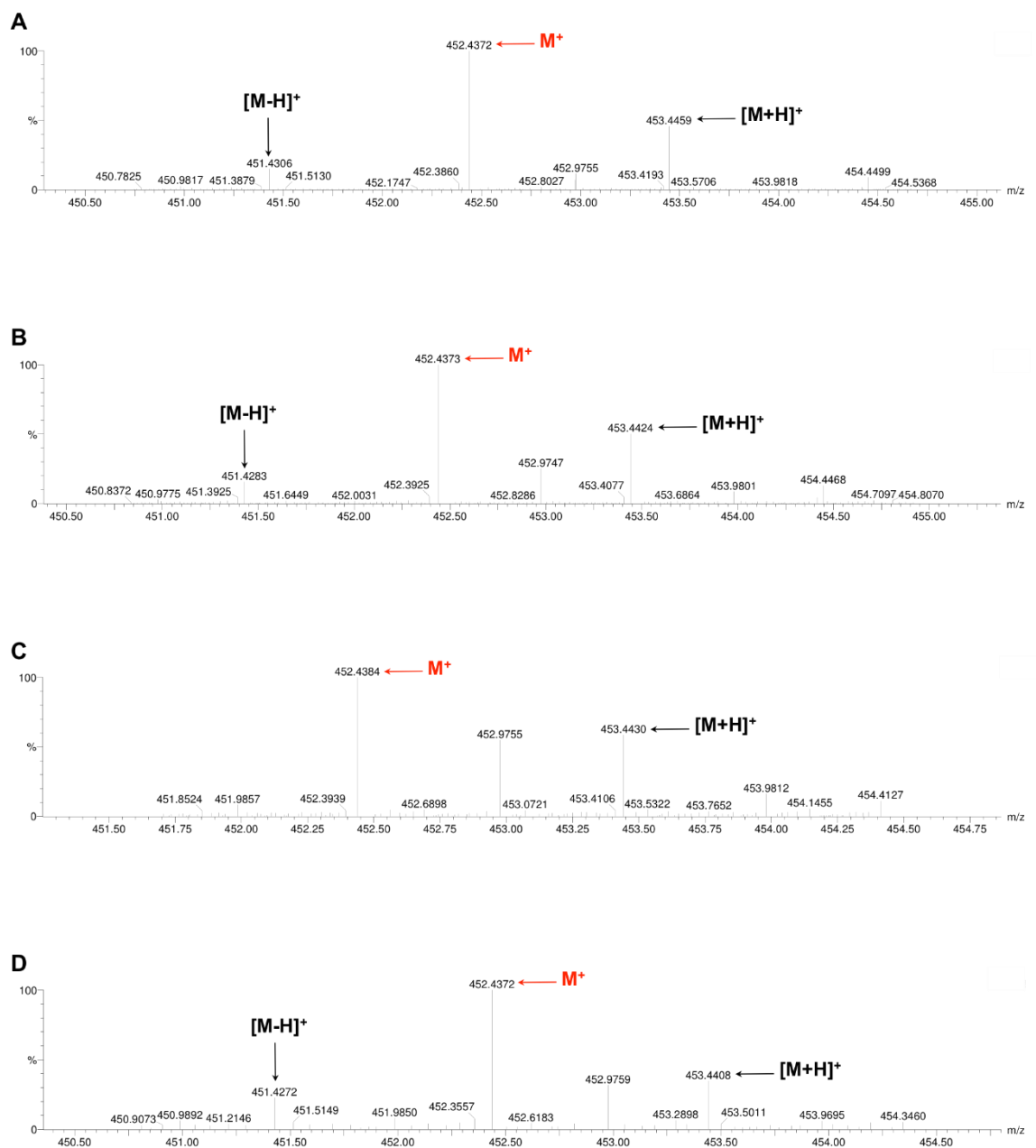


Figure 15 High resolution mass spectra from chemical ionization MS for **5** (A), **6** (B), **7** (C), and **8** (D). The parent ion is shown in red and other ionic species are shown in black.

5		
Position	δ_C , type	δ_H (<i>J</i> in Hz)
1	109.6, CH ₂	4.62 4.71
2	150.2, C	-
3	41.0, CH	2.10
4	33.6, CH ₂	1.34 1.45
5	37.7, CH ₂	1.87
6	135.1, C	-
7	124.8, CH	5.07
8	23.3, CH ₂	1.90
9	41.6, CH ₂	1.37
10	42.0, C	-
11	136.0, CH	5.35
12	134.1, CH	5.17
13	37.5, CH	2.07
14	36.4, CH ₂	1.16 1.34
15	28.7, CH ₂	0.86 1.48
16	43.9, CH	1.11
17	38.4, C	-
18a	35.4, CH ₂	[eq] 1.44, <i>m</i>
18b		[ax] 1.30, <i>ddd</i> (13.2, 12.6, 3.8)
19a	32.3, CH ₂	[eq] 1.62, <i>m</i>
19b		[ax] 1.14, <i>m</i>
20	37.6, CH	1.93, <i>m</i>
21	152.2, C	-
22	46.4, CH ₂	[eq] 2.04, <i>m</i> [ax] 1.86, <i>m</i>
23	19.2, CH ₃	1.65
24	16.2, CH ₃	1.56
25	23.8, CH ₃	1.07
26	147.0, CH	5.80
27	111.3, CH ₂	4.93
28	21.8, CH ₃	0.97
29	18.4, CH ₃	0.66, <i>s</i>
30	106.0, CH ₂	4.61, <i>d</i> (4.8)
31	18.5, CH ₃	1.04, <i>d</i> (7.2)
32	19.9, CH ₃	1.00
33	13.8, CH ₃	0.80

Table 2 ¹³C and ¹H NMR Spectroscopic Data (¹H 500 MHz, ¹³C 125 MHz, CDCl₃) for 5.

6			7 ^a			8	
Position	δ_C , type	δ_H (J in Hz)	Position	δ_C , type	δ_H (J in Hz)	δ_C , type	δ_H (J in Hz)
1	109.3, CH ₂	4.53	1	109.7, CH ₂	4.64, 4.73	109.7, CH ₂	4.64, 4.73
		4.70	2	150.1, C	-	150.0, C	-
2	149.6, C	-	3	41.2, CH	2.15	41.2, CH	2.16
3	41.8, CH	2.13	4	33.6, CH ₂	1.50	33.6, CH ₂	1.41, 1.51
4	33.0, CH ₂	1.98	5	31.8, CH ₂	1.87	31.9, CH ₂	1.90
		2.08	6	155.2, C	-	154.7, C	-
5	124.1, CH	5.10	7	40.8, CH	2.00	39.8, CH	2.06
6	139.0, C	-	8	33.5, CH ₂	1.40	34.2, CH ₂	1.35, 1.51
7	34.7, CH	2.51	9	39.2, CH ₂	1.28	37.7, CH ₂	1.89
8	29.5, CH ₂	1.23	10	41.8, C	-	135.4, C	-
9	39.7, CH ₂	1.23	11	135.9, CH	5.29	124.5, CH	5.16
10	41.3, C	-	12	134.3, CH	5.15	28.5, CH ₂	2.01
11	135.9, CH	5.29	13	37.5, CH	2.01	28.5, CH ₂	2.01
12	134.4, CH	5.15	14	36.0, CH ₂	0.97, 1.16	124.5, CH	5.16
13	37.5, CH	2.01	15	32.3, CH ₂	1.21, 1.46	135.4, C	-
14	36.0, CH ₂	0.97	16	56.8, CH	1.62	40.0, CH ₂	2.06
		1.16	17	148.9, C	-	26.9, CH ₂	2.12
15	32.3, CH ₂	1.21	18	31.2, CH ₂	2.05	124.5, CH	5.08
		1.46	19	24.2, CH ₂	1.30, 1.42	135.4, C	-
16	56.8, CH	1.62, <i>m</i>	20	35.0, CH	1.61	37.7, CH ₂	1.90
17	148.9, C	-	21	36.4, C	-	33.6, CH ₂	1.41, 1.51
18	31.2, CH ₂	2.05, <i>m</i>	22	27.1, CH ₃	0.87	154.7, C	-
19a	24.1, CH ₂	[eq] 1.42, <i>m</i>	23	19.1, CH ₃	1.66	131.6, C	-
19b		[ax] 1.30, <i>m</i>	24	107.6, CH ₂	4.66, 4.73	17.9, CH ₃	1.62
20	35.0, CH	1.61, <i>m</i>	25	23.8, CH ₃	1.03	19.2, CH ₃	1.65
21	36.4, C	-	26	147.2, CH	5.76	107.5, CH ₂	4.68, 4.75
22b	27.1, CH ₃	[ax] 0.87, <i>m</i>	27	111.2, CH ₂	4.90, 4.93	16.2, CH ₃	1.60
23	19.9, CH ₃	1.68	28	21.2, CH ₃	0.93	16.2, CH ₃	1.60
24	18.2, CH ₃	1.53	29	109.3, CH ₂	4.46, 4.55	16.2, CH ₃	1.60
25	23.8, CH ₃	1.03	30	21.9, CH ₃	0.73	25.9, CH ₃	1.69
26	147.2, CH	5.76	31	16.0, CH ₃	0.77	19.9, CH ₃	1.00
27	111.2, CH ₂	4.90	32	20.0, CH ₃	1.02	19.9, CH ₃	1.00
		4.93	33	20.6, CH ₃	0.98	20.4, CH ₃	1.00
28	21.2, CH ₃	0.93					
29a	109.3, CH ₂	4.63, <i>d</i> (3.5)					
29b		4.46, <i>d</i> (3.5)					
30	21.9, CH ₃	0.73, <i>d</i> (4.8)					
31	16.0, CH ₃	0.77, <i>d</i> (7.2)					
32	19.6, CH ₃	1.00					
33	19.8, CH ₃	0.93					

^aNOE measurements to determine relative configuration of the cyclohexane ring were not performed since the ¹H and ¹³C chemical shifts for the ring positions are identical to 6.

Table 3 ¹³C and ¹H NMR Spectroscopic Data (¹H 500 MHz, ¹³C 125 MHz, CDCl₃) for 6, 7^a, and 8.

C₃₃ Botryococcene^a		Trimethylsqualene^b
Position	δ_c, type	δ_c, type
1	109.3, CH ₂	109.6, CH ₂
2	150.1, C	150.1 ^c , C
3	40.8, CH	41.2, CH
4	33.4, CH ₂	33.6, CH ₂
5	37.5, CH ₂	31.9, CH ₂
6	135.0, C	155.0 ^c , C
7	124.6, CH	39.8, CH
8	23.1, CH ₂	34.2, CH ₂
9	41.3, CH ₂	37.6, CH ₂
10	42.0, C	135.5 ^c , C
11	135.7, CH	124.5, CH
12	133.9, CH	28.5, CH ₂
13	37.3, CH	28.5, CH ₂
14	35.0, CH ₂	124.5, CH
15	33.4, CH ₂	135.3 ^c , C
16	40.1, CH	40.0, CH ₂
17	155.0, C	26.8, CH ₂
18	31.6, CH ₂	124.5, CH
19	33.4, CH ₂	135.2 ^c , C
20	41.0, CH	37.7, CH ₂
21	150.2, C	33.6, CH ₂
22	109.5, CH ₂	40.9, CH
23	19.0, CH ₃	150.3 ^c , C
24	15.9, CH ₃	109.6, CH ₂
25	23.5, CH ₃	19.2, CH ₃
26	146.8, CH	107.5, CH ₂
27	111.1, CH ₂	16.2, CH ₃
28	21.1, CH ₃	16.2, CH ₃
29	107.2, CH ₂	16.2, CH ₃
30	18.9, CH ₃	19.2, CH ₃
31	19.8, CH ₃	19.9, CH ₃
32	19.7, CH ₃	19.9, CH ₃
33	20.2, CH ₃	20.4, CH ₃

^aAll ¹³C chemical shifts shown for C₃₃ botryococcene taken from a previous publication¹²⁷ for comparison to 5 in Table 2.

^b¹³C chemical shifts shown for trimethylsqualene were determined in the current study except where noted.

^cThese individual ¹³C chemical shifts not determined in this study. Presented ¹³C chemical shifts are from a previous publication⁸ for comparison to 8 in Table 3.

Table 4 ¹³C assignments for previously identified hydrocarbons used for comparison to 5-7 and 8. δ denotes ¹³C chemical shifts determined for each hydrocarbon molecule.

5			
Position	$^1J_{C-H}$ (J in Hz), type	NOE	HMBC (1H)
1	153.0, CH ₂ 153.0	nd ^a nd	3, 23
2	-	-	1, 3, 4, 23, 32
3	127.8, CH	nd	1, 4, 23, 32
4	124.2, CH ₂ 129.6	nd nd	3, 5, 32
5	120.6, CH ₂	nd	4, 7, 24
6	-	-	5, 24
7	151.2, CH	nd	8, 9, 24
8	124.2, CH ₂	nd	7, 9
9	127.8, CH ₂	nd	8
10	-	-	9, 11, 12, 25, 26, 27
11	147.0, CH	nd	9, 12, 13, 25, 26
12	146.4, CH	nd	11, 13, 14, 28
13	119.4, CH	nd	11, 12, 14, 28
14	125.4, CH ₂ 127.2	nd nd	13, 28
15	121.2, CH ₂ 127.8	nd nd	14, 16, 33
16	126.0, CH	nd	29, 33
17	-	-	15, 16, 18, 19, 22, 29, 33
18	123.0, CH ₂ 126.6	18a, 19a, 20, 22b, 33 18b, 19a, 20, 22b, 29, 33	19, 22, 29
19	125.4, CH ₂ 130.8	18a, 19a, 29, 33 18b, 19b, 20, 31	18, 31
20	124.2, CH	18b, 18a, 19a, 22b, 31	18, 22, 30, 31
21	-	-	19, 22, 31
22	114.0, CH ₂ 124.8	18b, 20, 22a, 33 16, 22b, 29, 30, 31	29, 30
23	124.8, CH ₃	nd	1
24	125.4, CH ₃	nd	5, 7
25	126.6, CH ₃	nd	11, 26
26	138.0, CH	nd	9, 11, 25, 27
27	168.0, CH ₂	nd	
28	126.0, CH ₃	nd	12
29	125.4, CH ₃	18a, 19b, 22a, 30, 33	22
30	143.4, CH ₂	22a, 29, 31	22
31	126.0, CH ₃	19a, 20, 30	19
32	125.4, CH ₃	nd	3, 4
33	125.4, CH ₃	nd	15, 16

^aNot determined for these positions.

Table 5 ^{13}C and 1H NMR Spectroscopic Data for 5.

6		
Position	$^1J_{C-H}$ (J in Hz), type	NOE
1	154.8, CH ₂	nd ^a
	153.6	nd
2	-	-
3	127.8, CH	nd
4	122.4, CH ₂	nd
	130.2	nd
5	150.0, CH	nd
6	-	-
7	124.2, CH	nd
8	124.2, CH ₂	nd
9	125.4, CH ₂	nd
10	-	-
11	146.4, CH	nd
12	147.6, CH	nd
13	126.0, CH	nd
14	120.6, CH ₂	nd
	123.6	nd
15	112.2, CH ₂	nd
	137.4	nd
16	130.8, CH	14, 14a, 15, 15a, 18, 22, 29b, 30
17	-	-
18	128.4, CH ₂	15b, 29b, 29a
19	112.8, CH ₂	18, 19a, 20, 22
	124.2	18, 19b, 20, 22
20	124.2, CH	18, 19b 19a, 22, 30, 31
21	-	-
22	125.4, CH ₃	16, 19b, 19a, 20, 30, 31
23	125.4, CH ₃	nd
24	125.4, CH ₃	nd
25	126.6, CH ₃	nd
26	135.6, CH	nd
27	153.6, CH ₂	nd
	156.0	nd
28	126.6, CH ₃	nd
29	153.6, CH ₂	29a
	154.8	29b
30	124.8, CH ₃	16, 19b, 20, 22
31	124.8, CH ₃	19b, 19a, 20, 22
32	124.8, CH ₃	nd
33	125.4, CH ₃	nd

^aNot determined for this positions

Table 6 ¹³C and ¹H NMR Spectroscopic Data for 6.

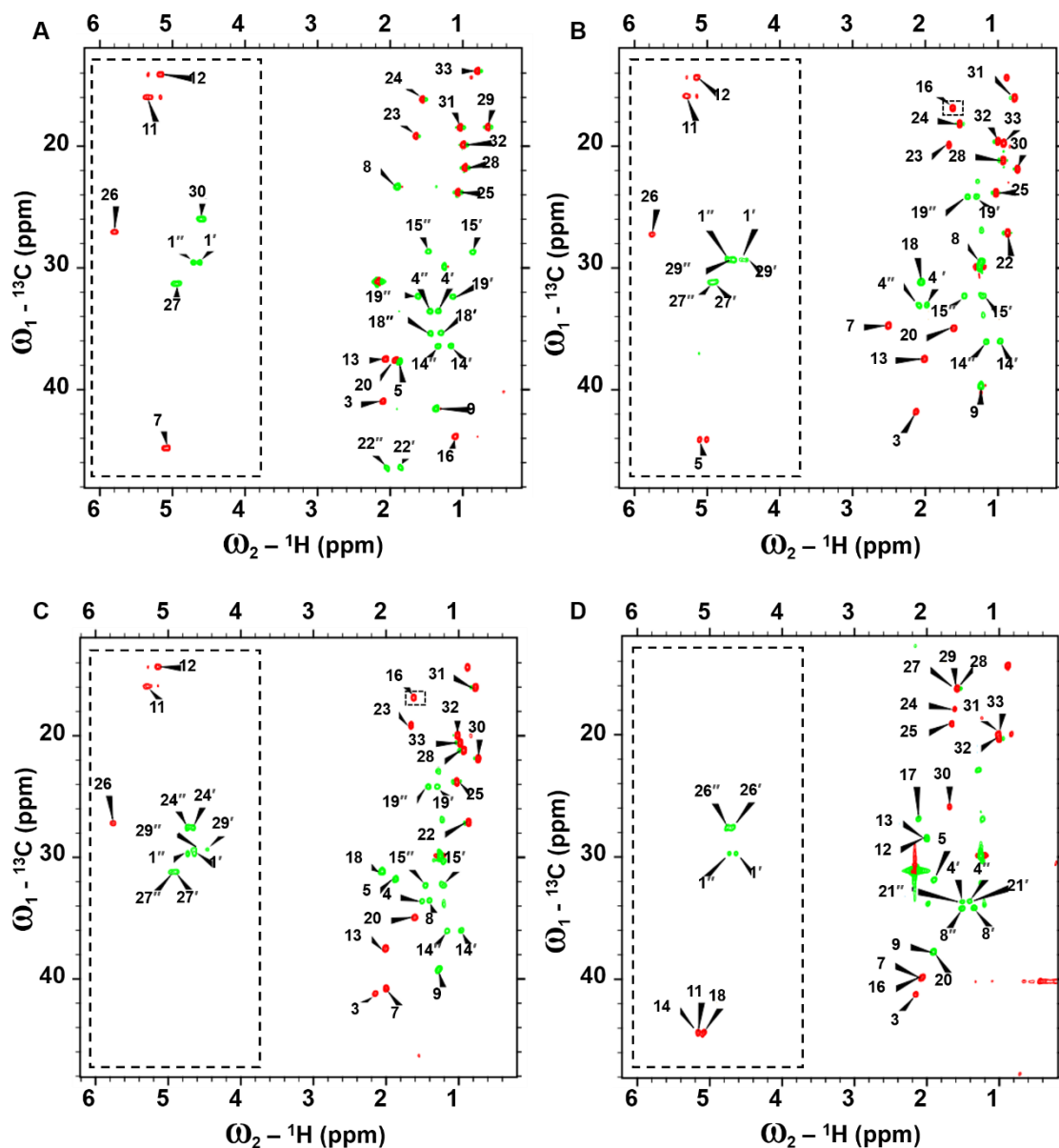


Figure 16 Multiplicity edited ^{13}C -HSQC spectra for 5 (A), 6 (B), 7 (C) and 8 (D). Positive (red) contours represent $-\text{CH}$ or $-\text{CH}_3$ correlations, while negative (green) contours indicate $-\text{CH}_2$ correlations. Peaks originating from the molecule of interest are marked with resonance assignment, while those from impurities are left unmarked. The spectra were acquired at reduced spectral width in the indirect dimension for deliberate aliasing of ^{13}C resonances. Aliased peaks are shown in dashed line boxes and the true ^{13}C chemical shifts were calculated as described in the Experimental section.

Here we have identified the differently cyclized molecule **5**, which describes a new scaffold not reported in the literature for botryococcenes. Molecule **5** has a unique arrangement whereby the six-membered ring is attached to the molecule backbone at position C-17. In previous reports, botryococcenes containing cyclic rings on the same side of the molecule as in **5** the C-16 position is usually seen to bear the cyclic ring^{125, 126, 157} as seen in braunicene^{157, 158} (Figure 11). A botryococcene with a similar methylenecyclohexane ring structure to **5** has been reported^{157, 159}. However, this molecule is a C₃₄ botryococcene with the methylenecyclohexane ring attached to the molecule backbone at C-6. For molecule **5**, the ring is formed from the connection of the methyl carbon in the terminal isoprene (C-22) unit and the third carbon (C-17) in the adjoining isoprene unit.

Due to this unique ring arrangement in **5**, chemical shifts originating from the ring positions differ significantly from other samples studied in this and other reports. The structure of **5** is elucidated by exclusive use of NMR experiments (Figures 16 and 17). By comparison, the ¹³C chemical shifts for C₃₃ botryococcene^{127, 160} and **5** are similar up to position C-13 (Tables 2 and 4 and Figure 18). Then the chemical shifts for C-14, C-15, and C-16 show slight deviation for **5** inferring that these two molecules are structurally similar until position C-16. Thereafter the NMR data is entirely different between the two molecules for carbons 17, 18, 19, 20, 21, and 22 of the ring structure (Table 2). Analysis by HMBC (Figures 17A and 19), COSY (Figures 17B and 20), HSQC-TOCSY (Figure 21), and H2BC (Figure 22) provide evidence that the six-membered ring is formed by C-17, C-18, C-19, C-20, C-21 and C-22. For example, long range HMBC correlations

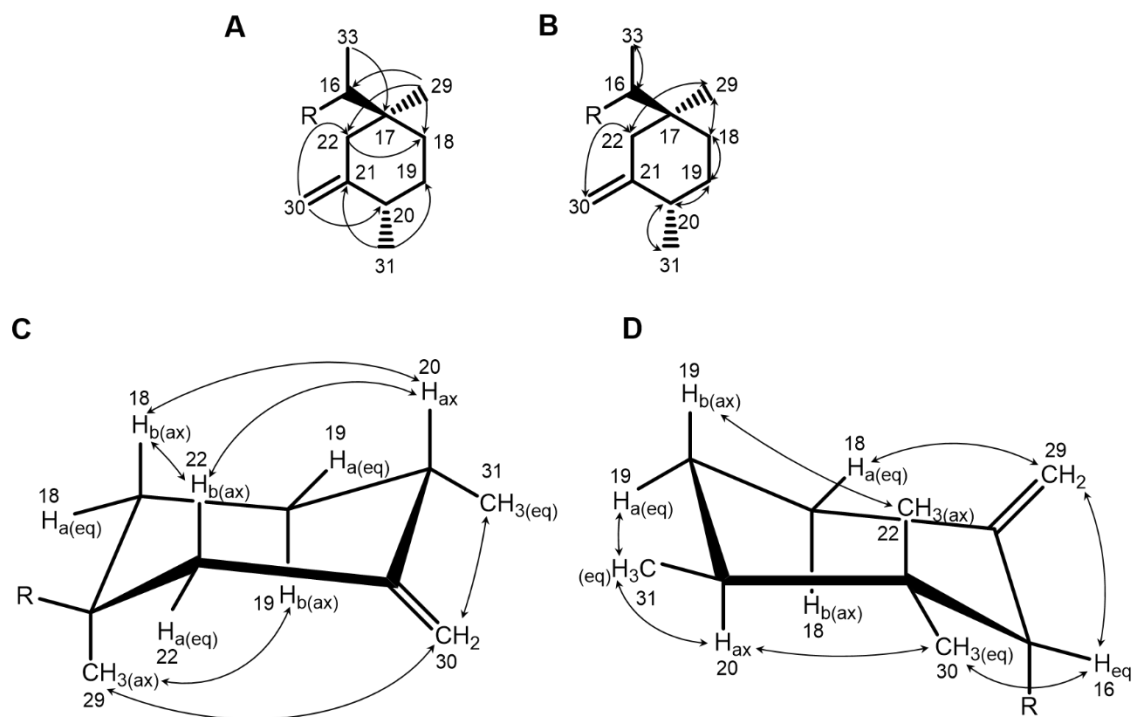
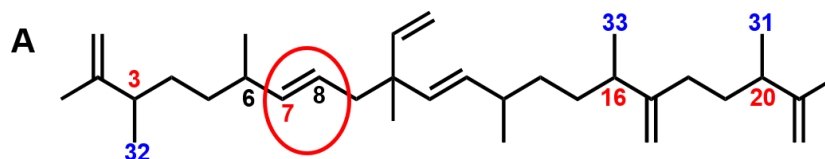
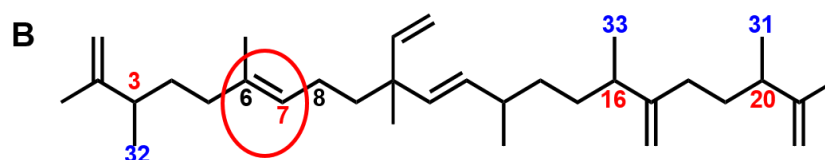


Figure 17 Selected correlations observed in 2D NMR experiments for 5. Only significant correlations shown. (A) HMBC correlations, (B) COSY correlations. C and D, relative configuration of the methylenecyclohexane ring of 5 and 6. Arrows indicate strong NOE connectivities within each ring. (C) Configuration of the methylenecyclohexane ring of 5. (D) Configuration of the methylenecyclohexane ring of 6.



Incorrect C_{33} botryococcene structure, molecule 3-2, as reported in Okada, S., Murakami, M., and Yamaguchi, K. (1997) *Phytochemistry* 8, 198-203.



Corrected C_{33} botryococcene structure, molecule 3-2. All NMR chemical shifts reported in Okada, S., Murakami, M., and Yamaguchi, K. (1997) *Phytochemistry* 8, 198-203 agree with this structure.

Figure 18 Correction of C_{33} botryococcene, molecule 3-2, structure reported in Okada, S., Murakami, M., and Yamaguchi, K. (1997) *Phytochemistry* 8, 198-203 (A) The incorrect structure. (B) The corrected structure.

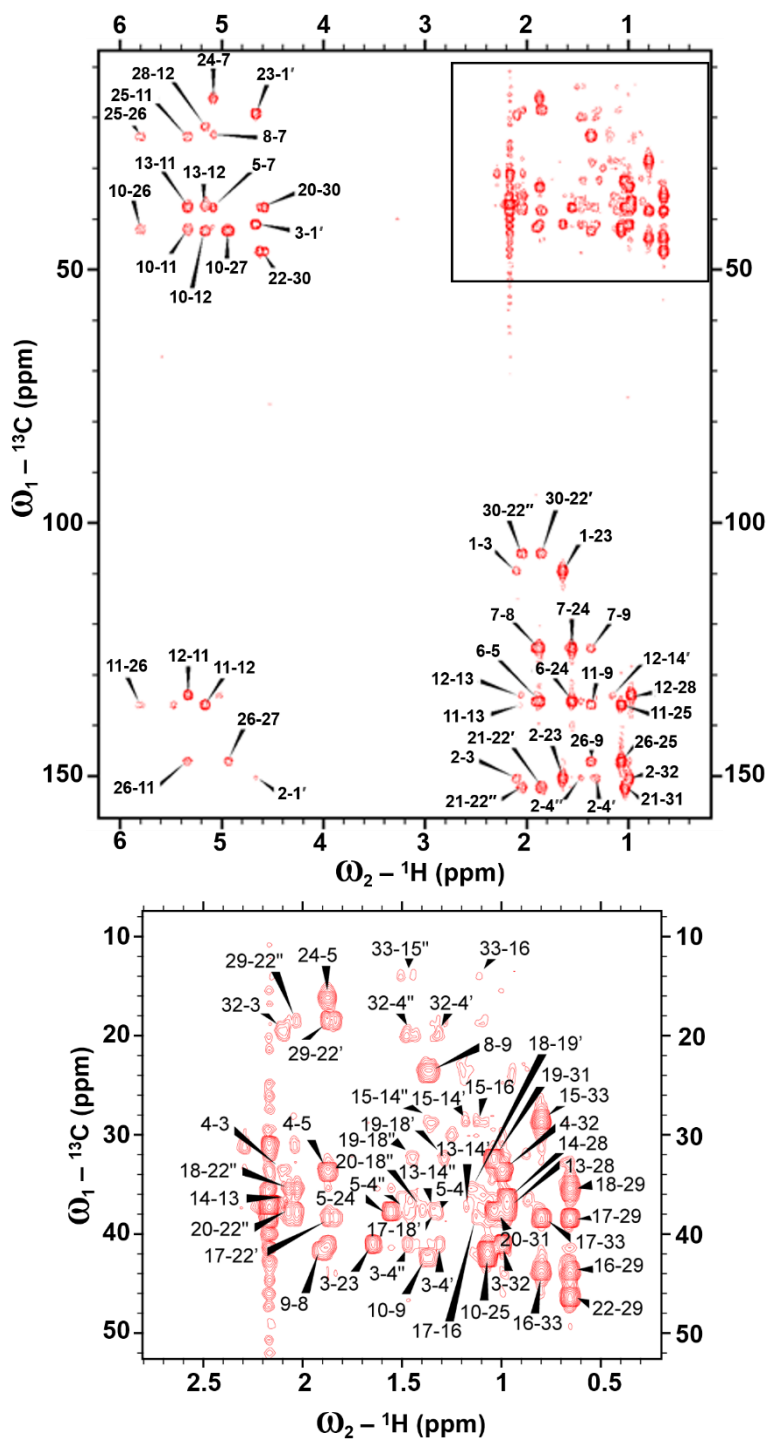


Figure 19 ^{13}C -HMBC spectrum acquired on **5**. The bottom panel is a blow up of the boxed region in the top panel.

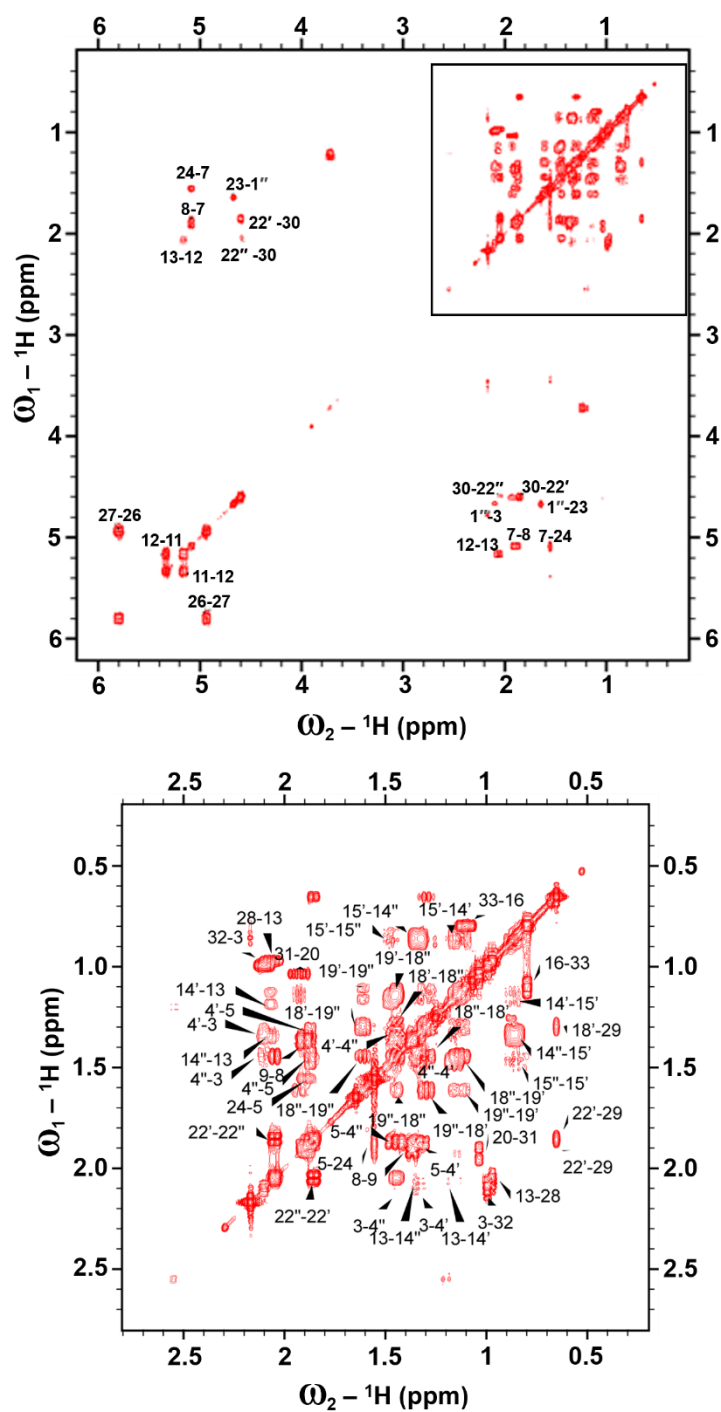


Figure 20 ^1H -COSY spectrum acquired on **5**. The bottom panel is a blow up of the boxed region in the top panel.

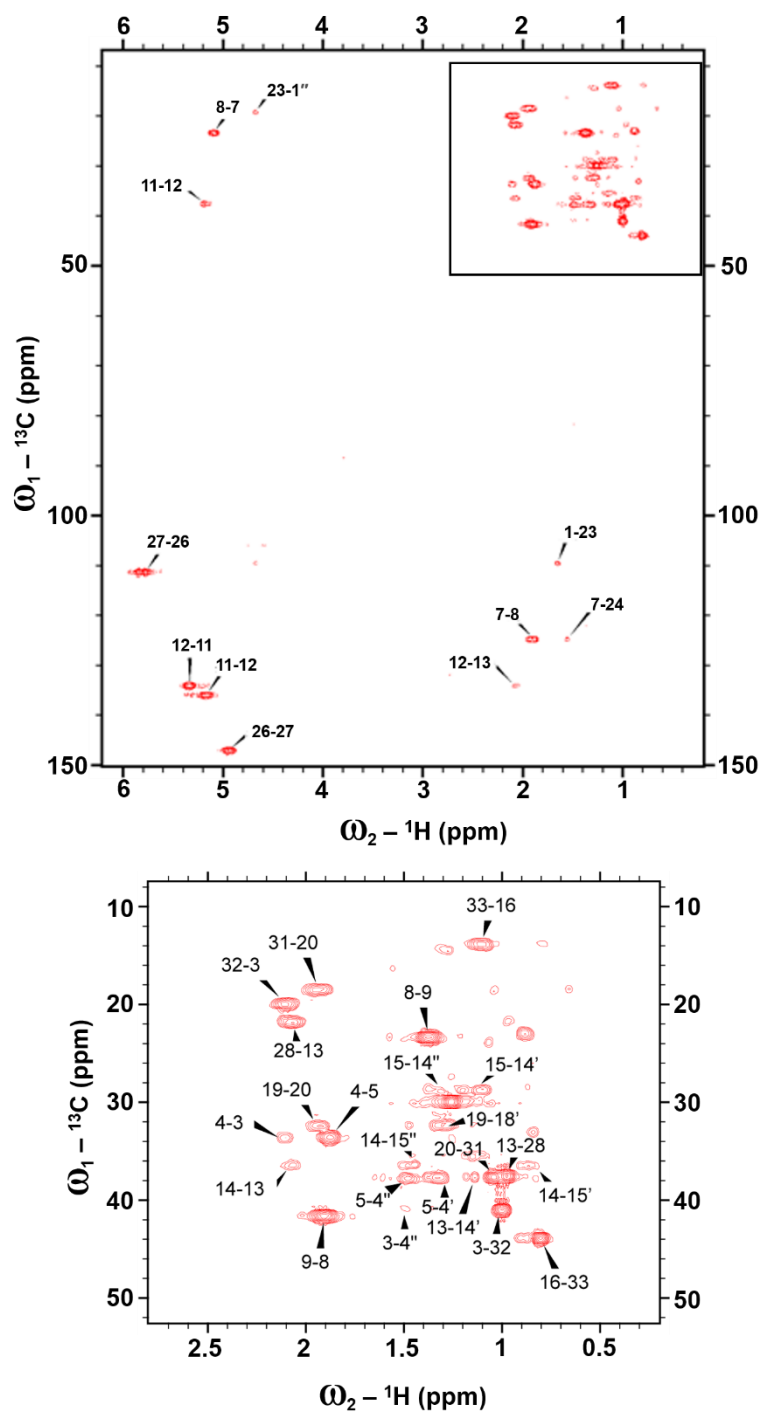


Figure 22 ^{13}C -H2BC spectrum acquired on **5**. The bottom panel is a blow up of the boxed region in the top panel.

between the H-29 methyl protons and C-16, C-18, and C-22 suggest such a ring formation. Moreover, COSY correlations indicate a connection between C-18 and C-19, and HMBC correlations from the H-30 exomethylene protons to C-20 and C-22 confirm this assignment. Those experiments also show the tertiary carbon C-20 methine is connected to the C-31 methyl group. Overall, our 2D NMR data demonstrates a novel botryococcene scaffold for **5**.

Additionally, the relative configuration of the methylenecyclohexane ring in **5** was determined using selective 1D NOESY experiments and complemented with *J*-couplings (Table 5). The observed NOEs between the axial positions H-18b, H-20 and H-22b, as well as between H-19b and the H₃-29 protons, in conjunction with the large vicinal *J*-coupling of 12.6 Hz seen for H-18b and H-19b suggest the cyclohexane ring exists in a stable chair conformation (Figure 17C). The *J*-coupling values for the acyclic portions of the molecule were not determined due extreme overlap that has also be seen in previous studies^{121, 127}.

Based on these studies, the ring formation for **5** can be speculated to occur as proposed in Figure 23. This mechanism involves C-16 methylation to generate a C-17 carbocation, ring closure between C-22 and C-17 creating a quaternary carbon at C-17 with production of a cyclohexyl cation at C-21, and finally hydrogen elimination at C-30.

The remaining two cyclic C₃₃ botryococcenes, **6** and **7** are very similar except for the location of a double bond. This double bond between C-5 and C-6 in **6** is seen to be relocated to between positions C-6 and C-24 in **7**. This can be readily seen in the multiplicity-edited HSQC spectra acquired on these samples (Figure 16B and C).

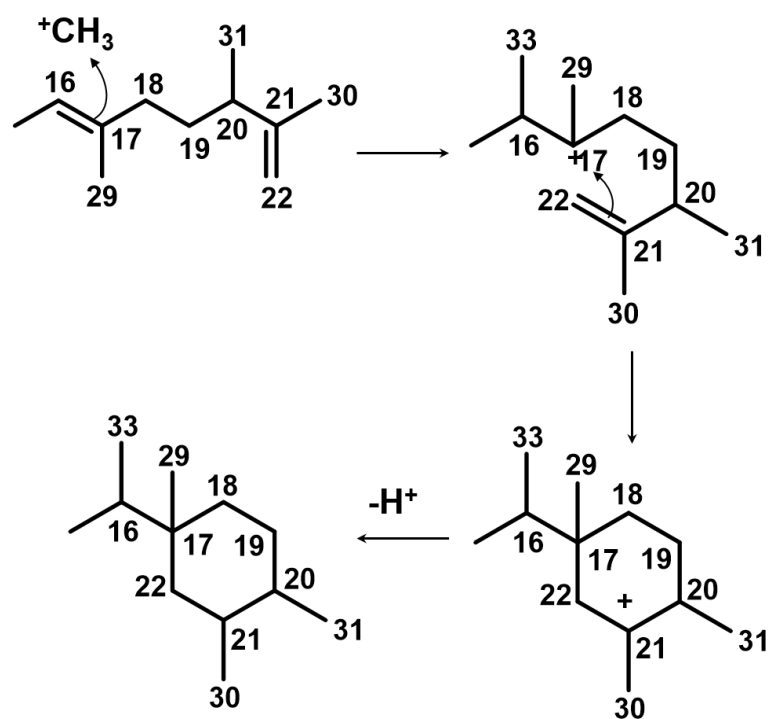


Figure 23 Proposed mechanism for ring formation in 5.

The selective 1D NOESY experiments carried out on **6** (Table 6) and **7** (data not shown) indicate relative ring configurations (Figure 17D) similar to that of braunicene as previously reported^{126, 158}. This conclusion is also supported by the fact that all HSQC peaks corresponding to the cyclohexane ring (C-16 to C-21) of **6**, **7**, and braunicene superimpose on each other (Tables 1 and 3).

The planar structure of **8** identified in our study closely resembles trimethylsqualene¹³⁰ (Tables 4 and 5) with the following differences. The double bond between C-23 and C-24 of trimethylsqualene is shifted to between C-22 and C-23 in **8** generating the corresponding differences in chemical shifts seen for C-22 to 24, and C-30 (Table 3). The structural confirmation for molecules **5-8** reported here provide proof of the impressive capacity of *B. braunii* race B to produce a variety of botryococcene and squalene isomers.

2.3 Experimental Section

2.3.1 General experimental procedures

A Chirascan CD Spectrometer (Applied Photophysics) was used to collect ECD data. NMR spectra were obtained using 500 and 800 MHz Bruker Avance III HD spectrometers equipped with a 5 mm inverse detection TXI probe. GC-MS analysis was performed using a Bruker 436-GC-SCION SQ Premium system, and HRMS spectra were collected on a Waters Micromass AutoSpec Ultima MS in the CI mode. HPLC was conducted using a Cosmosil 5C₁₈-AR-II column (Nacalai Tesque; 20 x 250 mm, 8 mL/min), a Develosil 60 silica column (Nomura Chemical, Japan; 20 x 250 mm, 8

mL/min), and a KANTO silver nitrate (AgNO₃) silica column (Kanto Chemical; 4.6 x 250 mm, 1 mL/min).

2.3.2 Culturing and harvesting *Botryococcus braunii* race B

Race B of *B. braunii*, Showa (Berkeley) strain¹⁶¹, was grown in modified Chu13 medium¹⁶² at 22 °C under using 13-W compact fluorescent 65K lighting with an intensity of 280 $\mu\text{mol photons/m}^2/\text{s}$. Cultures were continuously aerated with filter sterilized 2.5% CO₂ enriched air, and grown under a 12 h light:12 h dark cycle. New cultures were inoculated with 100 mL of six-week-old cultures into 750 mL of fresh medium. After five weeks, the cells were harvested by vacuum filtration using 10 μm nylon-mesh, frozen in liquid nitrogen, and stored at -80 °C.

2.3.3 Extraction and purification of race B hydrocarbons

Five-week-old algal cells stored at -80 °C were freeze dried and hydrocarbons extracted first using *n*-hexane three times to remove extracellular hydrocarbons, followed by two extractions with CH₃:MeOH (2:1) to recover intracellular hydrocarbons. Solvent from both extracts was evaporated using a rotary evaporator, resuspended in a small volume of *n*-hexane, combined, and applied to a gravity-fed silica gel column. *n*-Hexane was used as the eluent for the gravity-fed silica gel column and hydrocarbons were collected as the eluent from the column prior to the pigment front. The collected eluent was evaporated to dryness using a rotary evaporator and the sample resuspended in acetone equal to 1 volume of the hydrocarbons without solvent. Hydrocarbons were

separated using reversed phase HPLC by injecting ~100 μ L of sample onto a Cosmosil 5C₁₈-AR-II column (Nacalai Tesque; 20 x 250 mm) with 100% MeOH as the mobile phase at a flow rate of 8 mL/min and detection at 210 nm. Individual peaks containing multiple molecules were collected, and these samples were further separated using normal phase HPLC by injecting samples onto a Develosil 60 silica column (Nomura Chemical; 20 x 250 mm) with 100% *n*-hexane as the mobile phase at a flow rate of 8 mL/min and detection at 210 nm. Each peak was collected individually, and the purity and molecular weight of each sample determined by GC-MS analysis. Any samples that still contained multiple molecules were further separated by normal phase HPLC using a KANTO silver nitrate (AgNO₃) silica column (Kanto Chemical, Japan; 4.6 x 250 mm) with 99.85%:0.15% *n*-hexane:MeCN as the mobile phase at a flow rate of 1 mL/min and detection at 210 nm. Retention times of the HPLC peaks were observed as follows: C₁₈ column HPLC retention times: Peak A, ~50 min; peak B, ~52 min; peak C, ~58 min; peak D, ~64 min; peak E, ~68 min; peak F, ~71 min; peak G, ~79 min; and peak H (molecule **5**), ~86 min. Silica column HPLC retention times: C₃₂ botryococcene, ~7 min; wolficene, ~7 min; braunicene, ~8 min; trimethylsqualene, ~9 min; tetramethylsqualene, ~10 min; showacene, ~11 min; **8**, ~11 min; dimethylsqualene, ~13 min; and monomethylsqualene, ~14 min. Silver nitrate silica column HPLC retention times: C₃₄ botryococcene isomer, ~4 min; **6**, ~5 min; and **7**, ~8 min. The approximate yields of the new isolated compounds based on total hydrocarbon amount were determined as 1.2 mg, 0.30 % for **5**; 2.4 mg, 0.60 % for **6**; 0.8 mg, 0.20 % for **7**; and 0.6 mg, 0.15 % for **8**.

Cyclic C₃₃-1 botryococcene (5): colorless liquid; ECD (0.000221 M, *n*-hexane) λ_{max} ($\Delta\epsilon$) 164 (-0.3) nm, Figure 24; ^1H NMR (500 MHz and 800 MHz, CDCl_3) and ^{13}C NMR (500 MHz and 800 MHz, CDCl_3) data, Tables 2 and 5; HRCIMS m/z 452.4372 $[\text{M}]^+$ (calcd for $\text{C}_{33}\text{H}_{56}$, 452.4382).

Cyclic C₃₃-2 botryococcene (6): colorless liquid; ECD (0.000221 M, *n*-hexane) λ_{max} ($\Delta\epsilon$) 164 (-1.9) nm, Figure 24; ^1H NMR (500 MHz and 800 MHz, CDCl_3) and ^{13}C NMR (500 MHz and 800 MHz, CDCl_3) data, Table 3; HRCIMS m/z 452.4373 $[\text{M}]^+$ (calcd for $\text{C}_{33}\text{H}_{56}$, 452.4382).

Cyclic C₃₃-3 botryococcene (7): colorless liquid; ECD (0.000221 M, *n*-hexane) λ_{max} ($\Delta\epsilon$) 164 (-1.9) nm, Figure 24; ^1H NMR (500 MHz and 800 MHz, CDCl_3) and ^{13}C NMR (500 MHz and 800 MHz, CDCl_3) data, Table 3; HRCIMS m/z 452.4384 $[\text{M}]^+$ (calcd for $\text{C}_{33}\text{H}_{56}$, 452.4382).

C₃₃-2 tetramethylsqualene (8): colorless liquid; ECD (0.000221 M, *n*-hexane) λ_{max} ($\Delta\epsilon$) 164 (2.0) nm, Figure 24; ^1H NMR (500 MHz and 800 MHz, CDCl_3) and ^{13}C NMR (500 MHz and 800 MHz, CDCl_3) data, Table 3; HRCIMS m/z 452.4372 $[\text{M}]^+$ (calcd for $\text{C}_{33}\text{H}_{56}$, 452.4382).

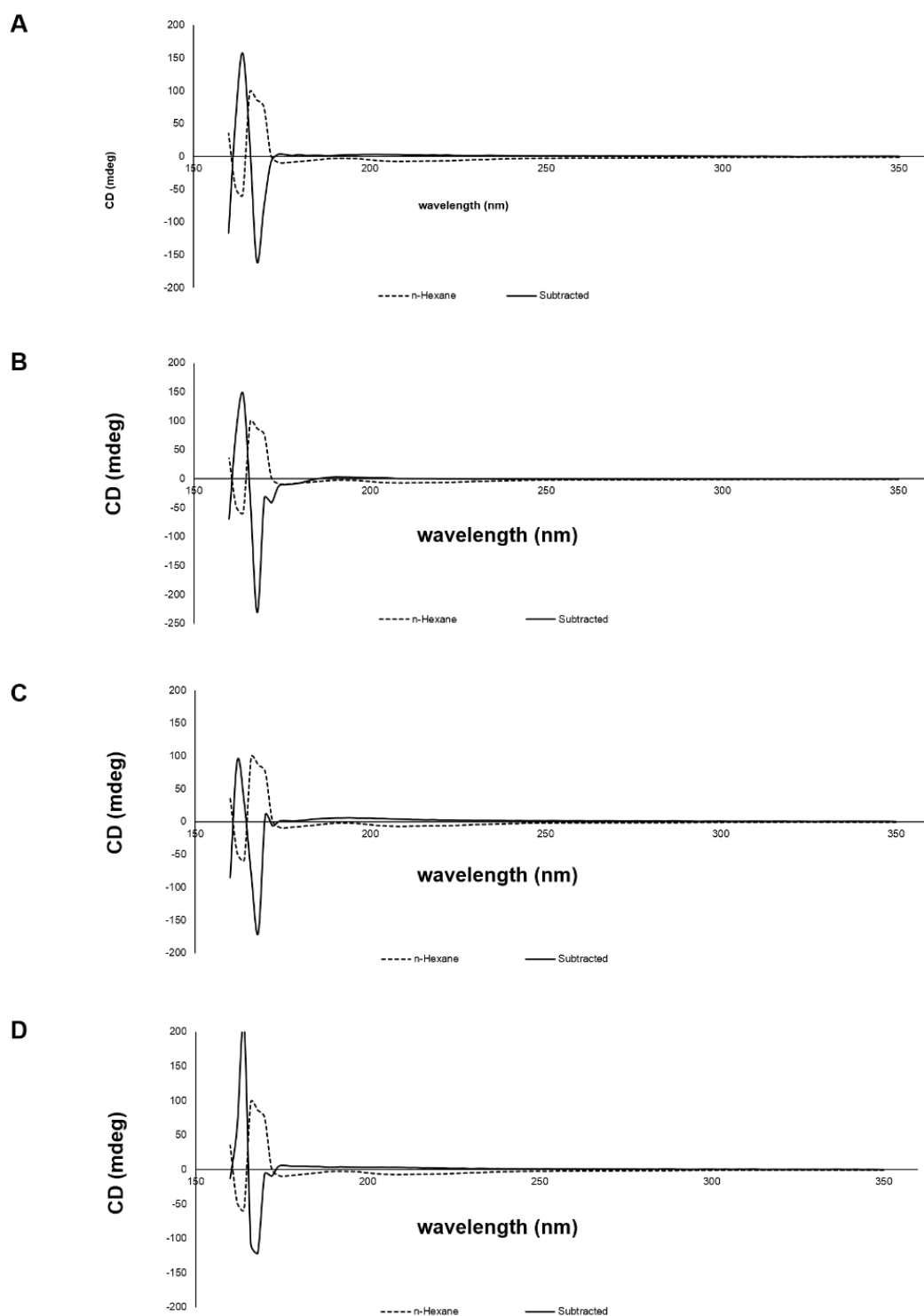


Figure 24 ECD spectra for molecules 5 (A), 6 (B), 7 (C), and 8 (D).

2.3.4 GC-MS analysis

GC-MS analysis was performed using a Bruker 436-GC-SCION SQ Premium system (Bruker Daltonics) with a 5% phenyl BR-5ms capillary column (30 m \times 0.25 mm, film thickness: 0.25 μ m) in EI (70eV) or CI (methane gas at 20 psi) mode. The GC temperature conditions were as follows: started at 220 °C and held for 1 min, then raised to 280 °C with a rate of 5 °C/min, and finally increased to 300 °C at a rate of 2 °C/min and held for 10 min. The GC injection port was held at 280 °C. The carrier gas was helium at a flow rate of 2.58 mL/min. For the MS, 250 °C and 200 °C were the temperatures of the interface and ion source, respectively.

3. RAMAN SPECTRA AND DFT CALCULATIONS FOR BOTRYOCOCCENE AND METHYLSQUALENE HYDROCARBONS FROM THE B RACE OF THE GREEN MICROALGA BOTRYOCOCCUS BRAUNII*

3.1 Background and rationale

Algae produced lipids are considered a potential renewable feedstock source to replace petroleum-based transportation fuels in order to address concerns over global climate changes due to fossil fuel derived greenhouse gas build up⁵⁴. One alga that produces large amounts of lipids in the form of liquid hydrocarbons is the green colony-forming microalga *Botryococcus braunii*⁵⁷. Hydrocarbons produced by *B. braunii* have been identified as major constituents of currently used petroleum and coal deposits indicating contribution by this organism to the formation of these deposits over geologic time^{61-63, 65, 163}. Thus, these *B. braunii* hydrocarbons can be converted into combustion engine fuels using conventional petroleum hydrocracking procedures^{20, 164, 165}.

B. braunii is divided into three chemical races, A, B, and L that are defined by the type of hydrocarbon produced⁵⁷. The B race of *B. braunii*, the focus of this study, produces 30-40 % of its dry weight as hydrocarbons in the form of two types of isoprenoid triterpenes; botryococcenes as the major accumulating hydrocarbons, and methylated squalenes as minor hydrocarbon components⁵⁷.

* Portions of this section have been reprinted with permission from: (1) Tatli M, Chun HJ, Camp Jr. CH, Li J, Cicerone MT, Shih W-C, Laane J, Devarenne TP (2017) Raman spectra and DFT calculations for botryococcene and methylsqualene hydrocarbons from the B race of the green microalga *Botryococcus braunii*. *Journal of Molecular Structure* **1147**:427-437. doi: <https://doi.org/10.1016/j.molstruc.2017.06.126>

Botryococenes and squalenes follow a very similar biosynthetic route and thus are very similar in structure with the exception of an ethene branch in the middle of the botryococcene structure that is not present in squalenes (Figure 25). Both botryococenes and methylated squalenes start as C₃₀ molecules, which are methylated at carbons 3, 7, 16, and 20 for botryococenes and carbons 3, 7, 18, and 20 for squalenes (Figure 25) to generate molecules ranging from C₃₀ - C₃₄⁵⁷. Additionally, botryococenes can be found in acyclic or cyclic forms with the ring structures on either end of the molecule, while methylsqualenes are always acyclic (Figure 25). This structural diversity has led to the identification of over twenty different botryococcene structures^{15, 57} and five different methylsqualenes ranging from monomethylsqualene to tetramethylsqualene (Figure 25)^{130, 155, 166}.

These structural differences between and among botryococenes and methylsqualenes should offer specific spectral signatures that could be used to distinguish between the different molecules. Such information could be useful for monitoring the production of specific hydrocarbons during large scale algal growth in relation to biofuel feedstock production. Additionally, molecule specific spectral information could be used for mapping cellular/colony locations of each molecule noninvasively by Raman spectroscopy. For example, our past studies analyzing the molecular structure of C₃₀, C₃₂, C₃₃, and C₃₄ botryococenes¹⁵¹, C₃₀ squalene¹⁶⁷, and the *B. braunii* L race hydrocarbons¹⁵⁰ show that Raman spectroscopy and DFT calculations can be used to identify spectral differences in the $\nu(\text{C}=\text{C})$ stretch for these molecules to differentiate between them. In the current study, we report the assignment of bond frequencies for seven botryococenes not

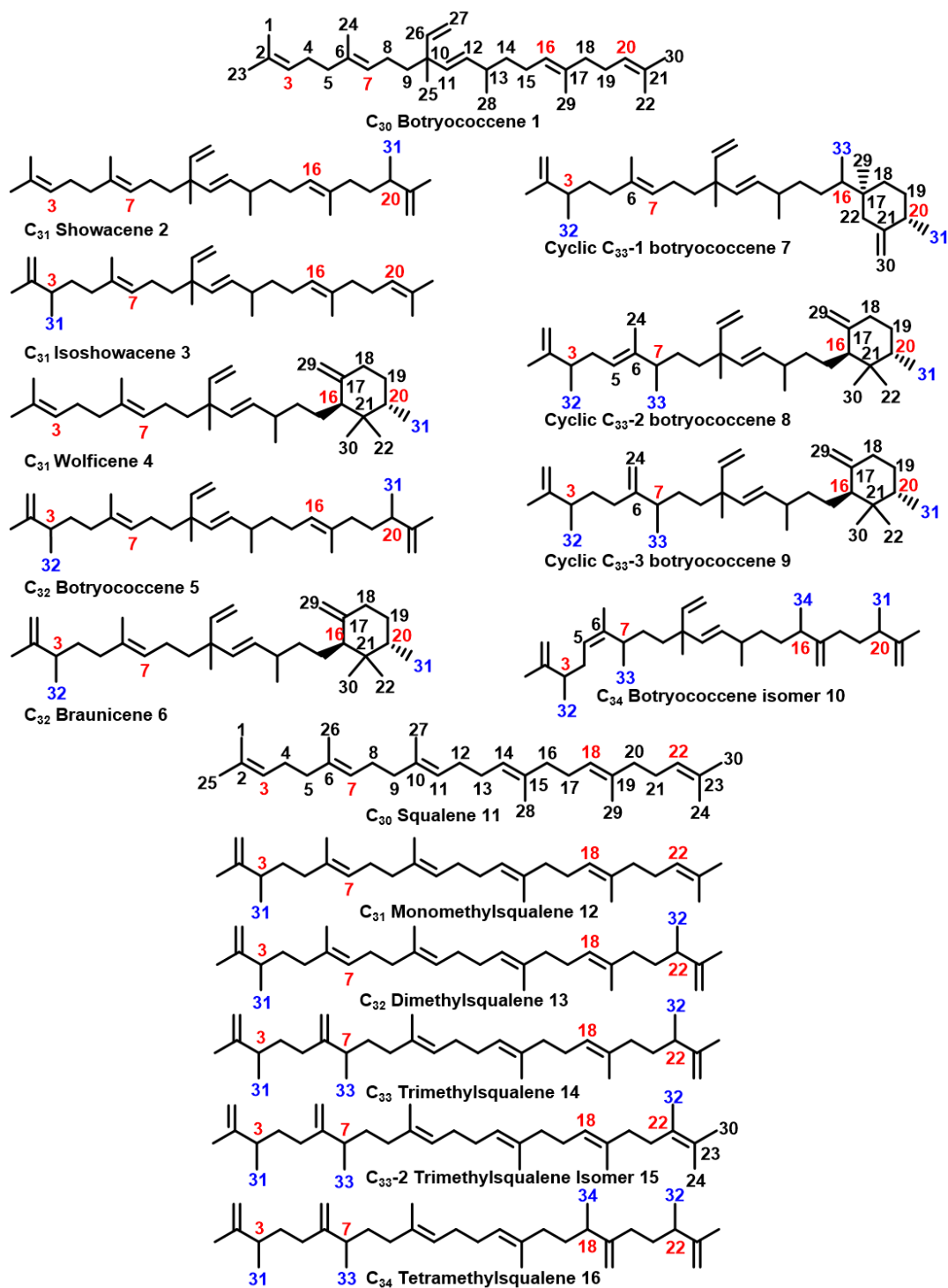


Figure 25 Hydrocarbons purified from the B race of *B. braunii* for use in Raman spectroscopy for this study.

before reported and five methylsqualenes based on experimental and DFT calculations. Additionally, ten botryococcenes are analyzed using various Raman spectroscopy techniques to show this approach can be used to distinguish between several of these molecules.

3.2 Results and discussion

3.2.1 Density function theory (DFT) computations for bond frequencies of B race hydrocarbons

The number of atoms (N) in these molecules ranges from 83 to 92, and thus the molecules possess 243 to 270 ($3N - 6$) vibrational modes of varying Raman intensity resulting in very complex spectra. The individual vibrational assignments are too lengthy to detail, but Tables 7 and 8 present a summary of the types of vibrations for each molecule. The C-H and C=C stretching vibrations are distinct but all of the other vibrational motions are strongly coupled. The computed spectra themselves are presented in Figures 26 to 30 and will be discussed below in comparison to the experimental spectra. The focus of this study is the application of Raman spectroscopy to the identification of these molecules and to distinguish between them. The most useful spectral region for this study is the $\nu(\text{C}=\text{C})$ stretch region in the 1600 to 1700 cm^{-1} range for these molecules. Other Raman bands in the lower frequency “finger print region” are also characteristic of several of these molecules and these will be discussed below.

Symbol	Vibration	Showacene 2		Wolficene 4		Braunicene 6		Cyclic C ₃₃ -1 7		Cyclic C ₃₃ -2 8		Cyclic C ₃₃ -3 9		C ₃₄ Botryococcene isomer 10	
		Wavenumber Range (cm ⁻¹)	Number of vibrations	Wavenumber Range (cm ⁻¹)	Number of vibrations	Wavenumber Range (cm ⁻¹)	Number of vibrations	Wavenumber Range (cm ⁻¹)	Number of vibrations	Wavenumber Range (cm ⁻¹)	Number of vibrations	Wavenumber Range (cm ⁻¹)	Number of vibrations	Wavenumber Range (cm ⁻¹)	Number of vibrations
$\nu(\text{CH}_3)$	CH ₃ stretch	2886-3045	24	2898-3025	24	2885-3025	24	2886-3044	24	2898-3027	27	2887-3019	24	2871-3045	27
$\delta(\text{CH}_3)$	CH ₃ deformation	1253-1463	24	1345-1474	24	1352-1474	24	1325-1477	24	1322-1474	27	1331-1472	24	1308-1466	27
$\rho(\text{CH}_3)$	CH ₃ rock	753-1222	16	792-1335	16	555-1329	16	885-1309	16	646-1322	18	477-1306	16	509-1300	18
$\tau(\text{CH}_3)$	CH ₃ torsion	100-285	8	108-319	8	102-335	8	99-290	8	150-316	9	138-315	8	145-302	9
$\nu(\text{CH}_2)$	CH ₂ stretch	2872-3117	20	2898-3116	20	2885-3116	22	2886-3116	24	2898-3118	20	2887-3125	24	2871-3119	22
$\delta(\text{CH}_2)$	CH ₂ deformation	1406-1463	10	1411-1474	10	1412-1474	11	1407-1477	12	1366-1474	10	1409-1472	12	1372-1466	11
$\omega(\text{CH}_2)$	CH ₂ wag	916-1377	10	1148-1384	10	907-1363	11	903-1380	12	909-1368	10	909-1366	12	912-1377	11
$t(\text{CH}_2)$	CH ₂ twist	693-1318	10	729-1332	10	692-1321	11	558-1321	12	691-1335	10	692-1331	12	693-1307	11
$\rho(\text{CH}_2)$	CH ₂ rock	737-1052	10	729-1033	10	692-1027	11	647-997	12	691-997	10	589-1047	12	536-1013	11
$\nu(\text{CH})$	=C-H stretch/-C-H stretch	2872-3117	8	2898-3044	8	2885-3115	8	2886-3044	8	2898-3027	9	2887-3024	8	2871-3117	9
$\omega_i(\text{CH})$	CH wag (in-plane)	1026-1406	8	1064-1416	8	1078-1378	8	995-1468	8	1061-1439	9	986-1350	8	1022-1408	9
$\omega_o(\text{CH})$	CH wag (out-of-plane)	818-1303	8	813-1345	8	852-1357	8	856-1322	8	857-1370	9	857-1020	8	693-1320	9
$\nu(\text{Ring})$	Six-membered ring stretch	-	-	741-1349	6	727-1352	6	736-1414	6	792-1415	6	692-1415	6	-	-
$b(\text{Ring})$	Six-membered ring bending (in-plane)	-	-	408-486	3	407-692	3	407-647	3	411-646	3	337-594	3	-	-
$b(\text{Ring})$	Six-membered ring bending (out-of-plane)	-	-	466-693	3	349-709	3	358-707	3	349-691	3	285-647	3	-	-
$\nu(\text{C}=\text{C})$	C=C stretch	1643-1679	6	1648-1679	5	1650-1670	5	1649-1671	5	1649-1664	5	1649-1663	5	1644-1667	6
$\nu(\text{C}-\text{C})$	C-C stretch	718-1406	24	713-1411	20	718-1407	21	713-1414	22	712-1415	22	692-1415	22	711-1416	27
$b(\text{C}-\text{C}-\text{C})$	C-C-C angle bend	135-632	11	283-693	8	173-709	9	232-707	10	192-691	9	186-714	11	145-629	14
$b(\text{C}=\text{C}-\text{C})$	C=C-C angle bend	288-632	10	388-693	5	248-709	6	301-707	6	257-691	6	271-714	6	255-629	7
$\omega_o(\text{C}-\text{CH}_3)$	C-CH ₃ wag (out-of-plane)	318-693	8	453-532	8	299-555	8	313-554	8	287-591	9	271-594	8	353-629	9
$\omega_i(\text{C}-\text{CH}_3)$	C-CH ₃ wag (in-plane)	272-590	8	297-444	8	231-537	8	266-502	8	234-589	9	246-542	8	239-338	9
$\tau(\text{C}-\text{C})$	Internal rotation (C-C)/Skeletal	9-285	16	5-319	18	8-335	19	8-290	20	10-316	19	12-315	20	9-302	18
$\tau(\text{C}=\text{C})$	Internal rotation (C=C)/Skeletal	52-285	4	50-319	3	53-335	2	55-290	2	59-316	2	55-315	1	51-302	6
Total			243		243		252		261		261		261		270

Table 7 Vibrations of *B. braunii* race B botryococcenes.

Symbol	Vibration	Monomethyl Squalene 12		Dimethyl Squalene 13		Trimethyl Squalene 14		Trimethylsqualene isomer 15		Tetramethyl Squalene 16	
		Wavenumber Range (cm ⁻¹)	Number of vibrations	Wavenumber Range (cm ⁻¹)	Number of vibrations	Wavenumber Range (cm ⁻¹)	Number of vibrations	Wavenumber Range (cm ⁻¹)	Number of vibrations	Wavenumber Range (cm ⁻¹)	Number of vibrations
$\nu(\text{CH}_3)$	CH ₃ stretch	2885-3025	24	2887-3025	24	2886-3043	24	2886-3045	27	2886-3024	24
$\delta(\text{CH}_3)$	CH ₃ deformation	1315-1463	24	1315-1462	24	1312-1464	24	1312-1476	27	1341-1463	24
$\rho(\text{CH}_3)$	CH ₃ rock	755-1275	16	735-1221	16	715-1306	16	766-1237	18	575-1307	16
$\tau(\text{CH}_3)$	CH ₃ torsion	72-232	8	74-229	8	87-237	8	70-267	9	87-229	8
$\nu(\text{CH}_2)$	CH ₂ stretch	2885-3116	22	2887-3118	24	2886-3125	26	2886-3125	24	2886-3125	28
$\delta(\text{CH}_2)$	CH ₂ deformation	1372-1463	11	1372-1462	12	1371-1464	13	1372-1476	12	1406-1463	14
$\omega(\text{CH}_2)$	CH ₂ wag	915-1367	11	914-1407	12	914-1377	13	914-1368	12	910-1372	14
$t(\text{CH}_2)$	CH ₂ twist	717-1315	11	716-1304	12	715-1306	13	715-1306	12	713-1307	14
$\rho(\text{CH}_2)$	CH ₂ rock	735-995	11	735-1031	12	734-988	13	736-803	12	736-942	14
$\nu(\text{CH})$	=C-H stretch/-C-H stretch	2885-3025	6	2887-3025	6	2886-3009	6	2886-3025	5	2886-3024	6
$\omega_l(\text{CH})$	CH wag (in-plane)	1018-1458	6	1031-1453	6	978-1444	6	1001-1408	5	991-1413	6
$\omega_o(\text{CH})$	CH wag (out-of-plane)	748-1302	6	735-1304	6	573-1306	6	782-1306	5	872-1314	6
$\nu(\text{C}=\text{C})$	C=C stretch	1654-1679	6	1654-1672	6	1652-1672	6	1652-1672	6	1652-1672	6
$\nu(\text{C}-\text{C})$	C-C stretch	717-1408	24	716-1407	25	715-1409	26	707-1412	26	713-1411	27
$b(\text{C}-\text{C}-\text{C})$	C-C-C angle bend	160-596	11	118-598	12	120-595	14	120-597	13	123-595	16
$b(\text{C}=\text{C}-\text{C})$	C=C-C angle bend	232-596	9	260-598	8	237-595	7	292-597	8	277-595	8
$\omega_o(\text{C}-\text{CH}_3)$	C-CH ₃ wag (out-of-plane)	456-583	8	305-580	8	307-586	8	267-597	9	471-579	8
$\omega_l(\text{C}-\text{CH}_3)$	C-CH ₃ wag (in-plane)	330-577	8	260-598	8	237-595	8	247-591	9	238-458	8
$\tau(\text{C}-\text{C})$	Internal rotation (C-C)/Skeletal	6-232	16	5-229	17	7-237	18	6-267	17	6-229	19
$\tau(\text{C}=\text{C})$	Internal rotation (C=C)/Skeletal	53-232	5	54-229	6	52-237	6	53-267	5	42-229	4
Total			243		252		261		261		270

Table 8 Vibrations of *B. braunii* race B squalenes.

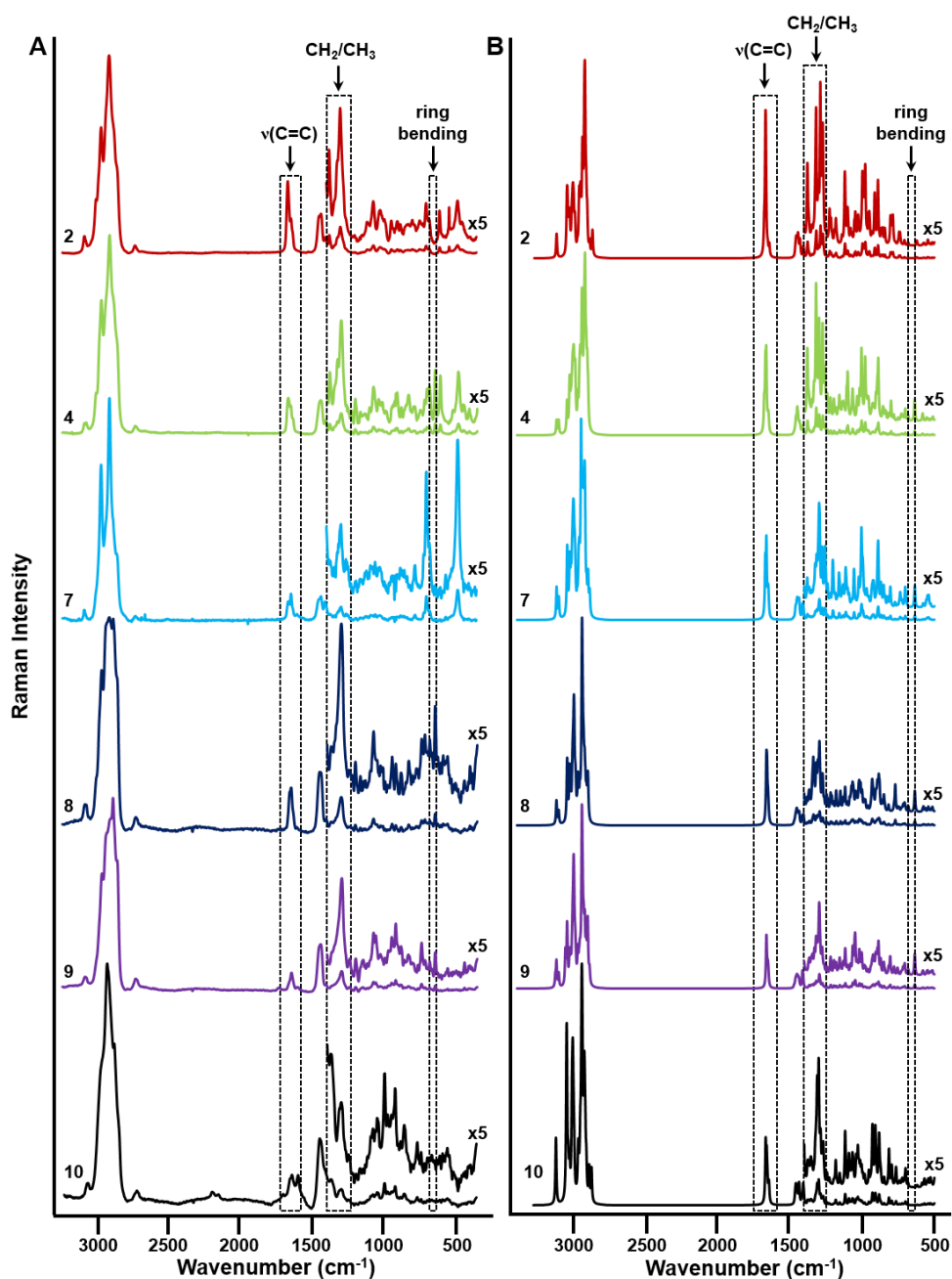


Figure 26 Experimental and calculated spectra for botryococenes. (A) The experimental Raman spectra for the indicated botryococenes. (B) The calculated Raman spectra for the indicated botryococenes. Spectral regions in dashed line boxes indicate the $\nu(\text{C}=\text{C})$ stretching analyzed in Figure 28A and B; CH_2/CH_3 bending analyzed in Figure 29A and B, and cyclohexane b(Ring) bending analyzed in Figure 30A and B.

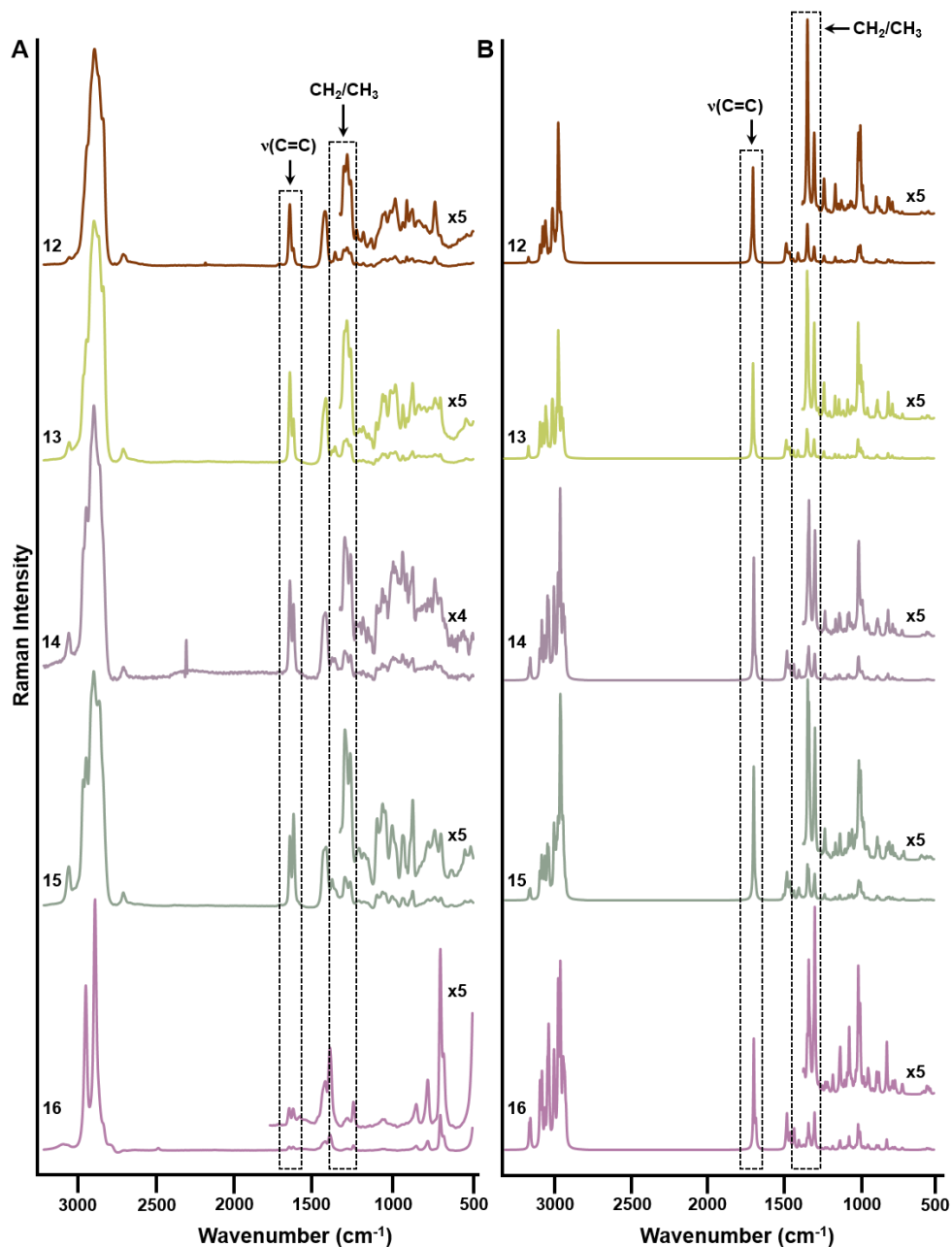


Figure 27 Experimental and calculated spectra for methylsqualenes. (A) The experimental Raman spectra for the indicated methylsqualenes. (B) The calculated Raman spectra for the indicated methylsqualenes. Spectral regions in dashed line boxes indicate the $\nu(\text{C}=\text{C})$ stretching analyzed in Figure 28C and D and CH_2/CH_3 bending analyzed in Figure 29C and D.

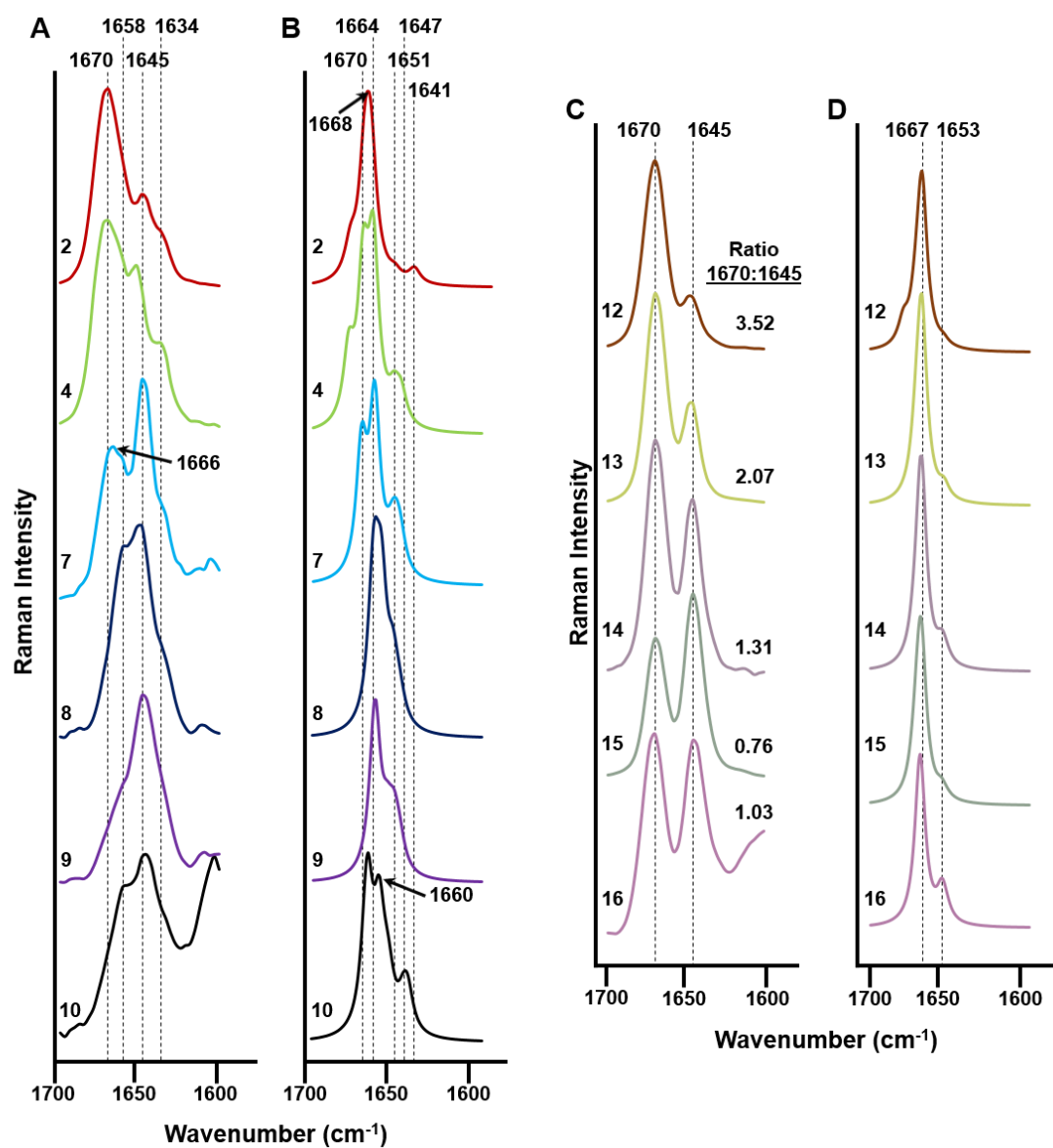


Figure 28 The experimental and calculated $\nu(\text{C}=\text{C})$ stretching Raman spectral regions for botryococcenes and methylsqualenes. (A) Experimental spectral region for the indicated botryococcenes. (B) Calculated spectral region for the indicated botryococcenes. (C) Experimental spectral region for the indicated methylsqualenes. (D) Calculated spectral region for methylsqualenes.

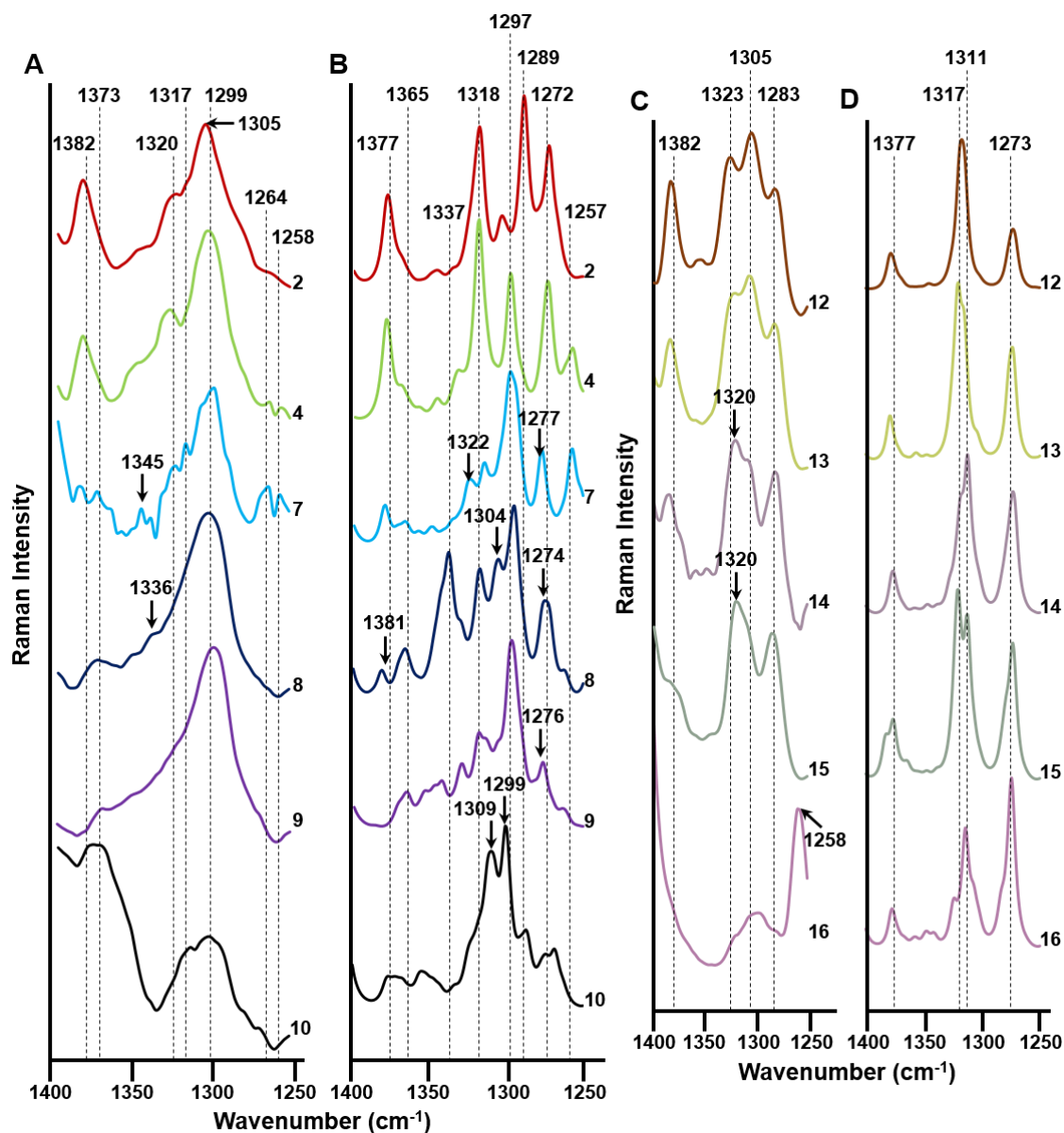


Figure 29 The experimental and calculated Raman spectral regions containing $\delta(\text{CH}_3)$ deformation, $\rho(\text{CH}_3)$ rock, $\omega(\text{CH}_2)$ wag, $t(\text{CH}_2)$ twist, $\omega_i(\text{CH})$ wag (in-plane), and $\nu(\text{C-C})$ stretch for botryococcenes and methylsqualenes. (A) Experimental spectral region for the indicated botryococcenes. (B) Calculated spectral region for the indicated botryococcenes. (C) Experimental spectra region for the indicated methylsqualenes. (D) Calculated spectral region for the indicated methylsqualenes.

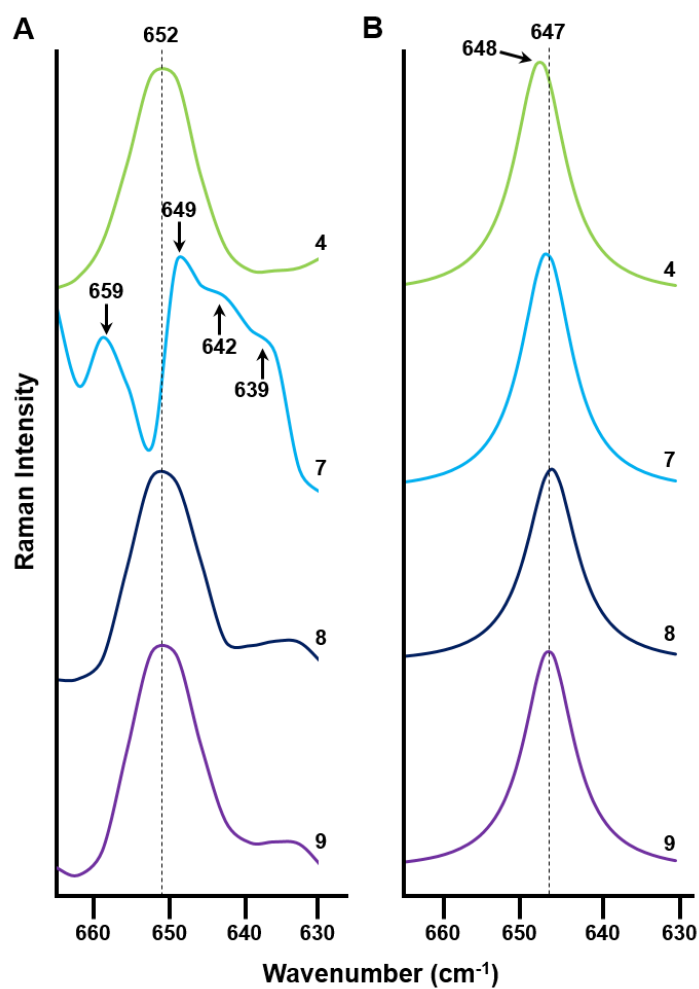


Figure 30 The experimental and calculated cyclohexane b(Ring) bending Raman spectral regions for botryococenes. (A) Experimental spectral region for the indicated botryococenes. (B) Calculated spectral region for the indicated botryococenes.

3.2.2 Comparison of experimental and calculated full Raman spectra for B race hydrocarbons

Based on our past Raman spectroscopy studies on *B. braunii* hydrocarbons from the B and L races^{150, 151} as well as non-methylated C₃₀ squalene¹⁶⁷, we have found several spectral regions that can be used to distinguish between these structurally similar molecules. Specifically, these regions are the $\nu(\text{C}=\text{C})$ stretching region between 1600 and 1700 cm^{-1} , and the CH₂/CH₃ bending, $\omega_i(\text{CH})$ wag (in-plane) and $\nu(\text{C}-\text{C})$ stretching region between 1250 and 1400 cm^{-1} (this region will be referred to as CH₂/CH₃ bending for simplicity). Additionally, based on the DFT calculations shown in Table 7 the cyclohexane b(Ring) bending (in-plane and out-of-plane) region between 630 – 665 cm^{-1} should provide a specific spectral signature to clearly distinguish the cyclic botryococcenes from the acyclic botryococcenes. These spectral regions were analyzed experimentally by subjecting two acyclic botryococcenes, C₃₁ botryococcene **2** and C₃₄ botryococcene isomer **10**, and 4 cyclic botryococcenes, C₃₁ wolficene **4**, C₃₃-1 cyclic botryococcene **7**, C₃₃-2 cyclic botryococcene **8**, C₃₃-3 cyclic botryococcene **9**, to line-scan Raman spectroscopy. Additionally, the computed spectra were compared to the experimental spectra. None of these botryococcenes have been studied by Raman spectroscopy previously.

Comparing the experimental (Figure 26A) and computed (Figure 26B) spectra for each botryococcene shows a strong correlation between the two types of spectra, with high similarity in the $\nu(\text{C}=\text{C})$ stretching and CH₂/CH₃ bending regions for all molecules. Comparing the spectra between acyclic and cyclic botryococcenes, the cyclohexane

b(Ring) bending region shows an obvious Raman band in the cyclic botryococenes in the 630 - 665 cm^{-1} region that is lacking from the acyclic botryococenes (Figure 26A and B). In addition, there are some differences between each molecule in the $\nu(\text{C}=\text{C})$ stretching and CH_2/CH_3 bending regions. For example, the intensity of the Raman bands varies between molecules in both of these regions, and some molecules have a single band while others have two bands in these regions (Figure 26A and B). These differences will be discussed in detail below.

The same comparative analysis was carried out for the methylsqualenes. As with the botryococenes, the experimental (Figure 27A) and computational (Figure 27B) spectra show a strong correlation for each methylsqualene. A comparison of the spectra between methylsqualenes indicates that in the CH_2/CH_3 bending region there is a high degree of similarity with two or three Raman bands for each molecule (Figure 27A and B). However, while each methylsqualene has a Raman band doublet in the $\nu(\text{C}=\text{C})$ stretching region, the intensity of each band differs for each molecule (Figure 27A). The ratio between these two bands should be a diagnostic signal for each molecule. These differences will be discussed in detail below.

3.2.3 Raman spectral regions that can distinguish between hydrocarbons

A detailed analysis of the $\nu(\text{C}=\text{C})$ stretching, CH_2/CH_3 bending, and cyclohexane b(Ring) bending regions for all hydrocarbons analyzed indicates these regions can be used to distinguish between several of these molecules. For the $\nu(\text{C}=\text{C})$ stretching region (1600 – 1700 cm^{-1}) of the botryococenes analyzed there is a clear distinction between the

predominant Raman band for **2** and **4** compared to **7**, **8**, and **9**. For **2** and **4**, there is a strong Raman band in the 1670 cm^{-1} region (Figure 28A). In **7** there is a Raman band of moderate intensity in the 1670 cm^{-1} region (at 1666 cm^{-1}), but the most prominent band is at 1645 cm^{-1} . The spectra for **8**, **9**, **10** show no Raman band in the 1670 cm^{-1} region, rather the strongest Raman band is shifted to 1645 cm^{-1} (Figure 28A). Additionally, bands at 1634 cm^{-1} in **2**, **4**, **7** and 1658 cm^{-1} in **8**, **9**, **10** can also be used to distinguish between these molecules (Figure 28A). These experimental observations are well supported by the Raman bands seen in the calculated spectra (Figure 28B).

The $\nu(\text{C}=\text{C})$ stretching region for the experimental analysis of methylsqualenes shows two Raman bands for each molecule at 1670 cm^{-1} and 1645 cm^{-1} (Figure 28C). It can be seen that as the methylation status of the squalenes increases from monomethylation (**12**) to tetramethylation (**16**) the intensity of the 1670 cm^{-1} band remains roughly the same while the 1645 cm^{-1} band increases (Figure 28C). The difference between these molecules is the position of the $\text{C}=\text{C}$ bonds. As the squalenes are methylated the backbone $\text{C}=\text{C}$ bonds at positions 2, 6, 18, and 22 migrate to exomethylene groups at carbon positions 1, 26, 29, and 24 (Figure 25). Thus, as the squalenes are methylated the number of backbone $\text{C}=\text{C}$ bonds decreases and the number of exomethylene $\text{C}=\text{C}$ bonds increases. From this information the 1670 cm^{-1} Raman band can be assigned to the backbone $\text{C}=\text{C}$ bonds and the 1645 cm^{-1} band can be assigned to the exomethylene $\text{C}=\text{C}$ bonds. The calculated 1670:1645 ratio for the intensities between these two bands shows there is a clear distinction in this value for each methylsqualene that could be used to distinguish between each molecule (Figure 28C). Interestingly, these experimental

observations are not well supported by the calculated spectra for the methylsqualenes, which shows a single predominant Raman band at 1667 cm^{-1} for each methylsqualene (Figure 28D).

As with our previous studies^{150, 151, 167}, for all molecules analyzed the DFT calculations predict a number of $\nu(\text{C}=\text{C})$ stretching frequencies equal to the number of $\text{C}=\text{C}$ bonds in each molecule (Table 7 and 8). However, the experimental spectra do not show all of these $\nu(\text{C}=\text{C})$ stretching Raman bands (Figure 28A and C). This is due to the fact that several $\text{C}=\text{C}$ stretching bands are very close in frequency, thus generating a few predominant bands caused by spectral overlap.

For the CH_2/CH_3 bending region ($1250 - 1400\text{ cm}^{-1}$), the experimental spectra have many similarities for the botryococcene analyzed with a main, broad Raman band centered around 1299 cm^{-1} (Figure 29A). However, a band at 1382 cm^{-1} can be used to distinguish **2**, **4**, and **7** from the other botryococcenes (Figure 29A). As was seen for the $\nu(\text{C}=\text{C})$ stretching frequencies, the calculated spectra (Figure 29B) and the calculated vibrations (Table 7) predict many CH_2/CH_3 vibrations in the $1250 - 1400\text{ cm}^{-1}$ region that area not clearly defined in the experimental spectra. Again, this is likely due to spectral overlap of vibrations within very close frequencies. Similarly, the CH_2/CH_3 bending region of the experimental spectra for the methylsqualenes shows a high degree of similarity for each molecule with several predominant Raman bands between 1283 cm^{-1} and 1323 cm^{-1} (Figure 29C). A band at 1282 cm^{-1} can distinguish **12**, **13**, and **14** from the other methylsqualenes while **16** has a band at 1258 cm^{-1} not seen in the other molecules

(Figure 29C). The calculated spectra confirm the main Raman bands in the 1273 - 1323 cm^{-1} region (Figure 29D).

The ring structures found in botryococcenes **4**, **7**, **8**, and **9**, should offer unique vibrations that can distinguish these molecules from the acyclic botryococcenes. In the experimental spectra, the 630 – 665 cm^{-1} region for the cyclohexane b(Ring) bending vibrations shows a Raman band at 652 cm^{-1} that is found in **4**, **8**, and **9**, while the main band for **7** is broad with a peak at 649 cm^{-1} and shoulders at 642 cm^{-1} and 639 cm^{-1} (Figure 30A). An additional band at 659 cm^{-1} can also distinguish **7** from the other cyclic botryococcenes (Figure 30A). These bands are not seen for the acyclic botryococcenes **2** and **10** (Figure 26). Molecules **4**, **8**, and **9** all have the same ring structure with cyclization at carbon 16 (Figure 25), and thus it is not surprising that the spectra for these molecules within this region are nearly identical (Figure 30A). For molecule **7**, the ring is cyclized at carbon 17 (Figure 25) generating the spectral differences seen. It should be noted that the intensity of the bands shown for **7** in Figure 30A have been scaled in order to have the bands stand out. However, in absolute values the intensity of these bands for **7** are much lower than that observed for **4**, **8**, and **9** within this ring bending region (Figure 26). The calculated spectra for the cyclic botryococcenes largely agree with the experimental spectra except for molecule **7**, which shows a single Raman band at 647 cm^{-1} (Figure 30B).

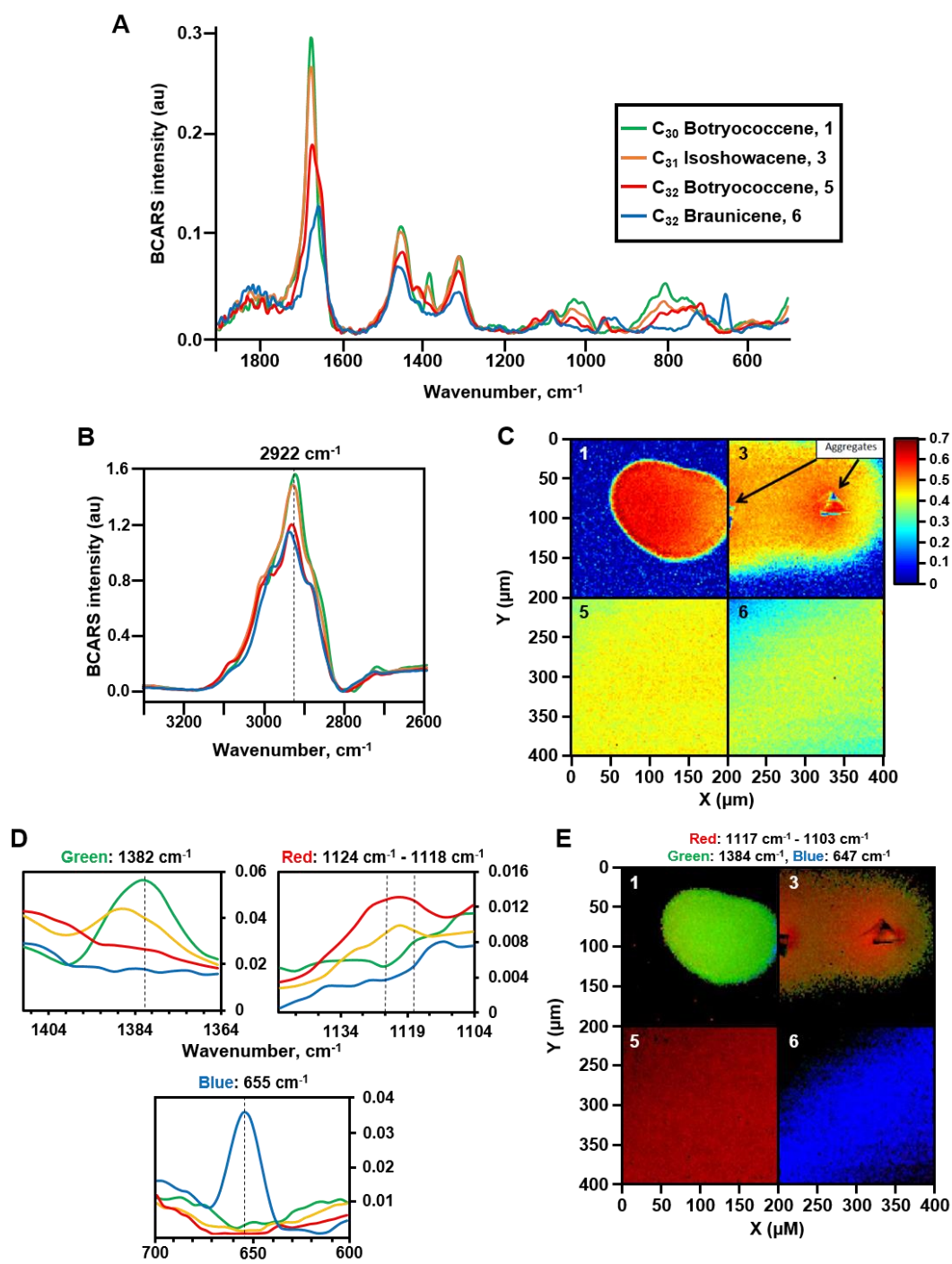
3.2.4 Distinguishing between botryococcene isomers using broadband CARS

We next attempted to distinguish between several botryococcenes using broadband coherent anti-Stokes Raman spectroscopy (BCARS). The botryococcenes used, C₃₀

botryococcene (**1**), C₃₁ isoshowacene (**3**), C₃₂ botryococcene (**5**), and C₃₂ braunicene (**6**), were chosen because analysis of these molecules should indicate if Raman spectroscopy can distinguish between the structurally similar acyclic isomers and one cyclic isomer. The BCARS spectra for these molecules in the fingerprint region (500 – 1900 cm⁻¹) indicate a high degree of similarity between the molecules (Figure 31A). However, several regions will be highlighted that can be used to distinguish each of these molecules.

First, each molecule was analyzed to confirm detection using the BCARS system by analyzing the intensity of the general C-H stretch for each molecule. Each botryococcene was spotted on a glass slide, analyzed by BCARS, and the intensity of the C-H stretching signal at 2922 cm⁻¹ (Figure 31B) mapped onto each molecule on the glass slide (Figure 31C). The intensity of the 2922 cm⁻¹ C-H stretching signal can be easily detected for each molecule indicating BCARS can be used for analysis of these molecules. Additionally, it can be seen that each molecule has different surface tension characteristics. The C₃₀ botryococcene (**1**) stayed as a tight oil drop on the glass slide while each of the other botryococcenes spread out more on the surface of the glass slide and were no longer contained in a tight oil drop such as for C₃₂ braunicene (**6**) (Figure 31C). Next, three spectral regions that should offer specificity to several of these botryococcenes were analyzed. These three regions were identified from the fingerprint region of each spectrum and then each botryococcene was analyzed for the detection of each signal. The first region analyzed was a Raman band at 1382 cm⁻¹ that is mainly found in the spectrum for C₃₀ botryococcene (**1**) (Figure 31D).

Figure 31 Spectral differentiation of C₃₀ botryococcene, C₃₁ isoshowacene, C₃₂ botryococcene, and C₃₂ braunicene by BCARS. (A) The experimental BCARS spectra for the indicated botryococcenes. (B) The BCARS C-H stretch spectra region for the indicated botryococcenes. Dotted vertical line marks wavenumber 2935 cm⁻¹ analyzed in C. The colored line indicated for each botryococcene applies to each spectra shown throughout the figure. (C) Detection of the indicated botryococcenes based on detection of the 2935 cm⁻¹ wavenumber. Scale of intensity detection is shown on the right side of this panel. (D) Specific BCARS spectral signatures for the indicated botryococcenes used for analysis in E. Vertical dotted lines indicate specific wavenumber or wavenumber region analyzed. (E) Each botryococcene spotted on a glass slide was scanned for each BCARS spectral signature shown in D, indicating that each molecule can be distinguished from each other by this method.



Each botryococcene was analyzed for a signal from the 1382 cm^{-1} band, and if it was detected the signal was assigned a green color for mapping to the hydrocarbons on the glass slide. It can be seen that the 1382 cm^{-1} signal was mainly found in C_{30} botryococcene (**1**) with lower detection in C_{31} isoshowacene (**3**), and no detection in either C_{32} botryococcene (**5**) or C_{32} braunicene (**6**) (Figure 31E). The second spectral region analyzed was a Raman band between $1124 - 1118\text{ cm}^{-1}$ that is mostly detected in C_{32} botryococcene (**5**) (Figure 31D). If the signal for the $1124 - 1118\text{ cm}^{-1}$ region was detected in any molecule it was given a red color for mapping to the hydrocarbons on the glass slide. This signal was mainly detected in C_{32} botryococcene (**5**) with lower detection in C_{31} isoshowacene (**3**), and no detection in C_{30} botryococcene (**1**) or C_{32} braunicene (**6**) (Figure 31E). Interestingly, C_{32} botryococcene (**5**) is mainly shown as an orange color which would be due to mixed detection of the green 1382 cm^{-1} signal and the red $1124 - 1118\text{ cm}^{-1}$ signal (Figure 31E). The third spectral region analyzed was a Raman band at 655 cm^{-1} that is specific for the cyclohexane b(Ring) bending in C_{32} braunicene (**6**) (Figure 31D). Detection of the 655 cm^{-1} signal in any molecule was assigned a blue color, and this signal is only detected in C_{32} braunicene (**6**) (Figure 31E).

3.3 Experimental Methods

3.3.1 Culturing of *B. braunii*

B. braunii race B, Showa (Berkeley) strain¹⁶¹, was grown at 22°C under continuous aeration with filter-sterilized air enriched with 2.5% CO_2 . A modified Chu 13 media was used for culturing¹⁶⁸. Cultures were exposed to a light intensity of $120\text{ }\mu\text{mole photons} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$.

$2 \cdot \text{s}^{-1}$ under a 12:12h light:dark cycle. Algal cultures were subcultured from a six-week-old culture using 100 ml of floating colonies for inoculation into 750 ml of fresh medium.

3.3.2 Hydrocarbon purification

B. braunii B race hydrocarbons were purified as previously described^{127, 166, 169}. Briefly, five-week-old freeze-dried algal colonies were treated three times with *n*-hexane to extract extracellular hydrocarbons, followed by two treatments with CHCl_3 :MeOH (2:1) to extract intracellular hydrocarbons. Solvents from both extracts were removed using a rotary evaporator, the hydrocarbons resuspended in a minimal amount of *n*-hexane, the extracts combined, and applied to a gravity-fed silica gel column using *n*-hexane as the mobile phase. The eluent collected prior to the pigment front was pooled as the total hydrocarbon fraction, the solvent removed using a rotary evaporator, the hydrocarbons resuspended in a small amount of *n*-hexane, and a crude separation of the hydrocarbons into eight peaks was done by reversed-phase HPLC using a Cosmosil 5C₁₈-AR-II column (20 mm x 250 mm) with 100% MeOH as the mobile phase at a flow rate of 8 ml/min and detection at 210 nm. Each peak from this separation contained multiple molecules, which were further separated using normal-phase HPLC using a Develosil 60 silica column (20 mm x 250 mm) with 100% *n*-hexane as the mobile phase at a flow rate of 8 ml/min and detection at 210 nm. Peaks that still contained multiple molecules were further separated by normal-phase HPLC using a KANTO silver nitrate (AgNO_3) silica column (4.6 x 250 mm) with 99.85%:0.15% *n*-hexane:acetonitrile as the mobile phase at a flow rate of 1ml/min and detection at 210 nm. Purity to at least 95% and molecular mass of each

sample was determined by GC-MS analysis, and molecule structure identity confirmed by NMR analysis¹⁶⁶.

3.3.3 Raman spectroscopy

Experimental Raman spectroscopy data was collected using line-scan and broadband coherent anti-Stokes Raman spectroscopy (BCARS). For all spectroscopy, each hydrocarbon was dissolved in *n*-hexane and ~100 µg of each sample was pipetted onto a glass microscope slide one drop at a time, allowing the *n*-hexane to evaporate after applying each drop. Once all *n*-hexane was evaporated the spectra for each hydrocarbon was collected.

Line-scan Raman spectroscopy analysis was carried out on the following purified hydrocarbons: C₃₁ showacene **2**, C₃₁ wolficene **4**, cyclic C₃₃-1 botryococcene **7**, cyclic C₃₃-2 botryococcene **8**, and cyclic C₃₃-3 botryococcene **9**, C₃₄ botryococcene isomer **10**, monomethylsqualene **12**, dimethylsqualene **13**, trimethylsqualene **14**, trimethylsqualene isomer **15**, and tetramethylsqualene **16**. For data acquisition, line-scan Raman was carried out as previously reported¹⁴³. Briefly, the output of a continuous-wave 532 nm laser (Spectra-Physics Millennia Xs) was focused on the sample as a uniform 133 × 1 µm² line. Epi-Raman was collected by a 60X objective (Olympus, N.A=1.2), imaged at the entrance slit of a spectrograph (Acton 2300) and recorded by a CCD detector (PIXIS 400BR). All spectra were collected at 35 mW of power using 4 min exposure. A fifth order polynomial least-squares fitting algorithm was used for spectral baseline correction.

BCARS analysis was carried out on the following purified hydrocarbons: C₃₀ botryococcene **1**, C₃₁ isoshowacene **3**, C₃₂ botryococcene **5**, and C₃₂ braunicene **6**. Each hydrocarbon was separately dried on a glass microscope slide within a gasket, then covered with a coverslip. BCARS-generated Raman spectra were collected on a 100 x 100 pixel grid with 2 micron/pixel spacing. Described previously¹⁷⁰, the BCARS instrument employs a 40 MHz dual-output fiber laser, with a 16 mW, 770 nm probe output of pulse duration 3.4 ps, and a 10 mW, 900 nm-centered supercontinuum output with 16 fs duration. This system provides Raman spectra at each image pixel with a pixel dwell time of 3.5 ms.

3.3.4 Theoretical computations

The vibrational frequencies and Raman intensities of all twelve molecules were calculated using the Gaussian 09 program, Revision A.02¹⁷¹. The density functional theory (DFT) B3LYP method^{172, 173} was utilized with a cc-pVTZ basis set for the computations. The calculations apply for the vapor-phase of these molecules, but little difference between these values and those for the liquid state is expected except for the low frequency modes below about 200 cm⁻¹. The standard Gaussian 09 protocol with the default values for bandwidths was used for producing the computed spectra in the figures. The Semichem AMPAC/AGUI program¹⁷⁴ was utilized to visualize the vibrational modes. A scaling factor of 0.969 was used for all frequencies above 1000 cm⁻¹ and 0.985 was used for the lower frequencies. These scaling factors values were found to be best suited for correlating the experimental and theoretical frequency values in our previous studies^{150, 151, 167, 175}.

Those previous studies also showed that on the whole the vibrational frequencies changed insignificantly for the different conformational orientations of these types of molecules.

4. POLYSACCHARIDE ASSOCIATED PROTEIN (PSAP) FROM THE GREEN MICROALGA *BOTRYOCOCCUS BRAUNII* IS A UNIQUE EXTRACELLULAR MATRIX HYDROXYPROLINE-RICH GLYCOPROTEIN

4.1 Background and rationale

An extracellular matrix (ECM) is a common feature in multicellular eukaryotic organisms and offers a range of functions from structural support to cell-to-cell communication^{176, 177}. The ECM in algae is used to hold cells into a multicellular organization that can contain differentiated cell types or all non-differentiated cells^{85, 87}. This multicellularity due to ECM development in green algae is thought to have been the first step in the evolution of land plants^{178, 179}. For colonial green microalgae, the ECM is integral to cell survival by offering a scaffold for holding cells into a colony, helping to regulate cell development and differentiation, and acting as a reservoir for molecules secreted from the cells^{57, 88, 89, 180}.

In terms of colony formation, some green microalgae form interconnected chains of cells through an ECM consisting mainly of the cell wall, while in others separate cells are implanted into an ECM made up of the cell wall plus a complex structure extending beyond the cell wall^{57, 58, 86, 87}.

Portions of this section have been reprinted with permission from: (1) Tatli M, Ishihara M, Heiss C, Browne DR, Dangott LJ, Vitha S, Azadi P, Devarenne TP (2018) Polysaccharide associated protein (PSAP) from the green microalga *Botryococcus braunii* is a unique extracellular matrix hydroxyproline-rich glycoprotein. *Algal Research* **29**:92-103. doi: <https://doi.org/10.1016/j.algal.2017.11.018>

In typical descriptions, the ECM consists of all material outside the cell membrane, including the cell wall⁸⁷. For simplicity, our discussions of the ECM in this study will refer only to material exterior to the cell wall.

The ECM in green microalgae can consist of many molecules ranging from glycoproteins to hydrocarbons^{57, 58, 86, 87}, however, the most common component is glycoproteins that are often cross-linked with each other⁸⁷. One of the best studied green microalgal ECMs is that of *Volvox carteri*, which consists solely of glycoproteins, can hold over 2,000 cells differentiated mainly into somatic cells with a few reproductive gonidia cells, and is highly enriched in 4-hydroxyproline (Hyp) rich glycoproteins (HRGPs) that are cross-linked for stabilization⁸⁷⁻⁹⁰. The use of HRGPs is a common feature in the ECMs and cell walls of both algae and land plants, and they are commonly *O*-glycosylated at Ser, Thr, and Hyp residues with arabinose and galactose^{79, 91-93}.

The green colony-forming microalga *Botryococcus braunii* is well known as a producer of hydrocarbon oils that can be converted into combustion engine fuels^{15, 57}. The three chemical races of *B. braunii* known as A, B, and L are differentiated by the type of hydrocarbons produced; odd carbon number alkenes in the A race^{116, 181-183}, triterpenoid botryococcenes and methylsqualenes in the B race^{128, 184, 185}, and the tetraterpenoid lycopadiene in the L race^{71, 100, 121}. In all three races an ECM holds approximately 100 - 200 undifferentiated cells into a colony, and the hydrocarbons are biosynthesized inside the cells for secretion and storage in the ECM^{15, 57}.

The *B. braunii* ECM shows some striking differences from the ECMs of other green microalgae such as *V. carteri*. For example, the main component of the *B. braunii*

ECM is an intricate network of long chain polyacetal hydrocarbons that are cross-linked through epoxide bridges to monomers of the liquid hydrocarbons in each race⁷²⁻⁷⁴. Additionally, the polyacetal hydrocarbons can be threaded through large macrocyclic carbon rings¹⁸⁶. These two types of linkages offer both a strong chemical bond (epoxide bridges) for structural integrity and mechanical linkages (carbon rings) for extensibility^{57, 73, 186}. The spaces within this cross-linked hydrocarbon network are filled with the liquid hydrocarbon oils produced by each race of *B. braunii*^{15, 57}.

In the B race of *B. braunii* the ECM also contains a retaining wall located near the outer edge of the ECM that encompasses the cross-linked hydrocarbon network⁵⁸. As the name implies, the retaining wall holds the liquid hydrocarbons around the cells and within the spaces created by the cross-linked hydrocarbon network⁵⁸. Radiating outward from the retaining wall is a network of 2-3 μm long polysaccharide fibers that completely circumscribe the colony^{58, 60, 187, 188}. These fibers consist mainly of galactose and arabinose with uncommonly abundant (1 \rightarrow 2),(1 \rightarrow 3) and (1 \rightarrow 2),(1 \rightarrow 4) branching connections between the sugars, possibly to limit degradation by bacteria cohabitating the environment⁵⁸.

Protein granules are localized to the base of these polysaccharide fibers where the fibers attach to the retaining wall, and it has been suggested the protein in these granules may be involved in fiber biosynthesis⁵⁸. During *B. braunii* cell division, portions of the ECM covering the apex of each cell and containing the retaining wall, polysaccharide fibers, and protein granules are broken off from the ECM and shed into the culture media^{58, 187, 189}. These fragments have the appearance of cup-shaped “shells”, are easily collected

from the media, and analysis indicates the associated protein granules consist of a single protein species⁵⁸.

Here we identify this protein, term it polysaccharide associated protein (PSAP), confirm its localization to the ECM by association with shell material, characterize it as an HRGP, identify the PSAP glycosylations, and propose possible PSAP functions in *B. braunii* race B ECM polysaccharide fibril formation.

4.2 Results

4.2.1 Identification of cDNA and protein sequences for polysaccharide associated protein

Analysis of the *B. braunii* race B shell material in our previous study showed a single protein species associated with the shells⁵⁸. This was confirmed in our current study (Figure 32A) and we termed this protein polysaccharide associated protein (PSAP). The molecular mass of PSAP was previously estimated at ~150 kDa⁵⁸, and a more detailed analysis estimated PSAP molecular mass at 162.2 kDa (Figure 32B).

In order to identify the PSAP protein and clone the associated cDNA, the PSAP protein from SDS-PAGE (Figure 32A) was excised from the gel, digested with trypsin, the generated peptides identified by LC-MS/MS, and the resulting peptides BLASTed against a computational translation of a *B. braunii* race B transcriptome (Figure 32A)¹⁹⁰. This process identified three contigs containing matches to PSAP peptides (Figures 33B and 34). A combination of these contigs did not yield a full length open reading frame,

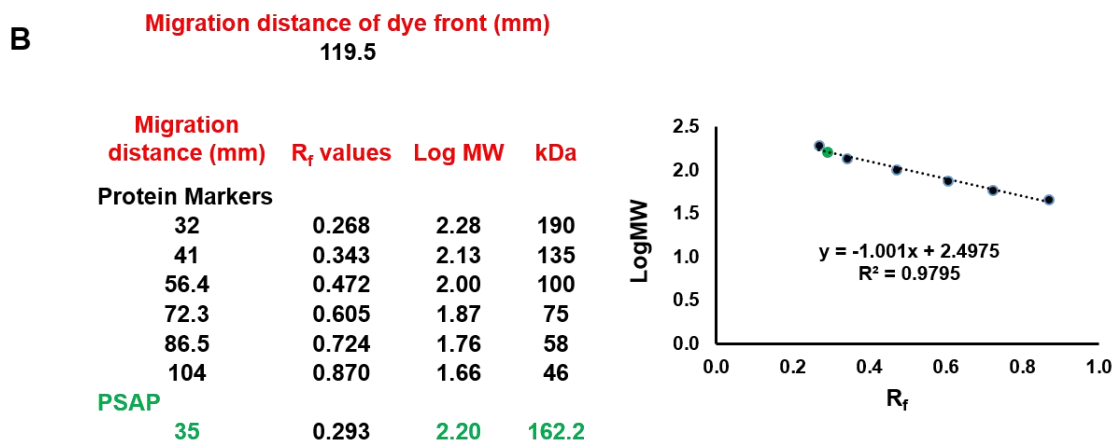
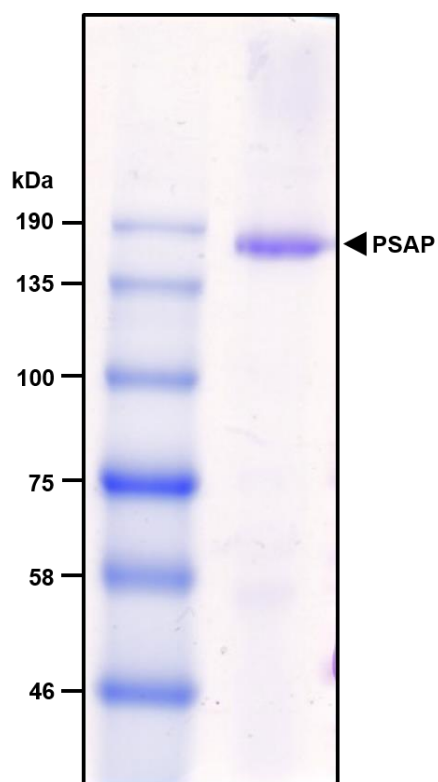


Figure 32 Analysis of PSAP molecular mass by SDS-PAGE. (A) The PSAP protein was extracted from shell material using SDS-PAGE sample buffer and separated by 12% SDS-PAGE followed by staining with Coomassie blue. (B) Calculation of PSAP molecular mass based on SDS-PAGE migration in comparison to molecular weight markers.

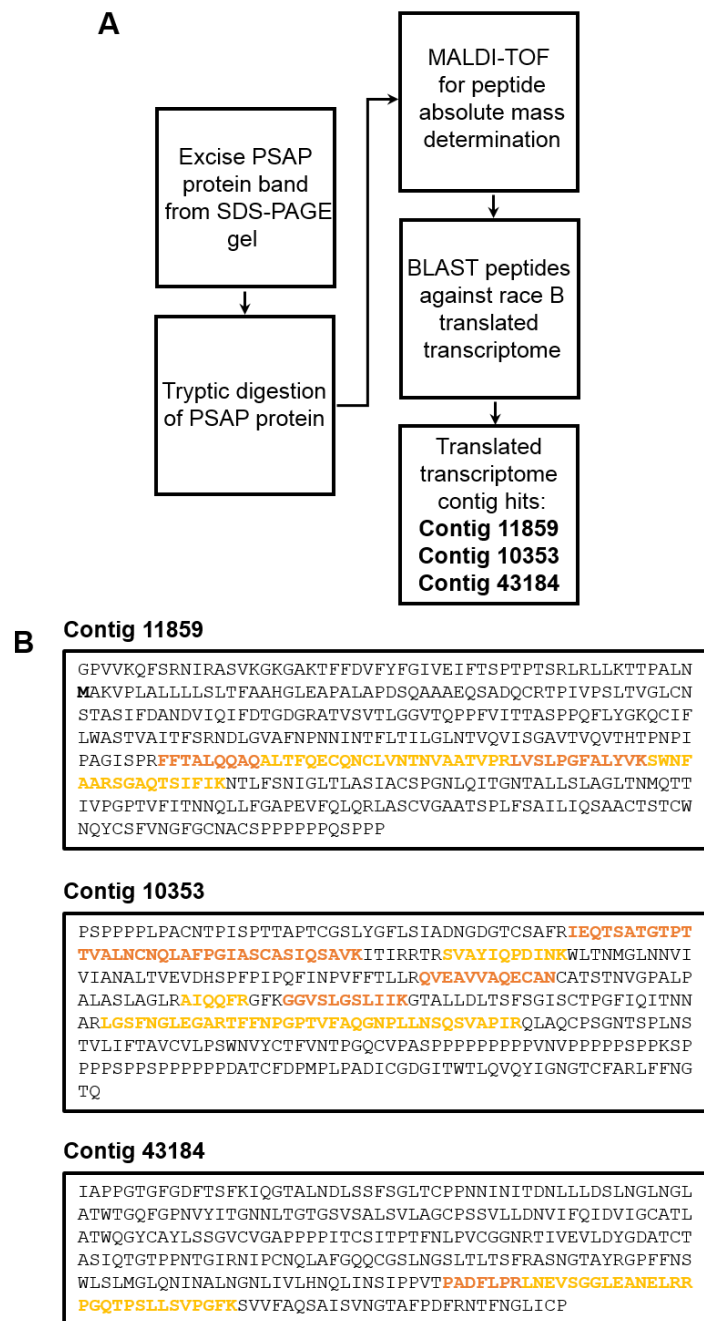


Figure 33 Identification of *B. braunii* race B transcriptome contigs matching PSAP MS peptides. (A) Step-by-step outline of PSAP MS analysis and identification of matching contigs. (B) Translation of the three *B. braunii* race B contigs that matched MS peptides from PSAP. PSAP peptides identified by MS that match the contigs are shown in alternating dark orange and light orange lettering. Methionine start site in contig 11859 is shown in black bold lettering.

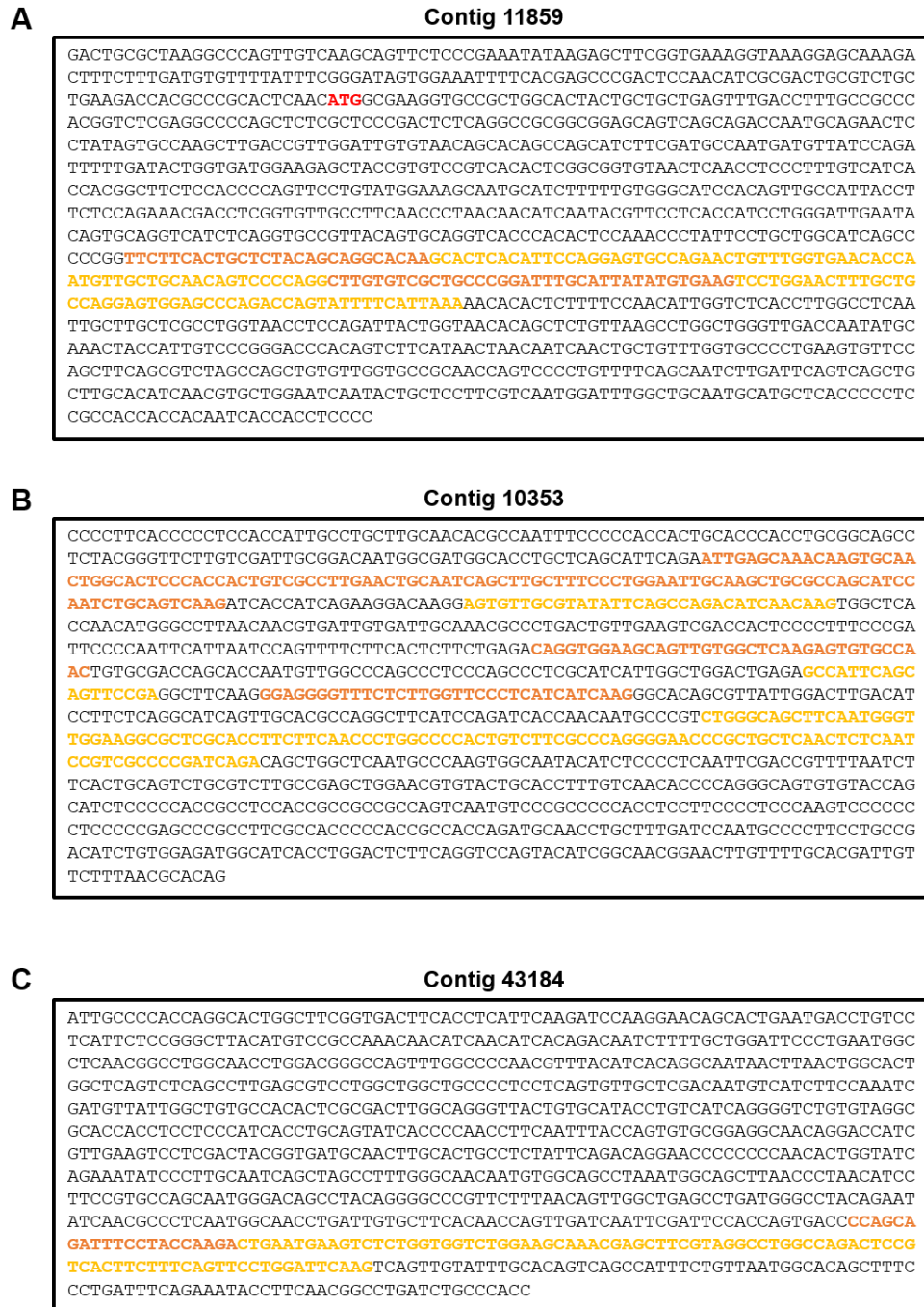


Figure 34 Nucleotide sequence of the three *B. braunii* race B contigs that matched MS peptides from PSAP. (A) Contig 11859, translation start codon shown in red lettering. (B) Contig 10353. (C) Contig 43184. Bases in alternating dark orange and light orange lettering correspond to the peptides found by MS analysis.

and thus further analysis was done to identify the full length PSAP sequence.

The full length *PSAP* cDNA and protein sequence were identified as laid out in (Figure 35A). Briefly, the three identified contigs were BLASTed against the *B. braunii* race B genome³⁹ to match the contigs to a 9.5 kb region of a single 401 kb genome scaffold (Figure 35B). The race B transcriptome was then aligned to the identified scaffold to map potential PSAP transcriptome reads to this scaffold (Figure 35B). The mapped transcriptome reads identified 11 exons and 10 introns spanning approximately 9 kb of genomic sequence (Figure 35B). The exons were assembled into a 4,928 bp cDNA sequence containing a *PSAP* open reading frame of 3,798 bp (Figure 36A) that encodes a protein of 1,266 amino acids (Figure 36B).

Analysis of the full length PSAP protein and cDNA sequence by BLAST against the entire GenBank database showed no identity to any proteins in the database. However, a domain search revealed an N-terminal 33 amino acid secretion signal peptide, three hydroxyproline (Hyp) rich motifs, and a predicted sugar binding domain (Figure 35C).

4.2.2 PSAP is a glycoprotein

The predicted molecular mass of PSAP is 132.8 kDa based on the open reading frame (Figure 36B) and the estimated PSAP molecular mass from shell material is 162.2 kDa (Figure 32B). This discrepancy in size suggests PSAP posttranslational modification such as glycosylation since many plant extracellular proteins are glycosylated¹⁹¹. Thus, glycosylation of PSAP was analyzed using an in-gel glycoprotein stain based on the periodic acid-Schiff (PAS) method. In this assay, periodic acid converts sugar vicinal diols

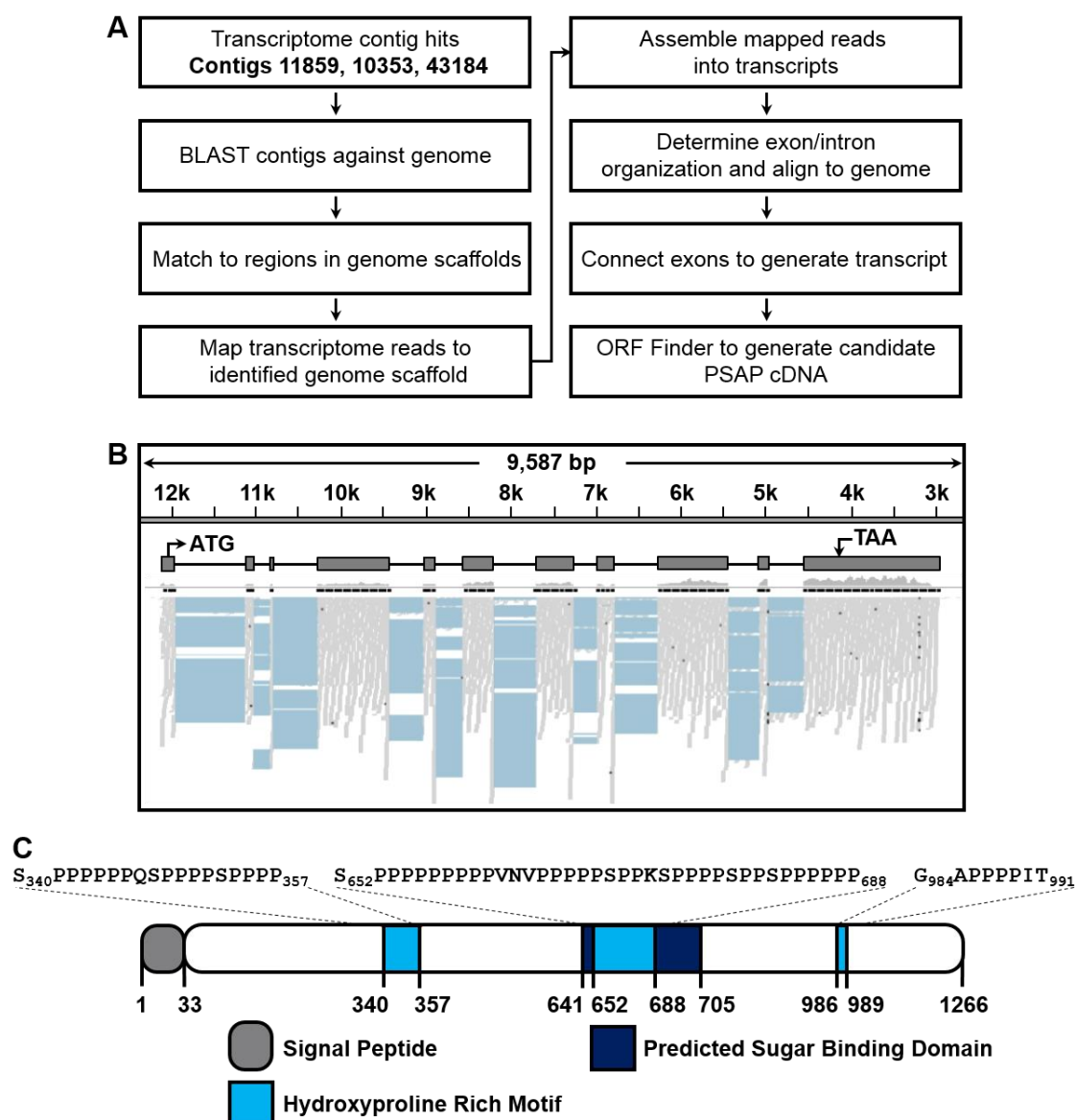


Figure 35 Identification of the PSAP cDNA and protein sequences using bioinformatics tools. (A) Bioinformatics steps followed to identify the full length PSAP cDNA. (B) Alignment of PSAP transcripts to genomic scaffold 141 to identify exon/intron junctions for assembling the PSAP cDNA. Solid gray line with kb markers indicates location of PSAP gene sequence within scaffold 141. Below the scaffold line the PSAP exons are indicated by gray boxes and introns by black lines. The translation start and stop locations are indicated. Transcriptome sequence reads aligned to the scaffold are shown as gray lines under the PSAP exons. Spaces where no transcriptome reads matched the PSAP sequence are shown as blue boxes. (C) Conserved domains found in the PSAP protein and the amino acid sequence of the Hyp-rich motifs.

ATGGCGAAGGTGCCGCTGGCACTACTGCTGCTGAGTTTGACCTTTGCCGCCCACGGTCTCGAAGCCCCAGCTCTCGCTCCGACATCTCAGGCCGCGGCGGA
CGAGTCAGACGAGCAATCGCAAGATCTCTATAGTGCCCAAGCTTGACCGTTGGATGTGTAAACAGCAGACCGAGCATCTTCGATGCCAATGATGTTATCCAGA
TTTTTGATACTGGTGAAGAGACTACCGGTGTCCGTCACACTCGCGCGGTAACTCAACCTCCCTTTGTCATCCACCGGGTCTCCACCAAGGCTCTCTGC
TATGGAAGCAATGCATCTTTTTGTGGGCATCCACAGTTGCCATTACCTTCTCCAGAAACGACCTCGGTGTTGCCCTCAACCCATAACACATCAATACGTT
CGTCCACCATCTCGGAGTGAATACAGTGCAGGTCACTCGAGTGCCTTACAGTGCAGGTCAACCCACATCTCAAAACCCCTATTCTCTGCTGGCATCAGCCCC
GGTTCCTCTCTACAGAGCACAAGCACTCACTCAGGAAGTGCAGAACTGTTTGGTGAACCAAACTGTTGCTGCACAACTGTCACCGGTTGTG
TCGCTGCCCGGATTGTCATTATATGTGAAGTCTCGGAACCTTTGCTGCCAGGAAGTGGAGGCCAGACAGTATTTTCATTAAACACACTCTTTTCCAACT
TCGTTCTCACTTTGGCTCAATTTGCTTGCCTGCCTGGTAACTCTCAGATTACTGTTAAACACAGCTCTGTTAAGCGTTGGCTGGGTTGACCAATATGCAAACT
CGATTGTCGGGAGCCCAAGCTTCTATAACTAACTCAACTGCTGTTTGGTGCCTGAAGTGTTCACAGTTCAGGCTTAGCCAGCTGTGTGTGGTGC
GCAACCACTCCCTGTTTTCAGCAATCTTGATTCAGTCAGCTGCTTGCACATCAACGTGCTTGAATCAATACTGCTCTCTGCTCAATGGATTGGCTGC
TGCACTGCTCAACCCCTCCGCGACCAACCAATCAACCACTCCCTCCCTCAACCCCTCCACCATTCGCTGTTGCAACACGCGCAATTTCCCCACACATGCAC
CCACTGCGGCAAGCTCTACGGGTTCTTGTCATTCGCAAACTGGCAATGGCAATGGCACTGACCATCTGAGTCAATCAGAAATGAGCAAAACAGTGCAACTGGCACTCCC
ACCACTGTGCGCTTGTAAGTGCATCAGCTGTGCTTTCCCTGGAAATGCAAGCTGCGCCAGCATCCAATCTGCAGTCAAGATCACCATCAAGGAAGCAAGGAG
TGTTGGGTATATTACCGCAGACATCAACAACTGGCTCAACCAATGGCGCTGCAACAGTGAATGTGATGTCAAAACCGCCCTGACTGTTGAAGTCGACCAT
CCGCTTTCGCGATTCCCAATTTCAATTAACAGTTTTCCTACTCTTCTGAGAAGGGAAGACAGTTGTGGCTCAAGAGTGTGCCAACTGTGCGACAGCG
ACCAATGTGTTGGCCAGCCCTCCAGCCCTCGCATATTGGTGGACTGAGAAGCACTCAGCAGTTCGCAAGGCTTCAAGGGAGGGGGTTTCTCTGGTTCCCT
CATCATCAAGGGCAGCGGTTATGGACTTGACATCTCTCGAGGCATGTTGCACGCGAGGCTCATCCAGATCAACCAATGACCCCTCTGGGCGAGCT
TCAATGGGTTGAAGGCGCTCGCACCTTCTTCAACCTGGCGCCACTGCTTCTCGCCAGGGGAACCCGCTGTCTCAACTCTCAATCCGTCGCCCGCATGAGA
CAGCTGGCTCAATGCCAAGTGCAGAAATACATCTCCCTCAATTCGACCGTTTAAATCTCTCGCAGTCTGTGGCTTGTGCCAGCTGGAAGAGTGTAATGCAC
CTTTGTCAACACCCAGGGCGAGTGTACCAAGCATCTCCCCACCGCTCCACCGCGCCGAGTCAATGTCCCGCCACCCTCTCTCCCTCCCAAGT
CCGCCCCCTCCCCAGCCGCTCTGCCACCCCCACGCCACAGATGCAACCTGCTTGTATCAATGCCCTTCTCGCGACATCTGTGGAGATGGCAT
ACCTGGAATCTTCAGGCGGACCATCTGGCAAGCAAGGAATGTTTGTGACAGATGTTCTTTAAGGCACAGGATATGGGACCCAGGCTAAGCGTTTACCGCTGC
CCAAGCCACTTCGAGGCGCACTATTTTCCGCGGCACTTGGCTGACCTTGAATGCCAACTGTACCTCTTTGTGCAAAACCGGCCAATTTAACTGCAT
CAACCTACTTTCAACCGCAGTTCAATGATGCTCGACCGCTCTGGGGTTGGATCCTTGAGACATTTAGTATGTGCTGTCTTGTGTTGGTATCAGACT
CATTTTCCCAATCTCTGTCAAATAGCCCGGTGTCTCTGGCAACACTCAAGAAAGTGTGTGCCATCCAAAGTATCGAGTGCATCAATTTGGCCTCTCTCC
GGATGTCACCCCTGTATCCCGACTGATAGGCTTCTCGGGCTGACGGGCTGAGGAGATCAATTTGGCCACCAAGGCATGGCTCGGTGACTTCACT
CATTCAAGATCCAAGGAACAGCACTGAATGACCTGTCTCTATCTCTCGGGTTCATAGTTCGCGCAAAACATCAACATCAGACAACTCTTTGTGCTGGAT
TCCTCGAATGGCTCATACGGCTCGGCAACTCGAGCGGGCGAGTTGGCCCAAGGTTTACATCAGGCAAACTAACTAATCGGCATGGCTGCTCATCGAGT
CTGTAGCGCTCTGGCTGGCTGCCCTCTCATGTTCTGTGCTGCAAACTGTATCTTCAAACTCATGTTATTTGGCTGTGCGACACTCGCGACTTGGCAGGTT
ACTGTGCATACCTGTCTACAGGGGTCTGTGTAGGGCGACCACTCTCTCCCATCCTGTATGCATCACTGCAATCTCAATTTACCGAGTGTGCGGAGGCAAC
AGGACACCTGTTGAAGTCTCGACTAGGTTGAGCAACTTGCAGCTGCTTCTTATCAGACAGGAACCCGCCCAACCTGGTATCAGAAATATCCCTTTCGAGT
TCAGCTAGCTTTGGGCAACAAATGGGACGCTTAATGGCAGTCAACCTTAACATCTCTCGTGAGCAATGGGACAGCTCAGAGGGCGGCTCTTTAT
ACAGTTGGCTGAGCCTGATGGCTCAGAAATATAACCGCTCAATGGCAACCTGATTGTGCTTCAACCAAGTGTATCAATTCGATTCCACCGATGACC
CCAGCAGATTTCTTCAAGAGTCAAGTAGCTCTGTGGTGGTGTGGAAGCAACAGCACTCTGAGCGCTGGGCAGCTCCGCTCACTCTTTGAGTCTCGG
ATTTCAGTCAGTTGTATTGTGACAGTCAGGCTCTTGTGTTAATGGCAGAGCTTTCCCTGATTTCAGAAATACCTTAACCGGCTGATCTGCCACCTTACC
AATTCATGAATTTCAAAACACCTTAATCTGAGGCTTTACCAAGGCTGTGATGCTCTGGTCAACCACTTCTCACTCAGTAAATACCTTATCTCT
CTCGGCACCAATCTTACACTCTCAGCATTTGATGCAATTCGAACTGGGCGAGTGCCTTACGGGCTTGTAAAGCAACATCAGCGGTATCATTAAATATAA
AGTTTCAATCTGCCCTATAATCCATCCACCTTTTCGCAAAATCTGCGCTTATATTTTAAACAGCTGGCCATAA

[illegible]

Figure 36 Nucleotide and protein sequence of PSAP. (A) PSAP ORF nucleotide sequence. Bases shown in alternating dark orange and light orange lettering correspond to the peptides found by MS analysis. Contig locations are identified by a black box for contig 11859, a red box for contig 10353, and a purple box for contig 43184. (B) PSAP protein sequence. The PSAP signal peptide is highlighted in green. The PSAP peptides identified by MS that match the contigs are shown in alternating dark orange and light orange lettering. They Hyp-rich motifs are shown in magenta lettering.

to aldehydes, which are bound by the Schiff reagent for detection under UV light¹⁹². Three different amounts of PSAP plus positive and negative controls were separated by SDS-PAGE and the gel stained for glycoproteins. Positive controls included glycosylated molecular weight markers and a mouse antibody (IgG). Negative controls included non-glycosylated molecular weight markers and the tomato AGC protein kinase Adi3 as expressed and purified from *E. coli*¹⁹³. In the glycoprotein staining assay all positive and negative controls showed the correct staining pattern, and the PSAP protein showed staining in a concentration dependent manner (Figure 37). This suggests PSAP is glycosylated and there may be ~29.4 kDa of glycosylation on PSAP to reach the observed 162.2 kDa size.

4.2.3 Identification of PSAP N-linked glycans

Since the glycan staining assay confirmed PSAP as a glycoprotein (Figure 37), the identity of the PSAP glycans was investigated. First, the possibility of *N*-linked glycosylation on Asn residues was analyzed. A computational analysis showed PSAP contains 96 Asn residues and 14 of these 96 (Figure 38) are found within the known NxS/T *N*-glycosylation consensus sequence¹⁹⁴. Which of the 14 Asn residues contained *N*-linked glycans was then analyzed experimentally. Briefly, PSAP was in-gel digested with chymotrypsin followed by enzymatic *N*-glycan release using PNGase A in an ¹⁸O-labeled dH₂O buffer, which will release the *N*-linked glycans, convert the glycan linked Asn to an ¹⁸O-labeled Asp, and introduce a 3 Da mass shift at the glycosylation site. The resulting peptides were analyzed by LC-MS to identify ¹⁸O-labeled peptides for *N*-linked

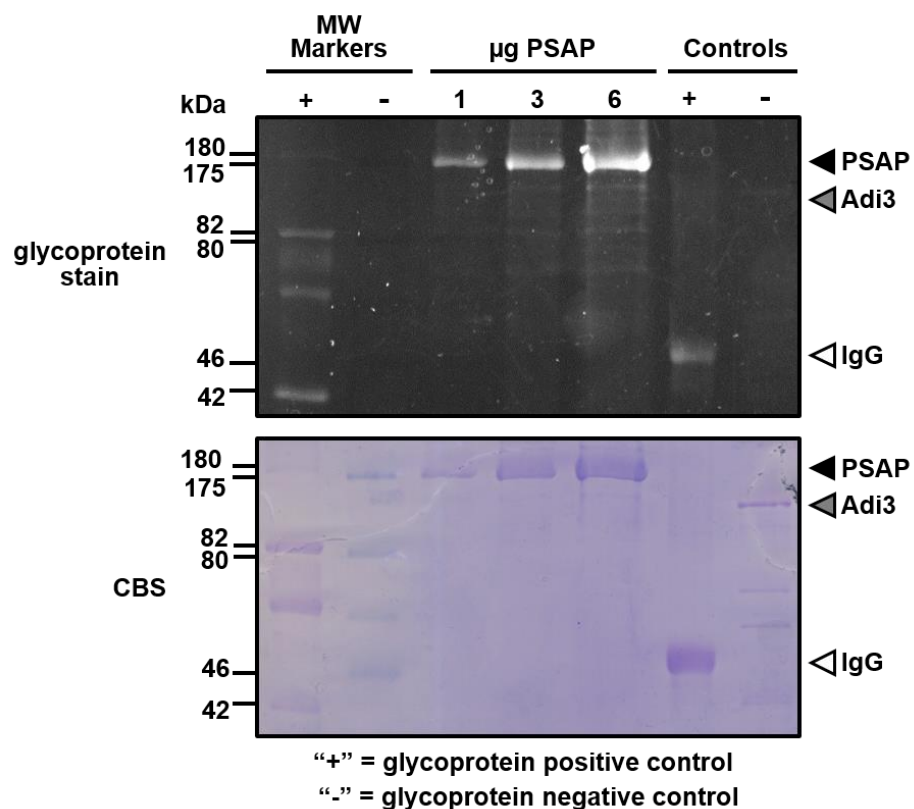


Figure 37 In-gel glycoprotein staining of PSAP. Three amounts of the PSAP protein plus positive and negative controls were separated by SDS-PAGE, treated with the glycoprotein stain, and visualized by UV light. The gel was stained with Coomassie blue (CBS) following the glycoprotein stain for visualization of all proteins in the assay. Positive controls include glycosylated molecular weight markers and a mouse antibody (IgG). Negative controls include non-glycosylated molecular weight markers and the tomato Adi3 protein.

MAKVPLALLLLSLTFAAHGLEAPALAPDSQAAAEQSADQCRTPIVPSLTVGLCHSTASIFDANDVIQIFD 70
 TGDGRATVSVTLGGVTQPPFVITTASPPQFLYGKQCIFLWASTVAITFSRNDLGVAFNPNNTINTFLTILG 140
 LNTVQVISGAVTVQVTHTPNPIPAGISPRFFTALQQAQALTFQECQNCCLVNTNVAATVPRLVSLPGFALY 210
 VKSWNFAARSGAQTSTFIKNTLFSNIGLTLASACSPGNLQITGNTALLSLAGLTNMQTTIVPGPTVFIT 280
 NNQLLFGAPEVFLQRLASCVGAAATSPLESAIILQSAACTSTCWNQYCSFVNGFGCNACSPPPPPQSP 350
 PPSPPPLPACNTPISPTTAPTCTGSLYGFLSIADNGDGTCSAFRIEQTSAATGTPPTVALNCNQLAFPGIA 420
 SCASIQSAVKITIRRTSVAYIQPDINKWLTNMGNNVIVIANALTVEVDHSPFPQPFINPVFFTLRLQ 490
 VEAVVAQECANCATSTNVGPALPALASLAGLRAIQFRGFGGVSLGSLIIGTALLDLTSFSGISCTPG 560
 FIQITNNARLGSFNGLEGARTFFNPGPTVFAQGNPLNSQSVAPIRQLAQCPSTSTPLSTVLIIFTAVC 630
 VLPSWNVYCTFVNTPGQCVPASPPPPPPPPVNVPPPPSPPKSPPPSPPPSPPPPPDATCFDPMPLPA 700
 DICGDGITWTLQVQYLGNGTCEARLFFNGTQLSLTAAQATCQATIFPGRTRCRTLNGNLYLFVTNQANLTA 770
 STYFQPQFNDGLTALGVGSLETIEYVLSLVVDHTPFPIPVQLAPVFLATLRNVFAIQVIESINFGLLPDV 840
 PPPVPRLVGLPGLTGVRRIIAPPGTGFGDFTSFKIQGTALNDLSSFSGLTCPNNINITDNLILLDSINGL 910
 NGLATWTGQFGPNVYITGNMLTGTGSVSAISVLACGPPSSVLLDNVIFQIDVIGCATLATWQGYCAYLSSG 980
 VCVGAPPPITCSITPTFNLPVCGGNRTIVEVLDYGDATCTASIQGTGTPPNTGIRNIPCNQLAFGQQCGS 1050
 LMGSLTLTSFRASNGTAYRGPFFNSWLSLMGLQNLNGLNGLVHLNQLINSIPPVTPADFLPRLNEVSG 1120
 GLEANELRRPGQTPSLLSVPGFKSVVFQAQSAISVNGTAFPPDFRNTENGLICPPYQFMFTNNPNLRSYQG 1190
 LDALGQPPFLPHVNTLFSPATPTLLTIDAIANWAGCPTGLSNTIDGIINIKVSFCPNPITTTFAQICAYI 1260
 LFNTCP 1266

Figure 38 PSAP N-linked on O-linked glycan analysis. Peptides highlighted in cyan and green represent the peptides identified by LC-MS/MS in N-linked glycan analysis. Cyan peptides do not contain predicted N-linked glycosylation sites. Green peptides contain predicted N-linked glycosylation sites. Asn (N) residues highlighted in red were confirmed to be an N-linked glycosylation site; N54, N718, N728, N767, N897, N930, N1006, N1052, N1064, N1155. Asn (N) residues highlighted in purple were confirmed to not be an N-linked glycosylation site; N615, N620, and N1212. Potential N-linked glycan site N1178, highlighted in magenta, was not found in any peptide identified in this study. The peptide highlighted in yellow represents the peptide identified to contain Hyp O-linked glycans; Hyp986, Hyp987, Hyp988, Hyp989 are shown in blue lettering. The protein sequence in the box identifies the peptide used for PSAP antibody development.

glycan site mapping. This analysis identified 25 peptides covering 40% of the PSAP protein sequence including 12 peptides containing 13 of the 14 potential *N*-linked glycan sites (Figure 38). A peptide containing potential *N*-linked glycan site N1178 was not identified in the analysis. Within the 12 identified peptides containing potential *N*-linked glycan sites, residues N54, N718, N728, N767, N897, N930, N1006, N1052, N1064, and N1155 were found to be *N*-linked glycan sites (Figure 38). Possible *N*-linked glycan sites N615, N620, and N1212 were found to not be glycosylated. Overall, 10 of the possible 14 *N*-linked glycan sites were confirmed to harbor glycans.

Next, structural details of the PSAP *N*-linked glycans were determined by releasing the glycans with PNGase A, purifying the free glycans¹⁹⁵, and subjecting the glycans to in-depth structural characterization using several MS techniques. Conversion of the glycans to partially methylated alditol acetates (PMAAs) followed by a GC-MS glycosyl linkage analysis showed the *N*-linked glycans contained mainly terminal, 2-linked, 3-linked, 6-linked, and 3,6-linked mannose (Table 9). The next most abundant sugar was 4-linked *N*-acetylglucosamine (GlcNAc) followed by terminal fucose and terminal arabinopyranose (Table 9).

In order to determine the size and glycosyl sequence of the PSAP *N*-linked glycans, the released and purified *N*-glycans were permethylated and profiled directly by NSI-MSn (Figure 39). The analysis revealed seven different glycan structures (Figure 40A). The main *N*-glycan components are high-mannose type structures, Man₄₋₇GlcNAc₂, that contain a base of (1 → 4)-connected GlcNAc with branches of mannose connected by (1 → 3)- and (1 → 6)-linkages (Figure 40A). Interestingly, three of the glycans were seen

Peak #	RT (min)	Peak Area	Assignment	Peak Area %
1	7.255	895846	T-Araf	0.6%
2	7.513	1571949	T-Arap	1.1%
3	7.725	6361609	T-Fuc	4.6%
4	8.24	369399	4-Fuc	0.3%
5	8.509	85705175	T-Man*	59.3%*
			T-Glc*	2.8%*
6	8.646	764836	T-Gal	0.6%
7	9.224	11076348	2-Man	8.0%
8	9.321	3721995	3-Man	2.7%
9	9.521	1811817	6-Man	1.3%
10	10.568	9919887	3,6-Man	7.2%
11	12.291	14740972	4-GlcNAc	10.7%
12	13.309	1001179	3,4-GlcNAc	0.7%
Total		137941012		100.0%

Araf: Arabinofuranose; Fuc: Fucose; Arap: Arabinopyranose; Man: Mannose; Glc: Glucose; Gal: Galactose; GlcNAc: N-acetylglucosamine; T-: Terminal; 2-: 2-linked; 3,6-: 3 and 6 linked (branched); 3,4-: 3 and 4 linked (branched).

*These sugars co-eluted on a DB-1 column but were well-separated on an SPB-column. The peak area % of terminal mannose and terminal glucose in the co-eluted peak was calculated based on peak area % of the corresponding peaks observed in SBP-column separation.

Table 9 Linkage analysis of the *N*-linked glycans on PSAP.

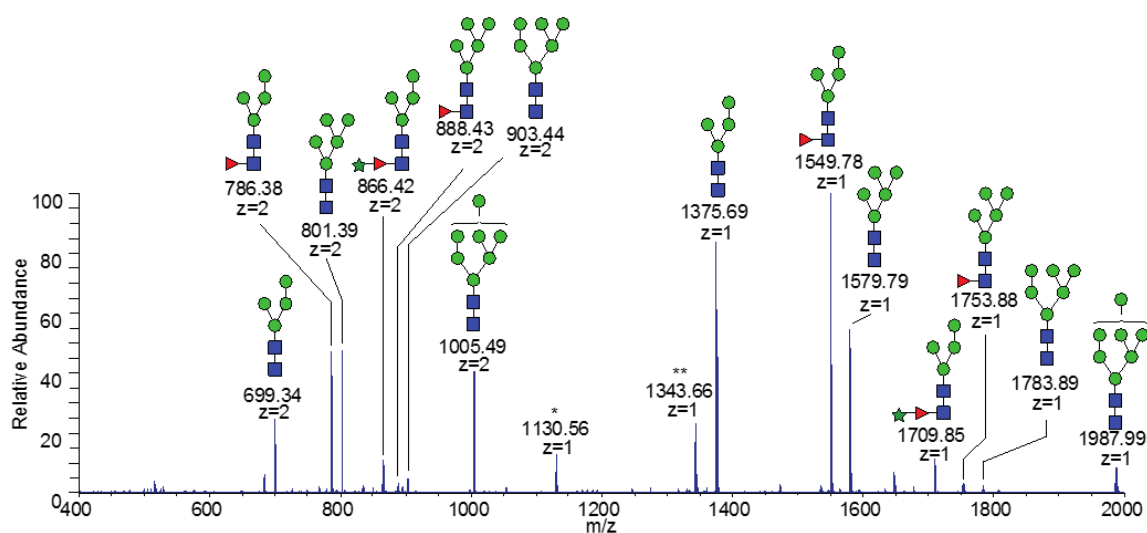


Figure 39 PSAP *N*-linked glycan analysis by NSI-MSn. *N*-linked glycans were released enzymatically from PSAP with PNGase A, permethylated, and profiled by NSI-MSn. Full MS profile of the released permethylated glycans showing a series of high-mannose type *N*-glycans and fucosylated *N*-glycans as the major components. In addition, a unique *N*-glycan carrying fucose and arabinose was detected at $m/z = 1709$ (1+) and $m/z = 866$ (2+). A single asterisk indicates an *N*-glycan fragment with Hex₄HexNAc₁. Two asterisks indicates the in-source fragment ion, presumably a Z_{1β} fragment ion from the precursor at $m/z = 1709$ (Hex₄HexNAc₂Pent₁Deoxyhex₁) due to a labile *O*-glycosyl linkage of the internal fucose.

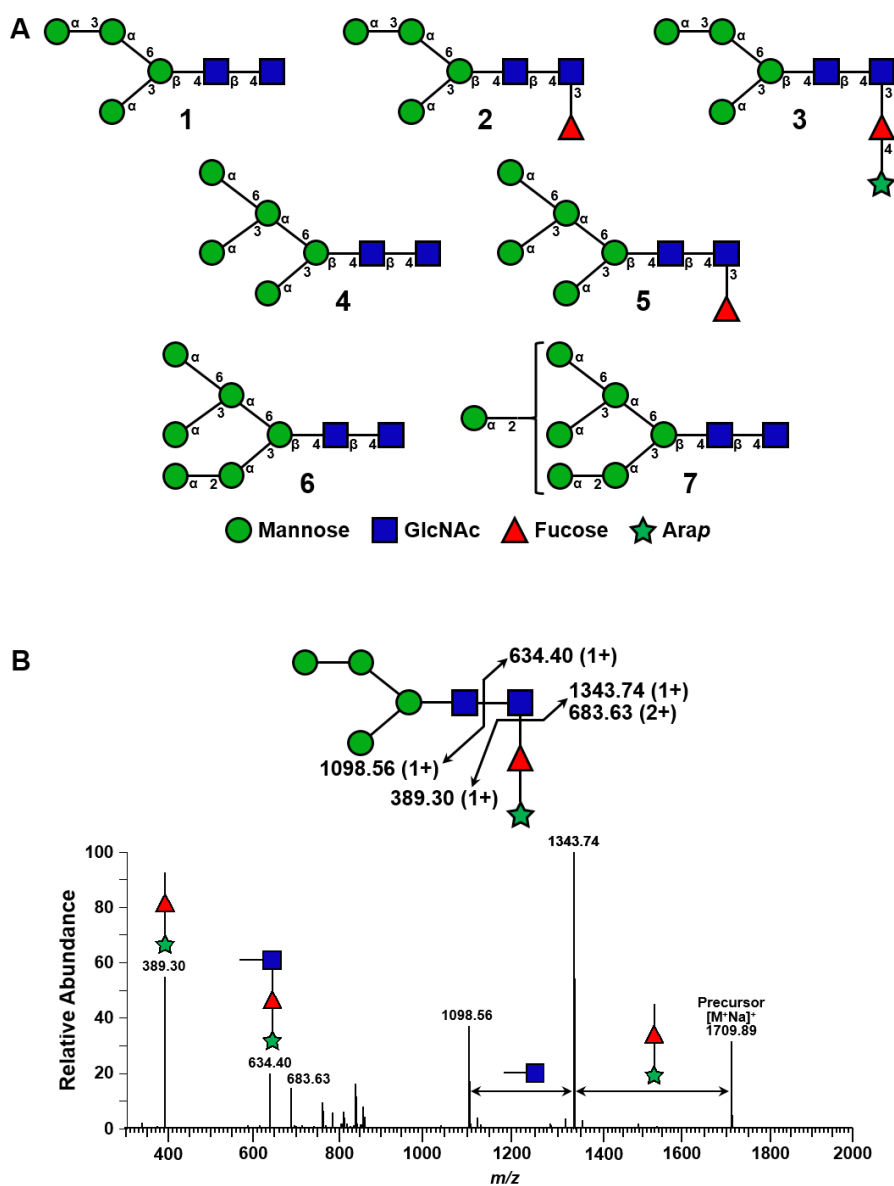


Figure 40 Structures of the *N*-linked PSAP glycans. (A) Structures of the seven *N*-linked glycans found on PSAP as determined by the MS data shown in Figure 39. (B) Tandem MS analysis of the unique fucose and arabinose containing *N*-glycan ion detected at $m/z = 1709.85$ shown in supplemental Figure 39. Key ions for identifying the location of arabinose and fucose are shown diagrammatically above the MS spectrum. Presence of a C-type fragment ion at $m/z = 389.30$, corresponding to an AraFuc disaccharide unit, and a Y-type fragment ion at $m/z = 634.40$, corresponding to an AraFucGlcNAc trisaccharide unit, indicated the AraFuc disaccharide is attached to the *N*-glycan chitobiose core.

to contain a fucose (glycans 2, 3, and 5 in Figure 40A) and one unique glycan contained an additional arabinopyranose (glycan 3 in Figure 40A). Tandem MS analysis was performed to identify the fucose and arabinose positions on the glycan ion at $m/z = 1709.85$ corresponding to $\text{Man}_4\text{GlcNAc}_2\text{Fuc}_1\text{Ara}_1$ (Figure 39). In the fragmentation analysis of this ion (Figure 40B), the presence of fragment ions at $m/z = 389.3$ (AraFuc) and $m/z = 634.4$ (AraFucGlcNAc) indicates the AraFuc disaccharide is attached at the reducing-end GlcNAc. This MS/MS data together with the glycosyl linkage analysis indicated the fucose is linked to the terminal GlcNAc at the 3 position (glycans 2, 3, and 5 in Figure 40A), and the arabinopyranose is linked to the fucose at the 4 position (glycan 3 in Figure 40A). The total molecular mass of the released glycans is predicted to be between 10-16 kDa.

4.2.4 Identification of PSAP hydroxyproline O-linked glycans

PSAP was also analyzed for mucin-type Ser or Thr *O*-linked glycosylation by chymotryptic digestion followed by LC-MS analysis of the resulting peptides, but this analysis did not identify any mucin-type *O*-linked glycans. Next, since PSAP contains many Pro residues, which are commonly hydroxylated to hydroxyproline (Hyp) and glycosylated in plant extracellular matrix proteins⁹¹, PSAP was analyzed for Hyp *O*-linked glycans. For this analysis, PSAP was reduced, alkylated, and digested with pronase, which should digest PSAP into single amino acids or linkages of a few amino acids. However, the PSAP Pro-rich regions will remain nearly intact since pronase is not active toward Pro.

The resulting digestion mixture was separated and analyzed by LC-MS (Figure 41A and B). The analysis identified one Pro-rich peptide containing four consecutive Pro residues at positions 986-989 (Figure 41B), all of which are hydroxylated to Hyp and carry arabinogalactan-type oligosaccharides consisting of 1,6-deoxyhexose, 4 hexoses, and 7 pentoses (Figure 41A and B). A linkage or composition analysis was not performed for the *O*-glycans, but the MS data (Figure 41A and B) indicates a pentose is attached to the peptide backbone and the 6-deoxyhexose is terminally linked. This observation is consistent with the Hyp *O*-glycans found on plant arabinogalactan proteins¹⁹⁶. Based on this analysis, the derived *O*-linked glycans on PSAP are shown in Figure 41C. It should be noted, at this time it is not clear which glycan is attached to each Hyp. The total molecular mass of the Hyp *O*-linked glycans is predicted to be between 1-2 kDa.

4.2.5 Estimation of PSAP molecular mass based on glycosylations

Taking into account the molecular mass of the naked PSAP protein and the glycosylations found on PSAP, the glycosylated molecular mass of PSAP is calculated to be between 143.8 and 150.8 kDa. This takes into account 132.8 kDa from PSAP, 10 – 16 kDa from *N*-linked glycans and 1 – 2 kDa from Hyp *O*-linked glycans. This value is below the calculated value of 162.2 kDa for PSAP based on SDS-PAGE migration (Figure 32B), suggesting there may be additional glycosylations or other modifications on PSAP to account for the ~11.4 – 18.4 kDa needed to reach the observed mass. It should be noted that if an SDS-PAGE gel is overloaded with PSAP a ladder of higher molecular mass

Figure 41 Analysis of PSAP Hydroxyproline *O*-linked glycans by LC-MS. PSAP was reduced, alkylated, digested with pronase, and analyzed by nano C₁₈-LC-MS. A glycopeptides ion, corresponding to G₉₈₄AP*P*P*P*IT₉₉₁ from PSAP carrying 1 deoxyhexose, 4 hexoses, and 7 pentoses eluted at 16.81 min of the LC-MS run. (A) The MS/MS-CID spectrum of the glycopeptide at $m/z = 1266.99$ (2+) (mono isotopic m/z of the ion is 1266.4869) showing a series of neutral losses of glycan moieties in the order of deoxyhexose, hexose, and then pentose, suggesting the pentose is attached to a hydroxyproline, whereas the hexose and deoxyhexose are on the non-reducing terminal end. (B) MS/MS-HCD spectrum (low mass region) of the same glycopeptide ion. The raw mass region of the HCD data was analyzed to determine amino acid content and sequence. The series of internal fragment ions from the peptide backbone allowed determination of the peptide sequence as G₉₈₄AP*P*P*P*IT₉₉₁ from PSAP. All Pro in the peptide were hydroxylated (hydroxyproline, Hyp). C, Hydroxyproline *O*-linked glycan structures found on PSAP on Hyp at positions 986-989 based on MS data shown in (A) and (B). Hex, hexose; Pent, pentose; Deoxyhex, deoxyhexose; P*, Hydroxyproline.

PSAP can be seen above the primary PSAP band. This can be seen in Figure 37 and may represent PSAP with different degrees of glycosylation.

4.2.6 PSAP contains a functional secretion signal peptide

Next we undertook a set of experiments to confirm that PSAP is indeed localized outside of *B. braunii* cells in the ECM and with the shell material. First, subcellular localization prediction tools^{197, 198} generated a 100% probability that PSAP is localized to the secretion pathway via the predicted signal peptide. Next, the functionality of the PSAP signal peptide as a secretion signal was tested through an N-terminal translational fusion to GFP (Figure 42A). This fusion was cloned into a galactose inducible yeast expression vector, expressed for 8 h in the presence of galactose followed by collection of the cells and media separately. Analysis of cellular and media proteins for the presence of GFP by α -GFP western blot showed a small amount of GFP was found in the media without the PSAP signal peptide, while a large amount of GFP was found in the media when expressed with the PSAP signal peptide (Figure 42B). These studies suggest the PSAP signal peptide is a functional secretion signal for exporting PSAP outside of *B. braunii* cells into the ECM.

4.2.7 Localization of PSAP to shell material

The second set of experiments performed to confirm PSAP localization to the ECM shell material were based on glycan detection with lectins coupled to fluorescence microscopy. Additionally, an α -PSAP peptide antibody was developed and also used to

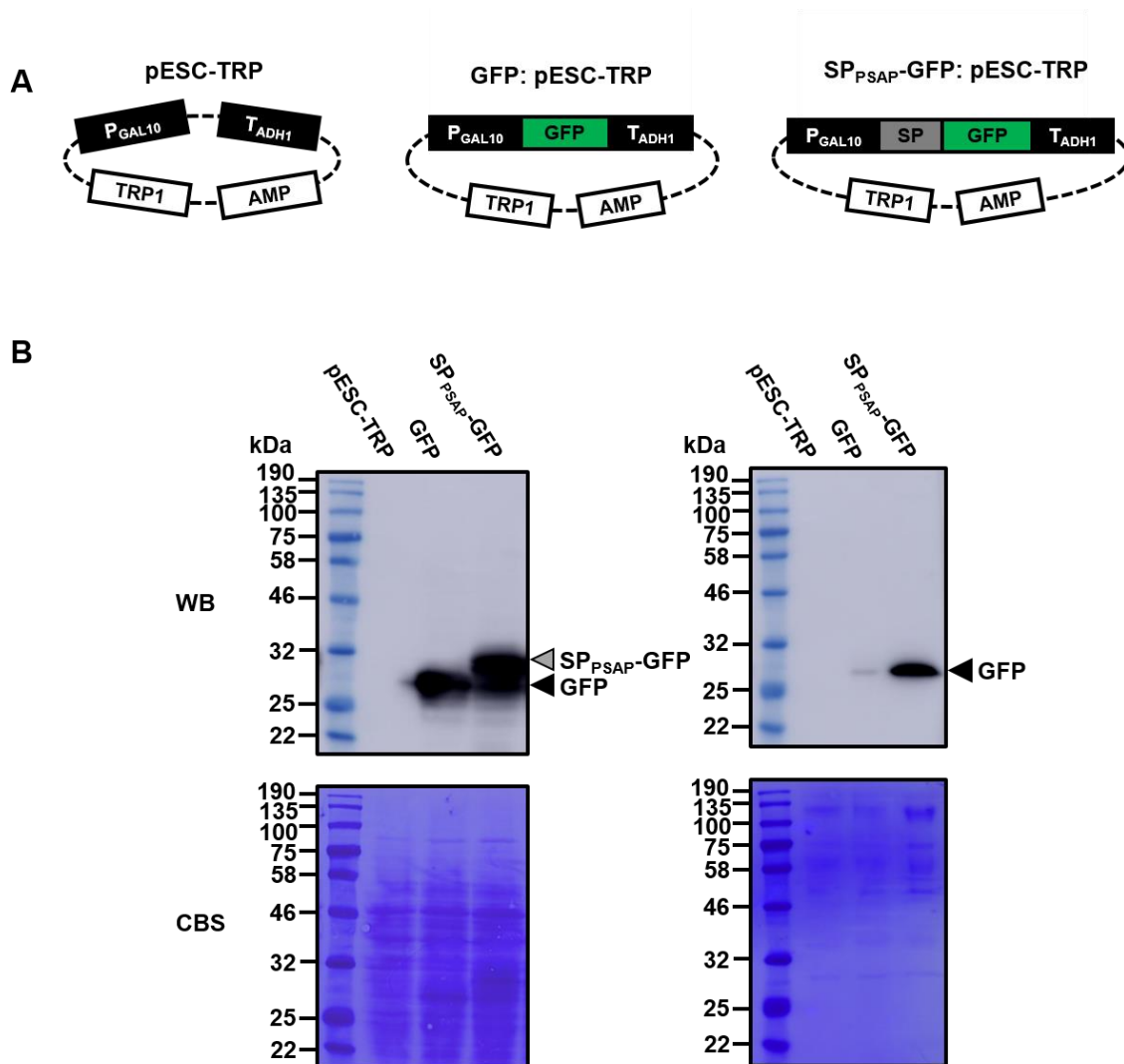


Figure 42 Functional analysis of the PSAP signal peptide. (A) The three constructs used for GFP secretion analysis in *S. cerevisiae*; pESC-TRP, GFP: pESC-TRP, and SP_{PSAP}-GFP: pESC-TRP. (B) Western blot detection of GFP from *S. cerevisiae* cell protein extracts and media protein extracts using the α -GFP-HRP antibody (1:3,000 for cell protein extracts and 1:1,500 for media protein extracts). WB, western blot; CBS, Coomassie blue stain of WB.

confirm PSAP localization to shell material. An initial analysis of shell material by electron microscopy showed many small dots on the shell surface (Figure 43) that resembles the granules seen in our previous studies⁵⁸. We speculated that these small dots are the PSAP protein in granules.

We have shown that PSAP is a mannose and GlcNAc containing glycoprotein (Figure 40A). The *Hippeastrum* hybrid lectin (HHL) binds specifically to α -1 \rightarrow 3 and α -1 \rightarrow 6 linked mannose structures¹⁹⁹ and the *Lycopersicon esculentum* lectin (LEL) specifically binds to GlcNAc²⁰⁰. All of these types of sugars are found in the PSAP glycans (Figure 40A). Both lectins were capable of detecting PSAP protein in shell material as well as a total protein extract by western blot using biotin labeled lectins and streptavidin-HRP (Figure 44A and B). Next, the lectins were used in fluorescence microscopy. Shell material was incubated separately with biotin labeled HHL or LEL, followed by incubation with streptavidin-Texas red and analysis by fluorescence microscopy. Both lectins were seen to bind to the shell material with the LEL lectin labeling small dots on the shells similar to the granules (Figure 45A, B, and C and Figure 46A, B, and C). The *Sambucus nigra* lectin (SNA), which is specific for sialic acid was used as a negative control (Figures 44C and 46D). A similar procedure was followed with the exception of following lectin binding with streptavidin-HRP incubation and detection with diaminobenzidine (DAB), which produces a dark staining that can be seen by light microscopy. The results show that the lectins labeled the shell material in a lectin dependent manner (Figure 47). These studies suggest the lectins are specifically binding

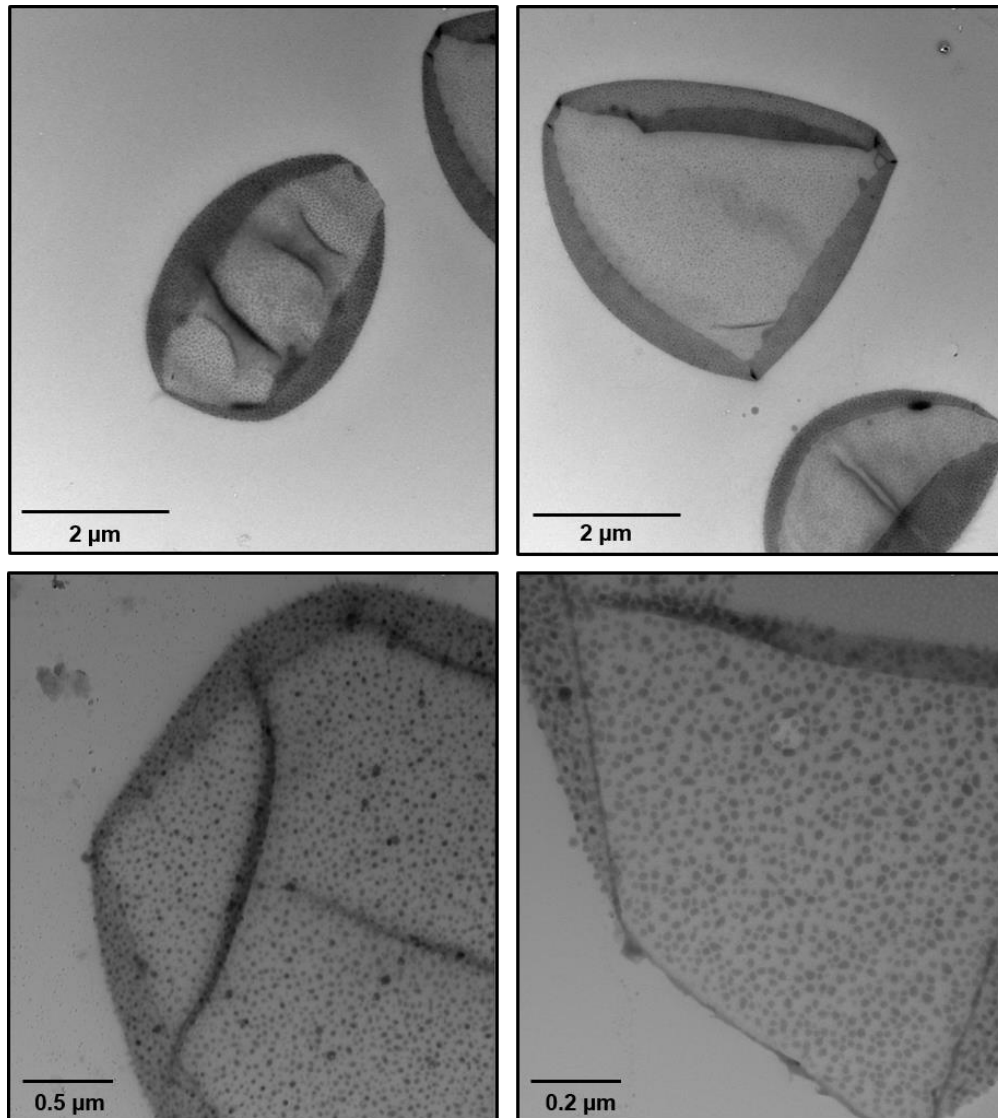


Figure 43 Electron micrographs of *B. braunii* race B shell material. Purified shells were resuspended in water and placed on an EM copper grid with carbon support film and allowed to dry. The shells were then visualized by TEM.

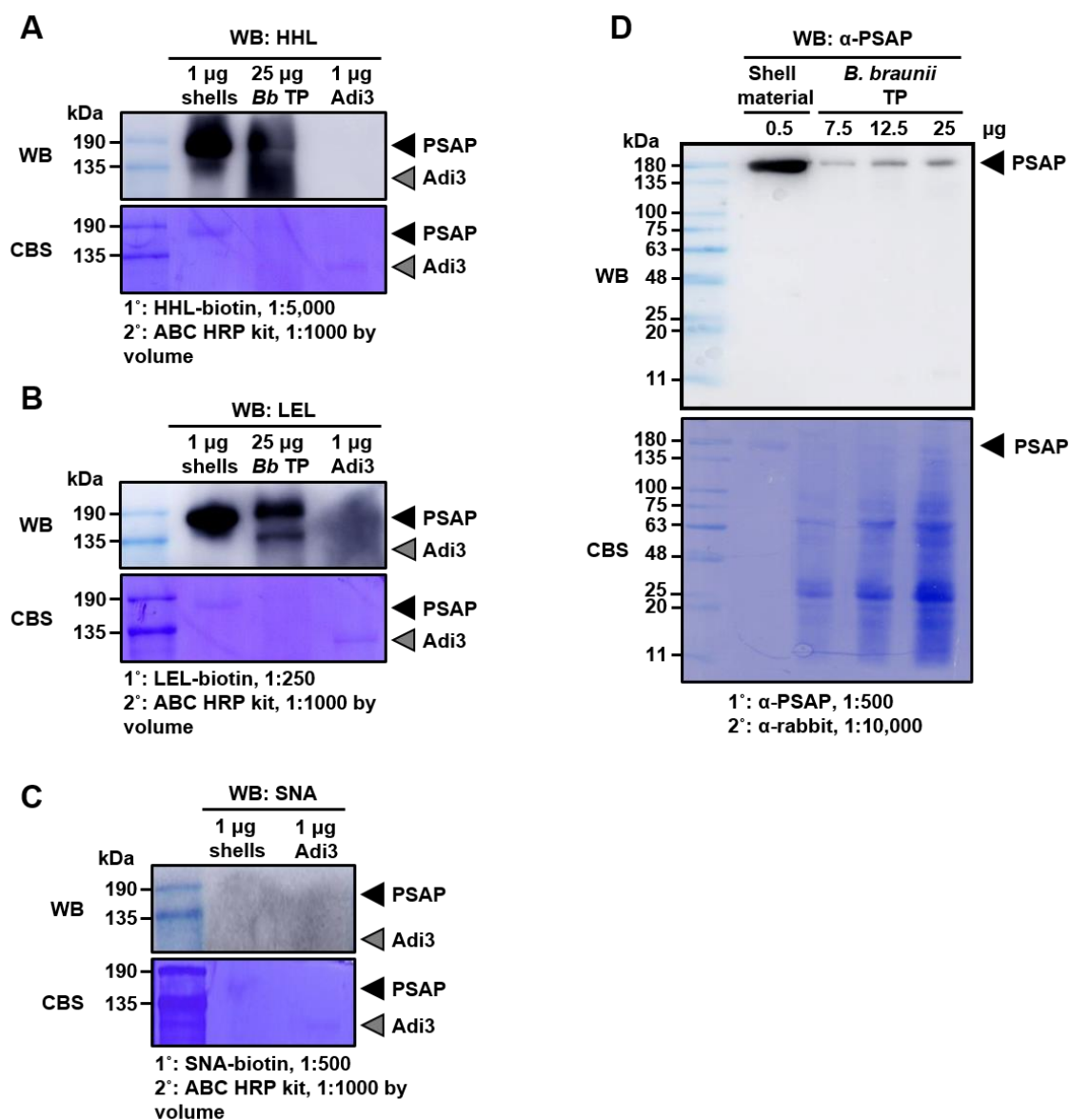
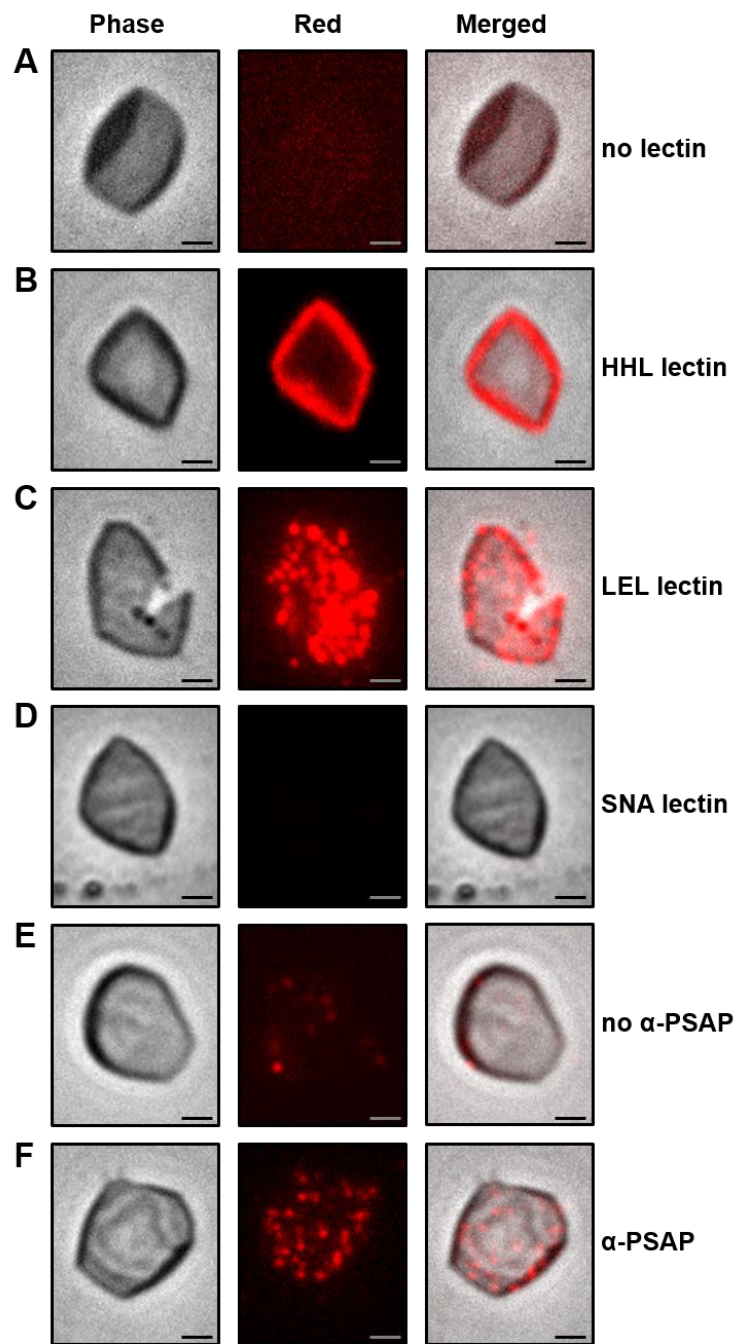


Figure 44 Lectin blot and western blot detection of PSAP from shell material. (A) HHL lectin blot. (B) LEL lectin blot. (C) SNA lectin blot. (D) Western blot detection of PSAP from shell material and a *B. braunii* race B total protein extract using the α-PSAP antibody. WB, western blot; CBS, Coomassie blue stain of WB; TP, total protein.

Figure 45 Detection of PSAP in shell material by fluorescence microscopy using PSAP glycan specific lectins and the α -PSAP antibody (α -PSAP). (A) Negative control. Shell material was incubated without a lectin followed by incubation with streptavidin-Texas red (1:1,000) and analyzed by fluorescence microscopy. (B) PSAP detection in shell material using the HHL lectin. Shell material was incubated with HHL (1:5,000) followed by incubation with streptavidin-Texas red (1:1,000) and fluorescence microscopy. (C) PSAP detection in shell material using the LEL lectin. Shell material was incubated with LEL (1:250) followed by incubation with streptavidin-Texas red (1:1,000) and fluorescence microscopy. (D) Lectin negative control. Shell material was incubated with the SNA lectin (1:500) followed by incubation with streptavidin-Texas red (1:1,000) and fluorescence microscopy. (E) Negative control for α -PSAP antibody detection of PSAP. Shell material was incubated without α -PSAP followed by incubation with α -rabbit-Texas red (1:1,000) and fluorescence microscopy. (F) PSAP detection in shell material using the α -PSAP antibody. Shell material was incubated with α -PSAP (1:100) followed by incubation with α -rabbit-Texas red (1:1,000) and fluorescence microscopy. Scale bars, 2 μ m.



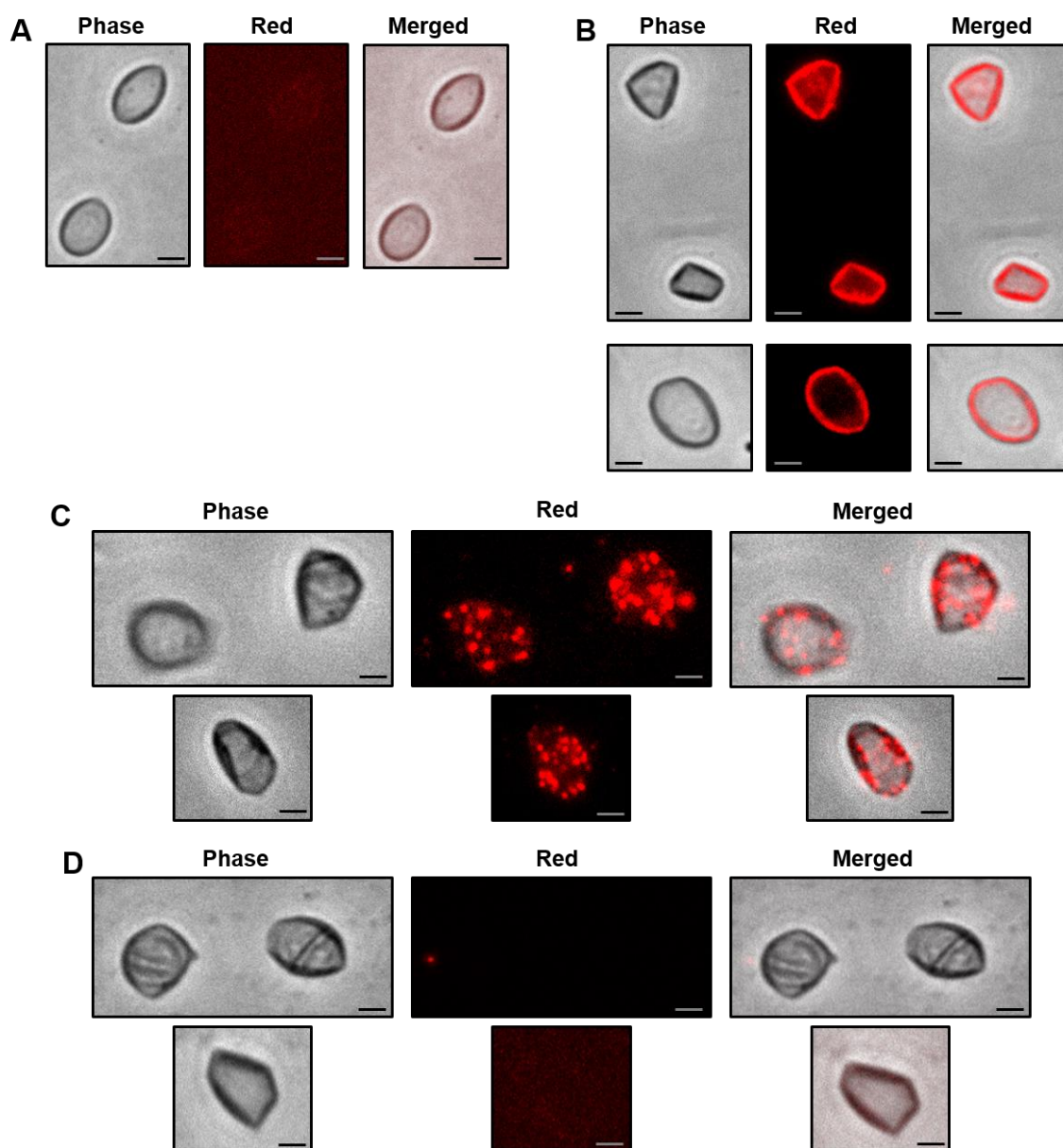


Figure 46 Detection of PSAP in shell material by fluorescence microscopy using lectins specific for PSAP glycans. (A) Negative control; no lectin used. (B) HHL lectin, 1:5,000; streptavidin-Texas red, 1:1,000. (C) LEL lectin, 1:250; streptavidin-Texas red, 1:1,000. (D) Negative control; SNA lectin, 1:500; streptavidin-Texas red, 1:1,000. Scale bar, 2 μm.

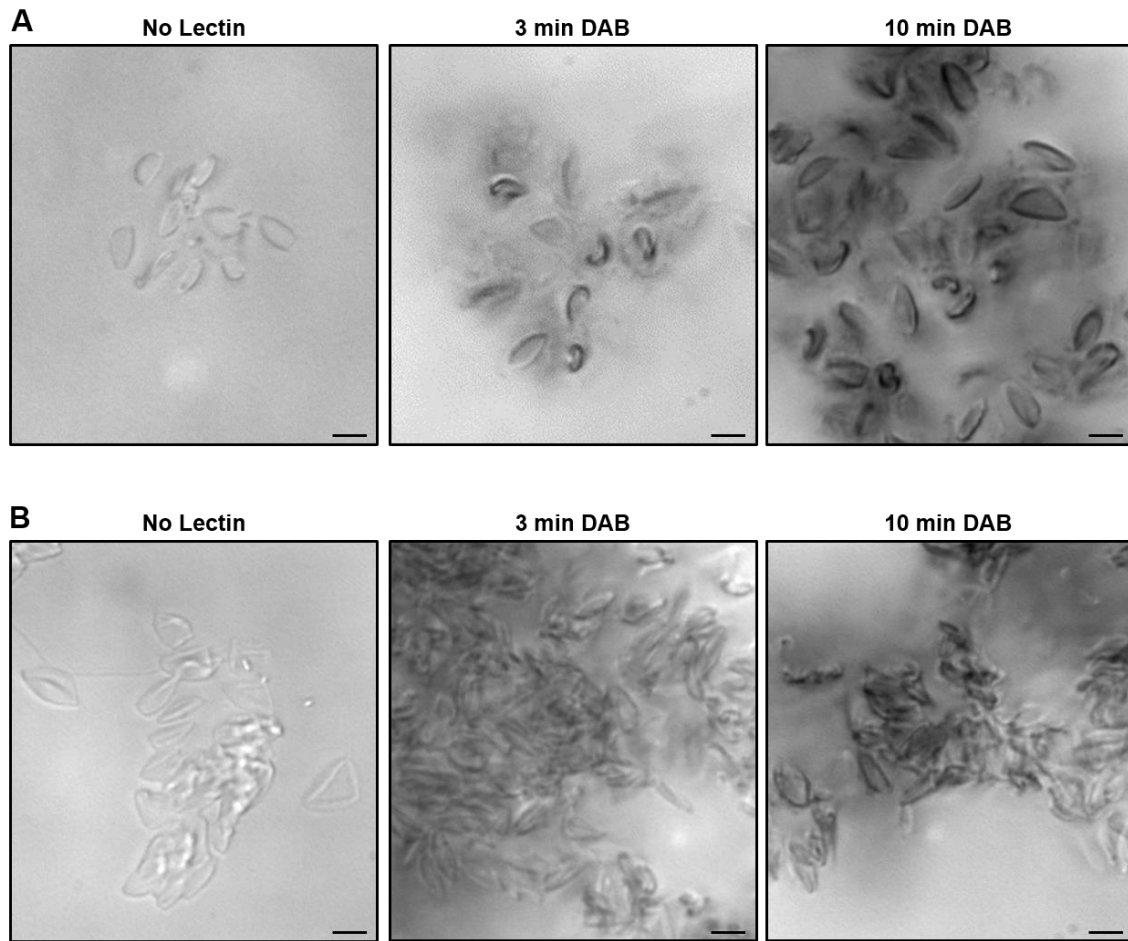


Figure 47 Light microscopy detection of PSAP in shell material using lectins specific for PSAP glycans and DAB. (A) PSAP localization in shells using the HHL lectin. Left panel, control with no lectin; middle panel, 3 min DAB exposure; right panel, 10 min DAB exposure. (B) PSAP localization in shells using the LEL lectin. Left panel, control with no lectin; middle panel, 3 min DAB exposure; right panel, 10 min DAB exposure. Scale bare, 5 μ m.

to PSAP in the shell material since PSAP is the only glycosylated protein associated with the shell material. Additionally, it is unlikely the lectins are binding to the polysaccharide fibers associated with the shell material since these polysaccharides do not contain the mannose or GlcNAc sugars identified by the lectins⁵⁸.

An α -PSAP antibody was also developed for use in PSAP location studies. A 16 amino acid peptide located at positions 854 – 869 (Figure 38) was used for antibody development. This region was chosen to avoid any of the identified PSAP glycosylations. The α -PSAP antibody appeared to be very specific, detecting PSAP from shell material and detecting only the PSAP protein from a total protein extract (Figure 44D). Incubation of the α -PSAP antibody with shell material followed by incubation with an α -rabbit-Texas red conjugated secondary antibody and fluorescence microscopy showed labeling of PSAP in small dots on the shell material reminiscent of the protein granules (Figure 45E and F and 48). The α -PSAP antibody was also used in immunogold labeling studies on thin slices of *B. braunii* colonies in order to localize PSAP on the surface of the retaining wall. However, gold labeling was not seen, possibly due to the α -PSAP antibody antigen not being surface exposed under the conditions used for these studies. Taken together, these studies suggest the PSAP protein is localized to the shell material and correlates to the protein granules seen in our previous study⁵⁸.

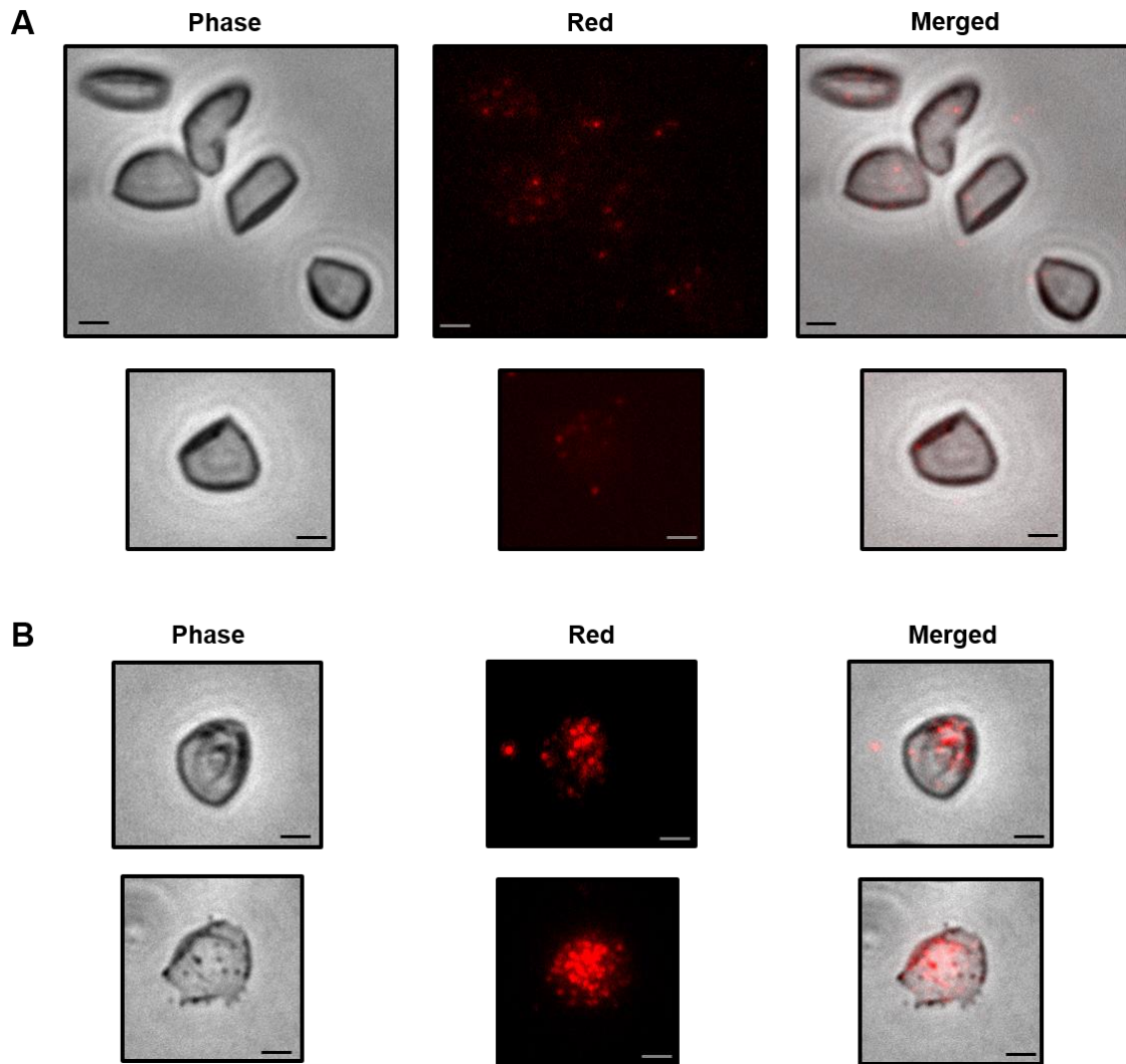


Figure 48 Detection of PSAP in shell material by fluorescence microscopy using the PSAP specific antibody (α -PSAP). (A) Negative control, no α -PSAP, anti-rabbit Texas red 1:1,000. (B) α -PSAP 1:100, anti-rabbit Texas red 1:1,000. Scale bar, 2 μ m.

4.3 Discussion

While the makeup of and potential biosynthesis mechanisms for the cross-linked hydrocarbon network portion of the *B. braunii* ECM has been well studied^{72-74, 186}, and the ECM polysaccharides have been studied in terms of ultrastructure, composition, and production levels^{58, 60, 100, 187, 189, 201-204}, proteins have not previously been identified as part of the *B. braunii* ECM. Here we have identified the unique glycosylated *B. braunii* ECM protein PSAP that is associated with the polysaccharide fibers in the ECM, and may have a role in the biosynthesis of these ECM polysaccharide fibers for this alga.

4.3.1 PSAP is a unique hydroxyproline-rich glycoprotein (HRGP)

We have shown here the *B. braunii* PSAP protein is *N*-glycosylated at ten Asn residues and *O*-glycosylated at four consecutive Hyp residues. In addition to this Hyp-rich region, PSAP contains 2 additional Hyp-rich regions, and thus PSAP is classified as an HRGP. Interestingly, the overall PSAP sequence does not match any other known sequences in the NCBI database at both the DNA and protein levels. Thus, PSAP appears to be a new and unique HRGP. Within the other *B. braunii* races, a PSAP homologue appears to be present in both the A and L races based on screening of unpublished transcriptome sequences generated in our lab. However, a full length PSAP-like transcript is not present in these transcriptomes, and the generation of A and L race genome sequences will be required to positively identify a PSAP homologue from these races.

HRGPs are found within the cell walls and ECMs of photosynthetic organisms ranging from green algae to land plants with a wide variety of roles including cell

expansion and responses to pathogens^{205, 206}. HRGPs are classified into three main categories based on the types of Hyp-rich motifs found in the proteins; arabinogalactan proteins (AGPs) with Ala-Hyp-Ala-Hyp repeats, extensins (EXTs) with Ser-Hyp₃₋₆ repeats, and Pro-rich proteins (PRPs) with Pro-Hyp-Val-Tyr repeats^{79, 91, 205, 206}. In all cases the Hyp residues are *O*-glycosylated to varying levels with arabinose and galactose in the non-contiguous Hyp repeats of AGPs consisting of up to 99% of the protein molecular mass, or with one to four arabinose units in the contiguous Hyp repeats of EXTs constituting 35-65% of the molecular mass^{91, 207}. The Hyp residues in PRPs are minimally (3%-10%) glycosylated with arabinose or are not glycosylated at all²⁰⁵.

While PSAP does not fit well into any of these specific HRGPs categories, PSAP appears to be most related to EXTs. For example, while the PSAP Hyp-rich motifs contain the typical Ser-Hyp₃₋₆ sequence found in EXTs, PSAP also contains shorter Ser-Hyp₂ and longer and Ser-Hyp₉ sequences (Figure 35C). However, none of the Hyp in these domains were found to be glycosylated and the Hyp region that was found to be glycosylated, Gly-Ala-Hyp-Hyp-Hyp-Hyp-Ile-Thr (Figure 41C), does not match the typical EXT glycosylated SerHyp₃₋₆ sequence. However, the identified PSAP Hyp-rich motif is glycosylated with one to four arabinose units as is generally found in EXTs⁹¹. Unlike typical EXT Hyp glycosylations, the PSAP Hyp glycosylations also contain galactose and 6-deoxyhexose (Figure 41C). Additionally, PSAP contains ~18% glycosylations by molecular mass, which is closer to that of PRPs.

This data raises the question whether the Pro residues in the first two PSAP Hyp-rich motifs are actually hydroxylated to Hyp, which in HRGPs typically occurs

posttranslationally, followed by glycosylation⁹¹. Since no glycosylations were found in these Pro rich regions, it is possible they were not converted to Hyp. No peptides covering these regions were found in the original peptide mass fingerprinting to identify the PSAP cDNA (Figure 33B), so it is not known at this time if they are converted to Hyp.

Another difference between PSAP and typical HRGPs is the high degree of *N*-glycosylation found in PSAP. Generally, plant HRGPs do not contain *N*-linked glycans¹⁹⁶, although some AGPs have been shown to have *N*-glycans²⁰⁸. On the other hand, PSAP appears to have the majority of its glycosylations at *N*-linked sites, which is not typical for HRGPs.

The PSAP Hyp *O*-linked glycans show some similarity to those in HRGPs from the green alga *Chlamydomonas reinhardtii*, which have been shown to contain two arabinoses with a terminal galactose²⁰⁹ like that found on PSAP (Figure 41C). However, in *C. reinhardtii* the galactose exists in the furanose configuration, while in PSAP the galactose is in the pyranose configuration (Figure 41C). PSAP is further differentiated from the *C. reinhardtii* *O*-glycans by containing a 6-deoxyhexose (Figure 41C).

4.3.2 PSAP has unique N-linked glycans

Some of the *N*-glycan structures found in PSAP are unusual. One rare feature is the presence of an internal fucose residue, i.e. a fucose that is further glycosylated. Usually, fucose is found as a terminal decoration of the chitobiose core of *N*-glycans²¹⁰. In mammalian glycoproteins, fucose is attached to the 6-position of the inner GlcNAc²¹¹, while in plants and insects fucose is normally attached to the 3-position of the inner

GlcNAc²¹². In fruit fly, fucose can also be attached to both the 3- and the 6-position of the inner GlcNAc¹⁹⁵. Internal fucose has so far only been found in invertebrates, such as *C. elegans*²¹³, *planaria*²¹⁴, and squid²¹⁵. In these cases, the fucose is substituted at *O*-4 with a galactose residue. We have found that *O*-4 of the core fucose in PSAP *N*-glycans is instead substituted by the pentose arabinose in the pyranose form (glycan 3 in Figure 40A). While xylose, another pentose sugar, is common in plant *N*-glycans^{196, 212}, arabinose has only been found in a few plant *N*-glycans²¹⁶, including carrot²¹⁷ and tomato²¹⁸. In summary, while there are few reports presenting *N*-glycans with internal fucose or *N*-glycans with arabinose, to our knowledge, this is the first time an *N*-glycan with an arabinose-fucose disaccharide attached to the chitobiose core has been identified.

4.3.3 Potential functional roles for PSAP

While PSAP appears to be an EXT-like protein, PSAP function is likely to be quite different from EXT proteins, which are involved in cell wall formation and expansion⁷⁹, since the PSAP protein is not associated with the cell wall, but rather the ECM. Thus, PSAP may be involved in the formation of the retaining wall or the polysaccharide fibers that extend from the retaining wall. When these studies began, we hypothesized PSAP was the enzyme responsible for forming the polysaccharide fibers by polymerizing sugar monomers or oligomers. This is supported by our finding that a paste-like substance, presumably sugar monomers/oligomers, were found to enter the cell side of the retaining wall near the PSAP protein granules and exit the protein granule layer as the polysaccharide fibers⁵⁸. This is further supported by the predicted PSAP sugar binding

domain (Figure 35C) that may bind the monomers/oligomers during polymerization. It is also possible PSAP does not have enzymatic activity and functions as an anchor for the polysaccharide fibers. Given that PSAP is the only protein associated with the shell material it may serve both roles. Many attempts to obtain purified PSAP protein for enzymatic studies were not successful. Purification of PSAP from shell material without a denaturing environment (i.e. no SDS) or attempts at expressing PSAP in heterologous hosts such as *E. coli*, *Saccharomyces cerevisiae*, *Nicotiana benthamiana*, and *Chlamydomonas reinhardtii* were unsuccessful. An alternative of expressing only the predicted PSAP sugar binding domain in *E. coli* or *S. cerevisiae* for *in vitro* characterization was also not successful.

This suggests PSAP may be an intrinsically disordered protein due to the Hyp-rich motifs, which is another characteristic of HRGPs²⁰⁵. The PSAP disordered regions may be leading to difficulties in heterologous expression. Indeed, an analysis of PSAP for disordered regions found that the Hyp-rich motifs are predicted to be highly disordered (Figure 49).

Thus, identifying a specific function for PSAP will require refinements in purifying PSAP from *B. braunii* shell material or in recombinant expression. Understanding PSAP function through knockdown, knockout, or overexpression in *B. braunii* is not possible since *B. braunii* transformation is not possible at this time.

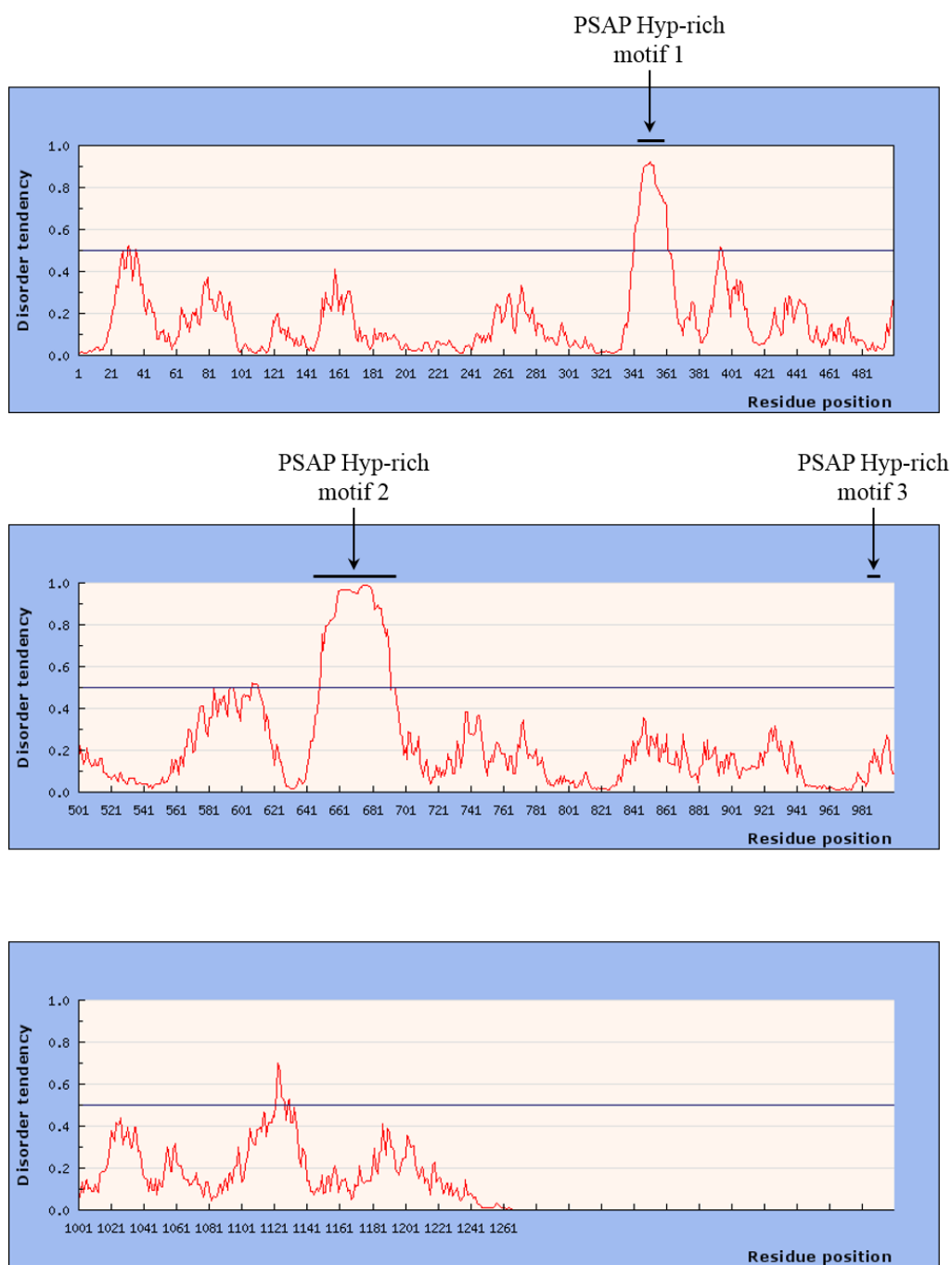


Figure 49 Identification of disordered regions in the PSAP proteins sequence. The PSAP protein sequence was analyzed by the IUPred web server (<http://iupred.enzim.hu/>) for the prediction of intrinsically unstructured regions in proteins. A value above the 0.5 line indicates disordered regions within the protein.

4.4 Experimental procedures

4.4.1 Identifying the PSAP full length cDNA

PSAP protein was extracted from 3.0 mg of lyophilized shell material by adding 200 μ l of extraction buffer (1 mM NaCl, 10 mM Tris, pH 6.8, 1x protease inhibitors) and 200 μ l of 2x SDS-PAGE sample buffer to the shell material, and placing the sample at 95°C for 5 min. A 4-15% gradient SDS-PAGE gel was used to separate 20 μ l of the extract, the protein bands visualized using Coomassie blue, the PSAP band excised from the gel, and the samples in-gel digested with trypsin overnight. The resulting peptides were separated and analyzed for mass and sequence identity by liquid chromatography-mass spectrometry (LC-MS/MS) at the University of Texas at San Antonio Health Science Center Institutional Mass Spectrometry Core Laboratory.

The identified peptides were queried against the *B. braunii* race B transcriptome¹⁹⁰ using TBLASTN and contigs 11859, 10353, and 43184 were found to contain matches to the peptides. These contigs were then queried against the race B genome³⁹ and the 401 kb scaffold #141 was identified to contain sequence matches to the contigs. To determine the full gene structure of PSAP, the race B RNA-seq reads were aligned to scaffold 141 using TopHat²¹⁹ and assembled into transcripts using Cufflinks²²⁰. This approach allowed identification of the 5'- and 3'- UTRs, the intron-exon boundaries, and the open reading frame for PSAP within the scaffold. The PSAP cDNA sequence has been deposited in GenBank, accession number MF36074.

Conserved domains within the PSAP protein were identified using the MyHits protein motif scan tool²²¹, the PSAP signal peptide was identified using the TatP²²²,

Phobius²²³, and SignalP²²⁴ signal peptide prediction tools, and secretion pathway prediction was analyzed using the YLoc subcellular prediction tool^{197, 198}.

4.4.2 In-gel glycoprotein staining

PSAP protein was extracted from shell material as described above. The extract was then separated by 8 % SDS-PAGE as described previously⁵⁸. The glycosylation status of PSAP in the gel was then analyzed using the Pro-Q Emerald 300 glycoprotein gel and blot stain kit (ThermoFisher) following the including instructions. Briefly, the SDS-PAGE gel containing PSAP was immersed in the fixing solution at room temperature for 45 min followed by one wash in the wash solution for 15 min with gentle agitation. The gel was then incubated in the oxidizing solution for 30 min and washed for 15 min. For staining, the gel was immersed in Pro-Q Emerald 300 staining solution while agitating for 2 hr followed by washing twice for 15 min. The gel was then visualized under UV light using a Bio-Rad ChemiDoc XRS.

4.4.3 PSAP N-linked glycan site mapping

N-linked site mapping for PSAP was determined based on previously described methods²²⁵. Coomassie blue-stained SDS-PAGE gel slices corresponding to PSAP were cut into ~1 mm³ pieces, and destained alternately with 40 mM ammonium bicarbonate and 100% acetonitrile until the color turned clear. This was followed by re-swelling the gel slices in 10 mM DTT and 50 mM ammonium bicarbonate at 55 °C for 1 hr. The DTT solution was exchanged with 55 mM iodoacetamide and incubated in the dark for 45 min

followed by washing twice alternately with 40 mM ammonium bicarbonate and 100% acetonitrile. The dehydrated gel was re-swelled with chymotrypsin in 50 mM ammonium bicarbonate on ice for 45 min, followed by protein digestion at 37 °C overnight. Peptides and the glycopeptides were extracted from the gel slices with successive additions of 20% acetonitrile in 5% formic acid, 50% acetonitrile in 5% formic acid, and 80% acetonitrile in 5% formic acid. The three collected solutions were dried and combined into one tube. The chymotrypsin in the sample was inactivated, and deglycosylation was carried out with 2 µL of PNGase A (Calbiochem) in 36 µL of ^{18}O -labeled water (H_2^{18}O) and 2 µL of 0.5 M citrate phosphate buffer pH 5.0. The sample was dried and resuspended in nanopure dH_2O with protease(s) to remove any possible C-terminal incorporation of ^{18}O from residual protease activity²²⁶. The sample was then dried and analyzed by mass spectrometry.

The peptides were analyzed by LC-MS/MS as previously described²²⁵. Briefly, an LTQ Orbitrap XL mass spectrometer (ThermoFisher) equipped with a nanospray ion source was used. One third of the labeled peptides was resuspended in 100 µl mobile phase A (0.1% formic acid in dH_2O) and filtered with 0.2 µm filters (Nanosep, PALL). The sample was then loaded onto a nanospray tapered capillary column/emitter (360 x 75x 15 µm, PicoFrit, New Objective) self-packed with C_{18} reverse-phase resin (10.5 cm, MICHROM Bioresources Inc.) in a nitrogen pressure bomb for 10 min at 1,000 psi (~5 µl load) and then separated *via* a 160 min linear gradient of increasing mobile phase B (80% acetonitrile, 0.1% formic acid in dH_2O) at a flow rate of ~500 nl/min with direct injection into the mass spectrometer. The LTQ was run in the automatic mode collecting a MS scan followed by data dependent MS/MS scans of the six highest abundant precursor ions.

The resulting data were analyzed manually as well as *via* a software search. For the software search, the raw data was analyzed against the PSAP protein sequence using the TurboSequest algorithm (Proteome Discoverer 1.4, ThermoFisher). The SEQUEST parameters were set to allow 10.0 ppm of precursor ion mass tolerance and 0.8 Da of fragment ion tolerance with monoisotopic mass. Digested peptides were allowed with up to three missed internal cleavage sites. The differential modifications of 57.02146 Da, 15.9949 Da and 2.98826 Da were allowed for alkylated cysteine, oxidation of methionine, and ^{18}O -labeled aspartic acid, respectively.

4.4.4 Free N-linked glycan preparation for linkage analysis

Free *N*-linked glycans were prepared from shell material based on previous studies¹⁹⁵ with slight modifications as follows. Shell material was homogenized using a dounce homogenizer with methanol:water (1:1 v/v) on ice. Lipids were extracted by adjusting the solvent mixture to give chloroform:methanol:water (4:8:3 v/v). The insoluble proteinaceous material was collected by centrifugation and the extraction procedure was repeated three times. The final pellet of insoluble protein was washed with cold acetone:water (4:1 v/v) two times, washed with cold acetone once, and dried under nitrogen. The sample was then reduced with 5 mM DTT for 1 h at 55 °C and carboxyamidomethylated with 15 mM iodoacetamide in the dark at room temperature for 45 min. The modified protein was dialyzed with a 4 kDa cutoff membrane (Millipore) against nanopure dH_2O at 4 °C overnight, dried in a SpeedVac concentrator, dissolved in 50 mM ammonium bicarbonate, and digested with chymotrypsin at 37 °C overnight. The

sample was then heated at 100 °C for 5 min to inactivate chymotrypsin, centrifuged at 2,000 x *g* at 4 °C for 15 min, and the supernatant collected and dried. The sample was then passed through a C₁₈ Sep-Pak cartridge, washed with 5% acetic acid to remove contaminants (salts, free sugar, etc.), the peptides and glycopeptides eluted in series with 20% isopropanol in 5% acetic acid, 40% isopropanol in 5% acetic acid, and 100% isopropanol. All eluates were then combined. The *N*-glycans were released with PNGase A (Calbiochem) at 37 °C overnight. After digestion, the sample was passed through a C₁₈ Sep-Pak cartridge, the released *N*-glycans eluted with 5% acetic acid, and dried by lyophilization.

4.4.5 N-linked glycan linkage analysis

For determination of sugar linkages, the free *N*-linked glycans were converted to partially methylated alditol acetates (PMAAs) as previously described²²⁷. Briefly, the samples were fully permethylated as previously described²²⁸, hydrolyzed with HCl:water:acetic acid (0.5:1.5:8 v/v) at 80 °C for 18 h, reduced with 1% NaBD₄ in 30 mM NaOH overnight, and acetylated with acetic anhydride/pyridine (1:1, v/v) at 100 °C for 15 min. The PMAAs were then separated and analyzed by gas chromatography-mass spectrometry (GC-MS) using an Agilent 7890A GC interfaced to a 5975C mass selective detector (MSD) in electron impact ionization mode. The GC separation of the PMAAs was performed on a 30 m SP2331 fused silica capillary column (Supelco) for neutral sugar derivatives, whereas a 30 m DB-1 fused silica capillary column (Agilent) was used for neutral and amino sugar derivatives.

4.4.6 N-linked oligosaccharide profiling

MS profiling of *N*-linked oligosaccharides was performed as described previously^{229, 230}. Briefly, the released *N*-glycans were permethylated and analyzed directly on MS instruments to identify size and structural details. Matrix-assisted laser desorption/ionization-time of flight-mass spectrometry (MALDI-TOF-MS) was performed in the reflector positive ion mode using α -dihydroxybenzoic acid (DHBA, 20 mg/ml solution in 50% methanol:water) as a matrix. The spectrum was obtained using an AB SCIEX TOF/TOF 5800 (AB SCIEX). Nanospray ionization MS/MS (NSI-MSn) analysis was performed using an LTQ Orbitrap XL mass spectrometer (ThermoFisher) equipped with a nanospray ion source. The permethylated samples were dissolved in 1 mM NaOH in 50% methanol and infused into the instrument at a constant flow rate of 0.5 μ l/ min. A full Fourier transform mass spectrometry (FTMS) spectrum was collected at a resolution of 30,000. The capillary temperature was set at 210 °C and MS analysis was performed in the positive ion mode. For total ion mapping (automated MS/MS analysis), the m/z range of 600 to 2,000 was scanned in ITMS mode in successive 2.8 mass unit windows that overlapped the preceding window by 2 mass units.

4.4.7 Hydroxyproline O-linked glycan analysis

For the analysis of hydroxyproline *O*-glycosylation, PSAP was extracted from 6 mg of shell material by boiling the material with SDS-PAGE buffer as described above. After protein extraction, Cys residues were carbamidomethylated and the protein was precipitated with acetone. The protein precipitate was washed with chloroform:methanol

(2:1 v/v) to remove lipids, resuspended in Tris-HCl pH8.2 with 10 mM CaCl₂, and digested with Pronase (Roche) at 37 °C for 48 h. The sample was dried, an aliquot dissolved in 200 µL of mobile phase A (0.1% formic acid in dH₂O), and the sample passed through a 0.2 µm filter (Nanosep, PALL).

For the LC-MS analysis of the pronase digest, the sample was separated on a nano-C₁₈ column (Acclaim pepMap RSLC, 75 µm x 150 mm, 2 µm particle size) via a 80 min gradient of increasing mobile phase B (80% acetonitrile, 0.1% formic acid in dH₂O) at a flow rate of 300 nl/min into the mass spectrometer. MS analysis was performed on an Orbitrap-Fusion equipped with an EASY nanospray source and Ultimate3000 autosampler LC system (ThermoFisher). For online MS detection, full MS data was first collected at a resolution of 60,000 in FT mode. MS/MS-CID data and MS/MS-HCD activation data (both in FT mode) were obtained for each precursor ion by data dependent scan (top speed scan, 3s). The resulting data were analyzed manually.

4.4.8 Functional analysis of the PSAP secretion signal peptide

The synthetic gene product, SP_{PSAP}-GFP, containing the PSAP signal peptide (SP) fused to the N-terminus of GFP was synthesized with *S. cerevisiae* codon optimization (ThermoFisher), inserted into the pMA-T plasmid using the *EcoRI* and *NotI* restriction sites, and subcloned into pESC-TRP using the same restriction sites for expression under the control of the inducible GAL10 promoter. Additionally, a GFP:pESC-TRP construct was generated for control experiments. The constructs were transformed into *S. cerevisiae* strain CKY457 (*MATa*, *leu2Δ1*, *ura3-52*, *trp1Δ63*, *his3Δ200* and *lys2-128δ*) via lithium

acetate transformation, positive transformants were grown at 30 °C in 40 ml SC-TRP media to $O.D_{600} = 0.6$, protein expression induced by replacing media with fresh media containing 2% galactose, and the cells grown for an additional 8 h. The cells were then centrifuged at 1,500 x g for 30 min at room temperature and both the cell pellet and supernatant (media) were saved for protein extraction. To extract protein from the cells, the cell pellet was resuspended in 600 μ L lysis buffer (0.1 M NaOH, 2% β -mercaptoethanol, 2% SDS, and 0.05M EDTA), incubated at 95 °C for 10 minutes, 4.56 μ L of 4M acetic acid added, incubated at 95 °C for 10 min, 150 μ L of loading dye added (0.25 M Tris-HCl pH 6.8, 50% glycerol, and 1x bromophenol blue), incubated at 95 °C for 10 minutes, and centrifuged at 3,380 x g for 2 min at room temperature. The collected media for each sample was run through a 0.22 μ m filter (Millipore Express PLUS Membrane, EMD Millipore) to eliminate any *S. cerevisiae* cells in the media and then lyophilized. For protein extraction from the lyophilized media, 2 ml dH₂O was added to each sample, the samples run through a 10 kDa cutoff centrifugal filter (Amicon Ultra, EMD Millipore) to eliminate reagents and concentrate the samples to ~60 μ L, and 20 μ L of 4x SDS-PAGE sample buffer added to each sample. Twenty-five microliters of the cell protein extract and all of the media protein extract were separated by 12% SDS-PAGE, transferred to a PVDF membrane, and analyzed by western blot following standard procedures. The α GFP-HRP antibody (1:3,000 for cell protein extracts and 1:1,500 for media protein extracts, Santa Cruz Biotechnology) was incubated with the membrane at 4 °C overnight with rotation. Detection and visualization were carried out using the

Amersham ECL Prime Western Blotting Detection Reagent and an Amersham Imager 600RGB, respectively.

4.4.9 Lectin blot

PSAP was extracted from shell material, separated by SDS-PAGE, transferred to a PVDF membrane as described above, and detection using lectins followed standard western blot procedures. The biotinylated lectin (10 µg) was first incubated with the membrane for 30 min at room temperature, followed by treatment with the VectorLabs VECTASTAIN Elite ABC HRP Kit following included instructions. Detection and visualization were carried out using the Amersham ECL Prime Western Blotting Detection Reagent (GE Healthcare) and an Amersham Imager 600RGB (GE Healthcare), respectively. The biotinylated lectins used for this analysis were *Hippeastrum* hybrid lectin (HHL), *Lycopersicon esculentum* lectin (LEL), and *Sambucus nigra* lectin (SNA), all from VectorLabs.

4.4.10 PSAP peptide antibody generation and western blot analysis

A polyclonal PSAP specific antibody was raised in rabbit by GenScript using the PSAP peptide 854-TGVRRRIAPPGTGFGD-869 with a Cys added at the N-terminus for conjugation to a carrier protein. Specificity of the antibody was tested by western blot using PSAP protein extracted from shell material as described above and a *B. braunii* total protein extract. The total protein extract was prepared by adding 350 µl of the extraction buffer (50 mM MOPS pH 7.3, 20 mM MgCl₂, 5 mM β-mercaptoethanol, 5 mM EGTA, 20%

glycerol) to 200 mg of race B tissue, mixing well, bead beating for 10 min, and removing the cell debris by centrifuging at 10,000 x *g* for 10 min at 4 °C. Following 12% SDS-PAGE, proteins were transferred to a PVDF membrane and western blot analysis followed standard procedures. The PSAP antibody (1:500) was incubated with the membrane at 4 °C overnight with rotation, followed by incubation with a goat anti-rabbit-HRP secondary antibody (1:3,000; ThermoScientific) for 2 h at room temperature. Detection and visualization were carried out using the Amersham ECL Prime Western Blotting Detection Reagent and an Amersham Imager 600RGB, respectively.

4.4.11 PSAP localization to shell material

Approximately 100 µg of lyophilized shell material was rehydrated in 200 uL PBS, centrifuged at 15,000 x *g*, and the PBS discarded. The shell material was then blocked with 5% dry milk for 30 min, washed with PBS three times, and incubated with the biotinylated HHL (1:5,000, 30 min, 25 °C), LEL (1:250, 30 min, 25 °C), and SNA (1:500, 30 min, 25 °C) lectins, and the PSAP antibody (1:100, 12 h, 4 °C). The shell material was then washed three times for 5 min at 25 °C with PBS + 0.05% Tween 20 (PBST).

Detection of the biotinylated lectins was carried out by incubating the shell material with a streptavidin-Texas red conjugate (VectorLabs) in PBS (1:1,000) for 30 min in the dark, washing three times with PBST, and imaging on a Nikon Eclipse Ti inverted epifluorescence microscope using a 100X objective (Plan Fluo, NA 1.40, oil immersion) with a 2.5 TV relay lens. A mercury lamp was used as the light source (X-Cite 200DC, Excelitas Technologies) within a cage incubator (InVivo Scientific) at 30 °C,

and images acquired using a cooled EMCCD (electron multiplying charge-coupled device) camera (iXon3 897, Andor, Belfast, Ireland). The images were acquired as phase-contrast (200 ms) and Texas red cube (200 ms).

The HHL and LEL lectins were also detected using diaminobenzidine (DAB). Following incubation with the lectins as above, the shell material was washed three times for 5 min at 25 °C with PBST. Ten microliters of Reagent A and ten microliters of Reagent B from the VECTASTAIN Elite ABC HRP Kit were added to 1 ml of PBS buffer, mixed well with the shell material, and allowed to stand for 30 min. The shell material was then washed three times with PBST, 550 µl of a DAB solution (25 µl of 1 % DAB in dH₂O, 25 µl of 0.3 % H₂O₂, 500 µl PBS, pH 7.2) was added, mixed well, and incubated for 3 min or 10 min at 25 °C. The shell material was then visualized using a Zeiss Axiophot light microscope equipped with a Plan Neofluar 100/1.3 oil immersion objective and a Coolsnap CF monochrome CCD camera (Photometrics) controlled by MetaView version 5.2 software (Molecular Devices) at the Texas A&M University Microscopy and Imaging Center.

For detection of the PSAP antibody, following incubation with the shell material as described above the sample was washed three times for 5 min at 25 °C with PBST. The sample was then incubated with a mouse anti-rabbit-Texas red conjugated secondary antibody (1:1,000) for 2 h at room temperature followed by three washes with PBST. The shell material was then imaged using the fluorescence microscope described above.

To visualize unstained shell material under transmission electron microscopy, lyophilized shell material was resuspended in water, applied to a carbon support film on

an EM grid, dried, and imaged on the 1200 EX transmission electron microscope (JEOL USA, Inc.). Images were captured on a 15C CCD camera (SIA).

5. CONCLUSIONS AND FUTURE DIRECTIONS

Though *B. braunii* produces large amounts of petroleum-equivalent liquid hydrocarbons, there is still little known about the general biology of *B. braunii* at the molecular level. The studies in this report add new structural insights and knowledge into understanding the diversity of race B hydrocarbons as well as the ECM and biosynthesis of polysaccharide fibrils in race B of *B. braunii*, which might potentially contribute to making *B. braunii* a cost effective oil producer in the future.

5.1 Hydrocarbon diversity of *B. braunii* race B

Previous studies have reported approximately twenty different botryococcene molecules ranging in carbon number from 30 to 37, which are based on the C₃₁, C₃₂, C₃₃, and C₃₄ botryococcene structures, and several methylated squalenes. In this study, we identified a novel cyclic C₃₃ botryococcene, C₃₃-1 cyclic botryococcene, and an isomer based on the trimethylsqualene structure, C₃₃-2 trimethylsqualene isomer, from the Showa strain of the B race of *B. braunii* (Figure 8). Using extensive 1D and 2D NMR experiments such as HSQC, HMBC, COSY, HSQC-TOCSY, and H2BC, we have described that C₃₃-1 cyclic botryococcene has a new scaffold not reported in the literature for any botryococcenes as the six-membered ring is attached to the molecule backbone at position C-17, while the same ring is attached at position C-16 in the previously reported botryococcenes (Figures 8 and 11). In addition, we have isolated and characterized two more C₃₃ cyclic botryococcenes, C₃₃-2 cyclic botryococcene and C₃₃-3 cyclic

botryococcene (Figure 8). Though the presence of these molecules in race B was reported previously, a structural characterization by NMR had not been done until our study. Additionally, 1D NOESY experiments and *J*-coupling values indicated the relative stereochemistry of the cyclohexane ring in the C₃₃ botryococcene isomers to be in the stable chair conformation (Figure 17). Here, we have shown that discrete botryococcene and methylsqualene isomers were isolated by using various analytical methods such as silver nitrate-coated silica-based HPLC column that is known to separate similar hydrocarbons to homogeneity and structural characterization of the new molecules was done by various 1D and 2D NMR techniques.

Overall, these findings provide proof of the impressive capacity of *B. braunii* race B to produce a variety of botryococcene and squalene isomers. The findings also provide valuable information that add to our knowledge about *B. braunii* biology at molecular level. Furthermore, more hydrocarbon purification attempts using different strains of *B. braunii* race B and growing these strains under different conditions along with the improvements in the analytical methods might reveal even more botryococcene and methylsqualene isomers.

5.2 *In vitro* Raman spectroscopy analysis of race B hydrocarbons

Distinguishing between structurally similar molecules by spectroscopy techniques such as the *B. braunii* race B hydrocarbons is challenging. Here we have shown that experimental and calculated Raman spectroscopy can be used to clearly differentiate between several of these hydrocarbons. The bond frequency assignments for ten purified

and characterized botryococcene molecules along with five methyl squalenes have been discussed in this report (Figure 25). Frequency assignments for seven of these botryococcene molecules, C₃₁ showacene **2**, C₃₁ wolficene **4**, C₃₂ braunicene **6**, cyclic C₃₃-1 botryococcene **7**, cyclic C₃₃-2 botryococcene **8**, cyclic C₃₃-3 botryococcene **9**, C₃₄ botryococcene isomer **10**, and the five methylsqualenes, monomethylsqualene **12**, dimethylsqualene **13**, trimethylsqualene **14**, trimethylsqualene isomer **15**, and tetramethylsqualene **16**, were not done until this study (Figure 25). Spectral comparison of six botryococcenes, **2**, **4**, **7**, **8**, **9**, and **10** by line-scan Raman spectroscopy indicated that the cyclic botryococcenes possess an obvious Raman band in the 630 - 665 cm⁻¹ region corresponding to the cyclohexane b(Ring) bending that is not present in the acyclic botryococcenes (Figure 26A), which is strongly correlated to the calculated spectra (Figure 26B). Thus, these results have shown that the cyclic botryococcenes can be spectrally differentiated from the acyclic botryococcene molecules. When the cyclic molecules are compared among each other, it can be seen that **7** has a distinct peak at 649 cm⁻¹ that is not present in the other cyclic molecules, **4**, **8**, and **9** (Figure 30A) and the calculated spectra largely agree with the experimental spectra (Figure 30B). This spectral difference in the cyclohexane b(Ring) bending region is not surprising as **7** has a unique ring structure compared to **4**, **8**, and **9** (Figure 25). In addition, two more regions the $\nu(\text{C}=\text{C})$ stretching region between 1600 and 1700 cm⁻¹ and the CH₂/CH₃ bending, $\omega_i(\text{CH})$ wag (in-plane), and $\nu(\text{C}-\text{C})$ stretching region between 1250 and 1400 cm⁻¹ revealed several differences when these six botryococcenes were compared (Figures 26A, 28A, and 29A), which is also in agreement with the calculated spectra (Figures 26B, 28B, and 29B).

Similarly, the five methylsqualene molecules, **12**, **13**, **14**, **15**, and **16**, show differences in both the $\nu(\text{C}=\text{C})$ stretching and CH_2/CH_3 bending regions when compared to each other, especially the band ratios of 1670 cm^{-1} and 1645 cm^{-1} in the $\nu(\text{C}=\text{C})$ stretching region is distinct for each methylsqualene and can be used to differentiate the methylsqualenes (Figures 27A, 28C, and 29C).

We have also used broadband coherent anti-Stokes Raman spectroscopy (BCARS) to spectrally differentiate between the botryococcenes C_{30} botryococcene (**1**), C_{31} isoshowacene (**3**), C_{32} botryococcene (**5**), and C_{32} braunicene (**6**) (Figure 25). Three spectral regions were determined to be used to differentiate these molecules. Each botryococcene was analyzed for the detection of each signal, which is assigned a different color. The Raman band at 1382 cm^{-1} was assigned a green color, the Raman band at $1124 - 1118\text{ cm}^{-1}$ was assigned a red color, and the Raman band at 655 cm^{-1} was assigned a blue color. When the four botryococcene molecules were analyzed for the detection of each signal, it was shown the Raman band at 1382 cm^{-1} is mainly found in the spectrum for **1** (Figure 31D and E), the Raman band at $1124 - 1118\text{ cm}^{-1}$ was mostly detected in **5** (Figure 31D and E), and the Raman band at 655 cm^{-1} that is specific for the cyclohexane b(Ring) bending was only detected in **6** (Figure 31D and E). Analyzing **5** for the detection of any of the three signals revealed an orange color which would be due to mixed detection of the green 1382 cm^{-1} signal and the red $1124 - 1118\text{ cm}^{-1}$ signal (Figure 31D and E).

In this report, we have created a spectral library for the above-mentioned botryococcenes and methylsqualenes and determined specific regions that we used to differentiate these molecules *in vitro*. Future studies will focus on mapping as many as of

these molecules *in vivo* by using their specific Raman signatures determined in this report. This is quite important for several reasons. First of all, particular hydrocarbons may be more suited for certain industrial applications, and thus more desirable. The use of Raman spectroscopy to identify production levels of these particular hydrocarbons within an algal culture would then inform best times for harvest to maximize hydrocarbon production. Second, the physiology of *B. braunii* botryococcene and methylsqualene biosynthesis is not well understood. Even though it has been shown that C₃₀-C₃₃ botryococcenes are mainly found in intracellular oil bodies and C₃₄ botryococcene is mainly found in the ECM, the cellular distribution of C₃₀-C₃₃ botryococcenes and how the botryococcenes are secreted out of the cell are still unknown. In addition, it is not known whether there is a difference in the composition of the different oil bodies inside *B. braunii* cells. For example, some oil bodies may be composed of only a single type of botryococcene while other oil bodies may be a mix of botryococcenes that eventually mature to C₃₄ botryococcene for export to the ECM. Another question of interest that can be answered by *in vivo* mapping using confocal Raman spectroscopy is whether the oil bodies with different hydrocarbon composition localized differentially within the cell, and this is influenced by developmental stage or growth conditions. Overall, these *in vivo* Raman studies would help to understand how the hydrocarbons are secreted to the ECM and add to our knowledge about the *B. braunii* biology at the molecular level and physiology of hydrocarbon production.

Several attempts we made to map some of the above-mentioned botryococcenes *in vivo* have not been successful due to limitations in the current Raman technologies and

methodologies for our studies and goals. However, with the fast growing developments in confocal Raman technology and methodology, these goals could be made possible in the near future.

5.3 Polysaccharide Associated Protein (PSAP) from the ECM of Green Microalga

Botryococcus braunii

We have reported here the biochemical characterization of the polysaccharide associated protein (PSAP). This is the first report related to the protein content of the *B. braunii* ECM. We have found that PSAP contains three hydroxyproline rich domains and a predicted sugar binding domain, but is not similar to any protein when its sequence was searched against the NCBI database. Therefore, PSAP seems to be a unique HRGP. PSAP is a glycoprotein and ten out of fourteen possible *N*-linked glycosylation sites on PSAP harbors glycans mainly formed of mannose and *N*-acetylglucosamine. Further analysis of the *N*-linked glycan structures has revealed that one of these glycans is unique as it contains an arabinose attached to fucose, which have not seen in other algae or plants (Figure 40A). Moreover, four hydroxyproline residues of PSAP were shown to carry short *O*-linked glycans formed of mainly galactose and arabinose, and one of these glycans has an additional 6-deoxyhexose (Figure 41C). We have also confirmed that PSAP is localized to shell material in the ECM using western blot and microscopy analyses (Figures 42 and 45). We suggest two functional roles for PSAP in this report. It either has a role in the biosynthesis of the polysaccharide fibrils or functions as an anchor for the polysaccharide

fibers and since PSAP is the only protein associated with the shell material in the ECM. Additinoally, it could have both functional roles.

The enzymatic activity of PSAP could not be tested because purified PSAP protein was not obtained directly from shell material without using a denaturant such as SDS, or after many heterologous expression attempts in several organisms including *E. coli*, *Saccharomyces cerevisiae*, *Nicotiana benthamiana*, and *Chlamydomonas reinhardtii*, which is possibly due to the intrinsically disordered regions that PSAP possesses (Figure 49). In addition, expression of only the predicted sugar binding domain of PSAP in *E. coli* or *Saccharomyces cerevisiae* was not successful.

Determining a specific function for PSAP *in vitro* might only be possible when purified PSAP protein is obtained either from shell material or heterologous expression in a different host. Identifying the function of PSAP through knockdown, knockout, or overexpression in *B. braunii* will only be possible when *B. braunii* can be transformed.

REFERENCES

- [1] Hannon, M., Gimpel, J., Tran, M., Rasala, B., and Mayfield, S. (2010) Biofuels from algae: challenges and potential, *Biofuels* 1, 763-784.
- [2] Bender, M. H. (2000) Potential conservation of biomass in the production of synthetic organics, *Resources, Conservation and Recycling* 30, 49-58.
- [3] Demirbas, M. F. (2006) Current technologies for biomass conversion into chemicals and fuels, *Energy Sour Part A*, 28, 1181-1188.
- [4] Dincer, K. (2008) Lower Emissions from Biodiesel Combustion, *Energy sources, part A: Recovery, utilization, and environmental effects* 30, 963-968.
- [5] Naik, S. N., Goud, V. V., Rout, P. K., and Dalai, K. A. (2010) Production of first and second generation biofuels: A comprehensive review, *Renewable and Sustainable Energy Reviews* 14, 578-597.
- [6] Sims, R. E., Mabee, W., Saddler, J. N., and Taylor, M. (2010) An overview of second generation biofuel technologies, *Bioresour Technol* 101, 1570-1580.
- [7] Chisti, Y. (2008) Biodiesel from microalgae beats bioethanol, *Trends Biotechnol* 26, 126-131.
- [8] Clark, J. H. (2007) Green chemistry for the second generation biorefinery-sustainable chemical manufacturing based on biomass, *J Chem Technol Biotechnol* 82, 603-609.
- [9] Kröger, M., and Müller-Langer, F. (2014) Review on possible algal-biofuel production processes, *Biofuels* 3, 333-349.

- [10] Chen, C. Y., Yeh, K. L., Aisyah, R., Lee, D. J., and Chang, J. S. (2011) Cultivation, photobioreactor design and harvesting of microalgae for biodiesel production: a critical review, *Bioresour Technol* 102, 71-81.
- [11] Chisti, Y. (2007) Biodiesel from microalgae, *Biotechnol Adv* 25, 294-306.
- [12] Brennan, L., and Owende, P. (2010) Biofuels from microalgae—A review of technologies for production, processing, and extractions of biofuels and co-products, *Renewable and Sustainable Energy Reviews* 14, 557-577.
- [13] Demirbas, A. H. (2009) Inexpensive oil and fats feedstocks for production of biodiesel, *Energy Educ Sci Technol Part A* 23, 1-13.
- [14] Falkowski, P. G., Barber, R. T., and Smetacek, V. V. (1998) Biogeochemical Controls and Feedbacks on Ocean Primary Production, *Science* 281, 200-207.
- [15] 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.
- [16] Sheehan, J., Dunahay, T., Benemann, J., and Roessler, P. (1998) A look back at the US Department of Energy's Aquatic Species Program—biodiesel from algae. National Renewable Energy Laboratory (NREL) report: NREL/TP-580-24190. Golden, CO.
- [17] Borowitzka, M. A. (1999) Commercial production of microalgae: ponds, tanks, tubes and fermenters, *Journal of Biotechnology* 70, 313-321.

- [18] Miron, A. S., Gomez, A. C., Camacho, F. G., Grima, E. M., Chisti, Y. (1999) Comparative evaluation of compact photobioreactors for large-scale monoculture of microalgae, *Journal of Biotechnology* 70, 249-270.
- [19] Pulz, O., and Scheinbenbogan, K. (1998) Photobioreactors: design and performance with respect to light energy input, *Advances in Biochemical Engineering/Biotechnology* 59, 123-152.
- [20] 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.
- [21] Chisti, Y. (2013) Constraints to commercialization of algal fuels, *J. Biotechnol.* 167, 201-214.
- [22] Chisti, Y. (2010) Fuels from microalgae, *Biofuels* 1, 233-235.
- [23] Vaccari, D. A. (2009) Phosphorus: a looming crisis, *Sci Am* 300, 54-59.
- [24] Vance, C. P. (2001) Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources, *Plant Physiol* 127, 390-397.
- [25] Demirbas, A., and Demirbas, M. F. (2011) Importance of algae oil as a source of biodiesel, *Energy Conversion and Management* 52, 163-170.
- [26] 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.
- [27] Beer, L. L., Boyd, E. S., Peters, J. W., and Posewitz, M. C. (2009) Engineering algae for biohydrogen and biofuel production, *Curr Opin Biotechnol* 20, 264-271.
- [28] Sheehan, J. (2009) Engineering direct conversion of CO(2) to biofuel, *Nat Biotechnol* 27, 1128-1129.

- [29] Waltz, E. (2009) Biotech's green gold?, *Nat Biotechnol* 27, 15-18.
- [30] Qin, S., Lin, H., and Jiang, P. (2012) Advances in genetic engineering of marine algae, *Biotechnol Adv* 30, 1602-1613.
- [31] Lü, J., Sheahan, C., and Fu, P. (2011) Metabolic engineering of algae for fourth generation biofuels production, *Energy and Environmental Science* 4, 2451-2466.
- [32] 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.
- [33] Radakovits, R., Jinkerson, R. E., Darzins, A., and Posewitz, M. C. (2010) Genetic engineering of algae for enhanced biofuel production, *Eukaryot Cell* 9, 486-501.
- [34] Kilian, O., Benemann, C. S., Niyogi, K. K., and Vick, B. (2011) High-efficiency homologous recombination in the oil-producing alga *Nannochloropsis* sp, *Proc Natl Acad Sci U S A* 108, 21265-21269.
- [35] Ghosh, A., Khanra, S., Mondal, M., Halder, G., N., T. O., Saini, S., Bhowmick, T. K., and Gayen, K. (2016) Progress toward isolation of strains and genetically engineered strains of microalgae for production of biofuel and other value added chemicals: A review, *Energy Conversion and Management* 113, 104-118.
- [36] Derelle, E., Ferraz, C., Rombauts, S., Rouze, P., Worden, A. Z., Robbens, S., Partensky, F., Degroeve, S., Echeynie, S., Cooke, R., Saeys, Y., Wuyts, J., Jabbari, K., Bowler, C., Panaud, O., Piegu, B., Ball, S. G., Ral, J. P., Bouget, F. Y., Piganeau, G., De Baets, B., Picard, A., Delseny, M., Demaille, J., Van de Peer, Y., and Moreau, H. (2006) Genome analysis of the smallest free-living eukaryote

- Ostreococcus tauri* unveils many unique features, *Proc Natl Acad Sci U S A* 103, 11647-11652.
- [37] Dutcher, S. K., Li, L., Lin, H., Meyer, L., Giddings, T. H., Jr., Kwan, A. L., and Lewis, B. L. (2012) Whole-genome sequencing to identify mutants and polymorphisms in *Chlamydomonas reinhardtii*, *G3 (Bethesda)* 2, 15-22.
- [38] Park, S., Polle, J. E., Melis, A., Lee, T. K., and Jin, E. (2006) Up-regulation of photoprotection and PSII-repair gene expression by irradiance in the unicellular green alga *Dunaliella salina*, *Mar Biotechnol (NY)* 8, 120-128.
- [39] Browne, D. R., Jenkins, J., Schmutz, J., Shu, S., Barry, K., Grimwood, J., Chiniquy, J., Sharma, A., Niehaus, T. D., Weiss, T. L., Koppisch, A. T., Fox, D. T., Dhungana, S., Okada, S., Chappell, J., and Devarenne, T. P. (2017) Draft Nuclear Genome Sequence of the Liquid Hydrocarbon-Accumulating Green Microalga *Botryococcus braunii* Race B (Showa), *Genome Announc* 5.
- [40] Mayfield, S. P., and Kindle, K. L. (1990) Stable nuclear transformation of *Chlamydomonas reinhardtii* by using a *C. reinhardtii* gene as the selectable marker, *Proc Natl Acad Sci U S A* 87, 2087-2091.
- [41] 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.
- [42] Chen, M., Schliep, M., Willows, R. D., Cai, Z. L., Neilan, B. A., and Scheer, H. (2010) A red-shifted chlorophyll, *Science* 329, 1318-1319.

- [43] Savile, C. K., and Lalonde, J. J. (2011) Biotechnology for the acceleration of carbon dioxide capture and sequestration, *Curr Opin Biotechnol* 22, 818-823.
- [44] Wijffels, R. H., Barbosa, M. J., and Eppink, M. H. M. (2010) Microalgae for the production of bulk chemicals and biofuels, *Biofuels, Bioprod. Biorefin.* 4, 287-295.
- [45] Durrett, T. P., Benning, C., and Ohlrogge, J. (2008) Plant triacylglycerols as feedstocks for the production of biofuels, *Plant J* 54, 593-607.
- [46] 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.
- [47] Yoshida, M., Tanabe, Y., Yonezawa, N., and Watanabe, M. M. (2012) Energy innovation potential of oleaginous microalgae, *Biofuels* 3, 761-781.
- [48] Keasling, J. D., and Chou, H. (2008) Metabolic engineering delivers next-generation biofuels, *Nat Biotechnol* 26, 298-299.
- [49] 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.
- [50] Keasling, J. D. (2008) Synthetic biology for synthetic chemistry, *ACS Chem Biol* 3, 64-76.
- [51] Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. (2003) Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids, *Nat Biotechnol* 21, 796-802.

- [52] Lee, S. Y., Kim, H. M., and Cheon, S. (2015) Metabolic engineering for the production of hydrocarbon fuels, *Curr. Opin. Biotechnol.* 33, 15-22.
- [53] Zhang, H., Liu, Q., Cao, Y., Feng, X., Zheng, Y., Zou, H., Liu, H., Yang, J., and Xian, M. (2014) Microbial production of sabinene--a new terpene-based precursor of advanced biofuel, *Microb Cell Fact* 13, 20.
- [54] Hu, Q., Sommerfeld, M., Jarvis, E., Ghirardi, M., Posewitz, M., Seibert, M., and Darzins, A. (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances, *Plant J* 54, 621-639.
- [55] Kasai, Y., Oshima, K., Ikeda, F., Abe, J., Yoshimitsu, Y., and Harayama, S. (2015) Construction of a self-cloning system in the unicellular green alga *Pseudochoricystis ellipsoidea*, *Biotechnol Biofuels* 8, 94.
- [56] Rezanka, T., Zahradnik, J., and Podojil, M. (1982) Hydrocarbons in green and blue-green-algae, *Folia Microbiologica* 27, 450-454.
- [57] Metzger, P., and Largeau, C. (2005) *Botryococcus braunii*: a rich source for hydrocarbons and related ether lipids, *Appl Microbiol Biot* 66, 486-496.
- [58] 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.
- [59] Eroglu, E., and Melis, A. (2010) Extracellular terpenoid hydrocarbon extraction and quantitation from the green microalgae *Botryococcus braunii* var. *Showa*, *Bioresour Technol* 101, 2359-2366.

- [60] Largeau, C., Casadevall, E., Berkloff, C., and Dhamelinourt, P. (1980) Sites of accumulation and composition of hydrocarbons in *Botryococcus braunii*, *Phytochemistry* 19, 1043-1051.
- [61] Moldwan, J. M., and Seifert, W. K. (1980) First discovery of botryococcane in petroleum, *J. C. S. Chem. Commun* 1, 912.
- [62] Mcirdy, D. M., Cox, R. E., Volkman, J. K., and Howel, V. J. (1986) Botryococcane in a new class of Australian non marine crude oils, *Nature* 320, 57.
- [63] 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 distribution and carbon isotope ratios, *Org. Geochem* 22, 1023.
- [64] Behrens, A., Schaeffer, P., Bernasconi, S., and Albrecht, P. (2000) 7,11-cyclobotryococca-5,12,26-triene, a novel botryococcene-related hydrocarbon occurring in natural environments, *Org Lett* 2, 1271-1274.
- [65] 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, 595.
- [66] 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.

- [67] 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.
- [68] 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.
- [69] 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.
- [70] 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 Physiology* 170, 702-716.
- [71] 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.
- [72] Metzger, P., Pouet, Y., Bischoff, R., and Casadevall, E. (1993) An aliphatic polyaldehyde from *Botryococcus braunii* (A race), *Phytochemistry* 32, 875-883.
- [73] Metzger, P., Rager, M. N., and Largeau, C. (2007) Polyacetals based on polymethylsqualene diols, precursors of algaenan in *Botryococcus braunii* race B, *Organic Geochemistry* 38, 566-581.

- [74] Bertheas, O., Metzger, P., and Largeau, C. (1999) A high molecular weight complex lipid, aliphatic polyaldehyde tetraterpenediol polyacetal from *Botryococcus braunii* (L race), *Phytochemistry* 50, 85-96.
- [75] Keegstra, K. (2010) Plant cell walls, *Plant Physiol* 154, 483-486.
- [76] Harholt, J., Suttangkakul, A., and Vibe Scheller, H. (2010) Biosynthesis of pectin, *Plant Physiol* 153, 384-395.
- [77] Scheller, H. V., and Ulvskov, P. (2010) Hemicelluloses, *Annu Rev Plant Biol* 61, 263-289.
- [78] Rose, J. K., and Lee, S. J. (2010) Straying off the highway: trafficking of secreted plant proteins and complexity in the plant cell wall proteome, *Plant Physiol* 153, 433-436.
- [79] Lamport, D. T., Kieliszewski, M. J., Chen, Y., and Cannon, M. C. (2011) Role of the extensin superfamily in primary cell wall architecture, *Plant Physiol* 156, 11-19.
- [80] Chabannes, M., Ruel, K., Yoshinaga, A., Chabbert, B., Jauneau, A., Joseleau, J. P., and Boudet, A. M. (2001) In situ analysis of lignins in transgenic tobacco reveals a differential impact of individual transformations on the spatial patterns of lignin deposition at the cellular and subcellular levels, *Plant J* 28, 271-282.
- [81] Alberts, B., Johnson, A., and Lewis, J. (2002) Cell junctions, cell adhesion, and the extracellular matrix, In *Mol. Biol. Cell* Fourth ed., Garland Science, New York.
- [82] Cooper, G. M. (2000) Cell walls and the extracellular matrix, In *The Cell: A Molecular Approach* Second ed., Sinauer Associates, Sunderland (MA).

- [83] Graham, L. E. (1996) Green algae to land plants: an evolutionary transition, *J. Plant Res.* 109, 7737–7742.
- [84] Popper, Z. A., and Tuohy, M. G. (2010) Beyond the green: understanding the evolutionary puzzle of plant and algal cell walls, *Plant Physiol* 153, 373-383.
- [85] Domozych, D. S., Ciancea, M., Fangel, J. U., Mikkelsen, M. D., Ulvskov, P., and Willats, W. G. (2012) The cell walls of green algae: A journey through evolution and diversity, *Front Plant Sci* 3, 82.
- [86] Umen, J. G. (2014) Green algae and the origins of multicellularity in the plant kingdom, *Cold Spring Harbor perspectives in biology* 6, a016170.
- [87] Domozych, D. S., and Domozych, C. E. (2014) Multicellularity in green algae: upsizing in a walled complex, *Frontiers in plant science* 5, 649.
- [88] Matt, G., and Umen, J. (2016) *Volvox*: A simple algal model for embryogenesis, morphogenesis and cellular differentiation, *Developmental biology* 419, 99-113.
- [89] Sumper, M., and Hallmann, A. (1998) Biochemistry of the extracellular matrix of *Volvox*, *International review of cytology* 180, 51-85.
- [90] Ender, F., Godl, K., Wenzl, S., and Sumper, M. (2002) Evidence for autocatalytic cross-linking of hydroxyproline-rich glycoproteins during extracellular matrix assembly in *Volvox*, *The Plant cell* 14, 1147-1160.
- [91] Hijazi, M., Velasquez, S. M., Jamet, E., Estevez, J. M., and Albenne, C. (2014) An update on post-translational modifications of hydroxyproline-rich glycoproteins: toward a model highlighting their contribution to plant cell wall architecture, *Frontiers in plant science* 5, 395.

- [92] Kieliszewski, M. J., and Lamport, D. T. (1994) Extensin: repetitive motifs, functional sites, post-translational codes, and phylogeny, *The Plant journal : for cell and molecular biology* 5, 157-172.
- [93] Kieliszewski, M. J., and Shpak, E. (2001) Synthetic genes for the elucidation of glycosylation codes for arabinogalactan-proteins and other hydroxyproline-rich glycoproteins, *Cellular and molecular life sciences : CMLS* 58, 1386-1398.
- [94] Chakraborty, T., Sen, A. K., and Pal, R. (2012) Chemical characterization and the stress induced changes of the extracellular polysaccharide of the marine cyanobacterium, *Phormidium tenue.*, *J. Algal Biomass Utiln* 3, 11-20.
- [95] Leung, M. Y., Liu, C., Koon, J. C., and Fung, K. P. (2006) Polysaccharide biological response modifiers, *Immunol Lett* 105, 101-114.
- [96] Parikh, A., and Madamwar, D. (2006) Partial characterization of extracellular polysaccharides from cyanobacteria, *Bioresour Technol* 97, 1822-1827.
- [97] Okutani, K. (1984) Antitumor and immunostimulant activities of polysaccharides produced by a marine bacterium of the genus *Vibrio.*, *Bull. Jpn. Soc. Sci. Fish* 50, 1035-1037.
- [98] Collic Jouault, S., Chevlot, L., Helley, D., Ratiskol, J., Bros, A., Siquin, C., Roger, O., and Fischer, A. M. (2001) Characterization, chemical modifications and in vitro anticoagulant properties of an exopolysaccharide produced by *Alteromonas infernus*, *Biochim Biophys Acta* 1528, 141-151.
- [99] Bender, J., and Phillips, P. (2004) Microbial mats for multiple applications in aquaculture and bioremediation, *Bioresour Technol* 94, 229-238.

- [100] 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, *Journal of Phycology* 26, 258-266.
- [101] Maki, M., and Renkonen, R. (2004) Biosynthesis of 6-deoxyhexose glycans in bacteria, *Glycobiology* 14, 1R-15R.
- [102] Kawachi, M., Tanoi, T., Demura, M., Kaya, K., and Watanabe, M. M. (2012) Relationship between hydrocarbons and molecular phylogeny of *Botryococcus braunii*, *Algal Research* 1, 114-119.
- [103] Demura, M., Ioki, M., Kawachi, M., Nakajima, N., and Watanabe, M. M. (2014) Desiccation tolerance of *Botryococcus braunii* (Trebouxiophyceae, Chlorophyta) and extreme temperature tolerance of dehydrated cells, *J Appl Phycol* 26, 49-53.
- [104] Ajikumar, P. K., Tyo, K., Carlsen, S., Mucha, O., Phon, T. H., and Stephanopoulos, G. (2008) Terpenoids: opportunities for biosynthesis of natural product drugs using engineered microorganisms, *Mol Pharm* 5, 167-190.
- [105] Langenheim, J. H. (1994) Higher plant terpenoids: A phytocentric overview of their ecological roles, *J Chem Ecol* 20, 1223-1280.
- [106] Vranova, E., Coman, D., and Gruissem, W. (2012) Structure and dynamics of the isoprenoid pathway network, *Mol Plant* 5, 318-333.
- [107] 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.

- [108] Sacchettini, J. C., and Poulter, C. D. (1997) Creating isoprenoid diversity, *Science* 277, 1788-1789.
- [109] Dereth, R. P., Jeanne, M. R., Bonnie, B., and Seichi, P. T. M. (2006) Biosynthetic diversity in plant triterpene cyclization, *Curr. Opin. Plant Biol.* 9, 305-314.
- [110] Kuzuyama, T. (2002) Mevalonate and nonmevalonate pathways for the biosynthesis of isoprene units, *Biosci Biotechnol Biochem* 66, 1619-1627.
- [111] Zhao L, C. W., Xiao Y, Liu HW, Liu P. (2013) Methylerythritol phosphate pathway of isoprenoid biosynthesis, *Annual Review Biochemistry* 82, 497-530.
- [112] Robinson, G. W., Tsay, Y. H., Kienzle, B. K., Smith-Monroy, C. A., and Bishop, R. W. (1993) Conservation between human and fungal squalene synthetases: similarities in structure, function, and regulation, *Mol Cell Biol* 13, 2706-2717.
- [113] Xu, W., Ma, X., and Wang, Y. (2016) Production of squalene by microbes: an update, *World Journal of Microbiology & Biotechnology* 32.
- [114] 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.
- [115] Poulter, C. D. (1990) Biosynthesis of non-head-to-tail terpenes – formation of 1'-1 and 1'-3 linkages, *Accounts Chem Res.* 23, 70-77.
- [116] Templier, J., Largeau, C., and Casadevall, E. (1984) Hydrocarbon formation in the green-alga *Botryococcus braunii* .4. Mechanism of non-isoprenoid hydrocarbon biosynthesis in *Botryococcus braunii*, *Phytochemistry* 23, 1017-1028.

- [117] Templier, J., Largeau, C., and Casadevall, E. (1987) Hydrocarbon formation in the green-alga *Botryococcus braunii*. 5. Effect of various inhibitors on biosynthesis of non-isoprenoid hydrocarbons in *Botryococcus braunii*, *Phytochemistry* 26, 377-383.
- [118] Dennis, M. W., and Kolattukudy, P. E. (1991) Alkane biosynthesis by decarbonylation of aldehyde catalyzed by a microsomal preparation from *Botryococcus braunii*, *Archives of Biochemistry and Biophysics* 287, 268-275.
- [119] Dennis, M., and Kolattukudy, P. E. (1992) A cobalt-porphyrin enzyme converts a fatty aldehyde to a hydrocarbon and Co, *Proceedings of the National Academy of Sciences of the United States of America* 89, 5306-5310.
- [120] Wang, X., and Kolattukudy, P. E. (1995) Solubilization and purification of aldehyde-generating fatty Acyl-CoA reductase from green-alga *Botryococcus braunii*, *Febs Letters* 370, 15-18.
- [121] Metzger, P., and Casadevall, E. (1987) Lycopadiene, a tetraterpenoid hydrocarbon from new strains of the green-alga *Botryococcus braunii*, *Tetrahedron Lett.* 28, 3931-3934.
- [122] 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.

- [123] 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.
- [124] Huang, Z., and Poulter, C. D. (1989) Isoshowacene, a C31 hydrocarbon from *Botryococcus braunii* var. Showa, *Phytochemistry* 28, 3043-3046.
- [125] Huang, Z., and Poulter, C. D. (1988) Isobraunicene, wolficene, and isowolficene. New cyclic 1'-3 fused isoprenoids from *Botryococcus braunii*, *J. Org. Chem.* 53, 5390-5392.
- [126] Huang, Z., Poulter, C. D., Wolf, F. R., Somers, T. C., and White, J. D. (1988) Braunicene, a novel cyclic C32 isoprenoid from *Botryococcus braunii*, *J. Am. Chem. Soc.* 110, 3959-3964.
- [127] Okada, S., Murakimi, M., and Yamaguchi, K. (1997) Characterization of hydrocarbons from the Yayoi strain of the green microalga *Botryococcus braunii*, *Phytochemistry* 8, 198-203.
- [128] Metzger, P., Casadevall, E., Pouet, M. J., and Pouet, Y. (1985) Structures of some botryococcenes: Branched hydrocarbons from the B-race of the green alga *Botryococcus braunii*, *Phytochemistry* 24, 2995-3002.
- [129] Wake, L. V., and Hillen, L. W. (1981) Nature and hydrocarbon content of blooms of the alga *Botryococcus braunii* occurring in Australian freshwater lakes, *Aust. J. Mar. Fresh. Res.* 32, 353-367.

- [130] 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.
- [131] Gardiner, D. J., and Graves, P. R. (1989) Introduction to Raman scattering, In *Practical Raman Spectroscopy*, Springer Berlin Heidelberg, Berlin.
- [132] Raman, C. V. (1928) "A new radiation" *Indian J. Phys.* 2, 387–398.
- [133] Ball, D. W. (1994) Rayleigh and Raman scattering, *Spectroscopy* 9, 20-21.
- [134] Schmitt, M., and Popp, J. (2006) Raman spectroscopy at the beginning of the twenty-first century, *Journal of Raman Spectroscopy* 37, 20-28.
- [135] Petry, R., Schmitt, M., and Popp, J. (2003) Raman spectroscopy—A prospective tool in the life sciences, *CHEMPHYSCHEM*. 4, 14-30.
- [136] Labanowski, J. K., and Jan, K. (1991) *Density functional methods in chemistry*, Andzelm J. W. ed., Springer, New York.
- [137] Mahadevan-Jansen, A., and Richards-Kortum, R. R. (1996) Raman spectroscopy for the detection of cancers and precancers, *J Biomed Opt* 1, 31-70.
- [138] Haka, A. S., Shafer-Peltier, K. E., Fitzmaurice, M., Crowe, J., Dasari, R. R., and Feld, M. S. (2005) Diagnosing breast cancer by using Raman spectroscopy, *Proc Natl Acad Sci U S A* 102, 12371-12376.
- [139] Huang, Z., McWilliams, A., Lui, H., McLean, D. I., Lam, S., and Zeng, H. (2003) Near-infrared Raman spectroscopy for optical diagnosis of lung cancer, *Int J Cancer* 107, 1047-1052.

- [140] Gierlinger, N., Keplinger, T., and Harrington, M. (2012) Imaging of plant cell walls by confocal Raman microscopy, *Nat Protoc* 7, 1694-1708.
- [141] Wu, H., Volponi, J. V., Oliver, A. E., Parikh, A. N., Simmons, B. A., and Singh, S. (2011) In vivo lipidomics using single-cell Raman spectroscopy, *Proc Natl Acad Sci U S A* 108, 3809-3814.
- [142] Hosokawa, M., Ando, M., Mukai, S., Osada, K., Yoshino, T., Hamaguchi, H. O., and Tanaka, T. (2014) In vivo live cell imaging for the quantitative monitoring of lipids by using Raman microspectroscopy, *Anal Chem* 86, 8224-8230.
- [143] Qi, J., and Shih, W. C. (2014) Performance of line-scan Raman microscopy for high-throughput chemical imaging of cell population, *Appl Opt* 53, 2881-2885.
- [144] Wei, X., Jie, D., Cuello, J. J., Johnson, D. J., Qiu, Z., and He, Y. (2014) Microalgal detection by Raman microspectroscopy, *TrAC Trends in Analytical Chemistry* 53, 33-40.
- [145] Parab, N. D. T., and Tomar, V. (2012) Raman spectroscopy of algae: A review *J Nanomedic Nanotechnol* 3, 131.
- [146] He, X. N., Allen, J., Black, P. N., Baldacchini, T., Huang, X., Huang, H., Jiang, L., and Lu, Y. F. (2012) Coherent anti-Stokes Raman scattering and spontaneous Raman spectroscopy and microscopy of microalgae with nitrogen depletion, *Biomed Opt Express* 3, 2896-2906.
- [147] Jaeger, D., Pilger, C., Hachmeister, H., Oberlander, E., Wordenweber, R., Wichmann, J., Mussnug, J. H., Huser, T., and Kruse, O. (2016) Label-free in vivo

- analysis of intracellular lipid droplets in the oleaginous microalga *Monoraphidium neglectum* by coherent Raman scattering microscopy, *Sci Rep* 6, 35340.
- [148] Wang, T., Ji, Y., Wang, Y., Jia, J., Li, J., Huang, S., Han, D., Hu, Q., Huang, W. E., and Xu, J. (2014) Quantitative dynamics of triacylglycerol accumulation in microalgae populations at single-cell resolution revealed by Raman microspectroscopy, *Biotechnol Biofuels* 7, 58.
- [149] Lee, T. H., Chang, J. S., and Wang, H. Y. (2013) Rapid and in vivo quantification of cellular lipids in *Chlorella vulgaris* using near-infrared Raman spectrometry, *Anal Chem* 85, 2155-2160.
- [150] 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. Struc.* 1129, 216-221.
- [151] 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.
- [152] Wang, C.-C., Chandrappa, D., Smirnoff, N., and Moger, J. (2015) Monitoring lipid accumulation in the green microalga *Botryococcus braunii* with frequency-modulated stimulated Raman scattering, pp 93291J-93291J-93295.

- [153] Brown, A. C., and Knights, B. A. (1969) Hydrocarbon content and its relationship to physiological state in the green alga *Botryococcus braunii*, *Phytochemistry* 8, 543-547.
- [154] Galbraith, M. N., Hillen, L. W., and Wake, L. V. (1983) Darwinene: A branched hydrocarbon from a green form of *Botryococcus braunii*, *Phytochemistry* 22, 1441-1443.
- [155] Huang, Z., and Poulter, C. D. (1989) Tetramethylsqualene, a triterpene from *Botryococcus braunii* var. *showa*, *Phytochemistry* 28, 1467-1470.
- [156] Gao, M., Simoneit, B. R., Gantar, M., and Jaffe, R. (2007) Occurrence and distribution of novel botryococcene hydrocarbons in freshwater wetlands of the Florida Everglades, *Chemosphere* 70, 224-236.
- [157] David, M., Metzger, P., and Casadevall, E. (1988) Two cyclobotryococenes. from the B race of the green alga *Botryococcus braunii*, *Phytochemistry* 27, 2863-2867.
- [158] Huang, Z., and Poulter, C. D. (1988) Braunicene. Absolute stereochemistry of the cyclohexane ring., *J. Org. Chem.* 53, 4089-4094.
- [159] Huang, Y., Street-Perrott, F. A., Perrott, R. A., Metzger, P., and Eglinton, G. (1999) Glacial-interglacial environmental changes inferred from molecular and compound specific $\delta^{13}\text{C}$ analyses of sediments from Sacred Lake, Mt. Kenya, *Geochim. Cosmochim. Acta* 63, 1383-1404.
- [160] Caron, B., Mark, A. E., and Poger, D. (2014) Some Like It Hot: The Effect of Sterols and Hopanoids on Lipid Ordering at High Temperature, *J Phys Chem Lett* 5, 3953-3957.

- [161] Nonomura, A. M. (1988) *Botryococcus braunii* var. *showa* (Chlorophyceae) from Berkeley, California, United States of America., *Japanese Journal of Phycology* 36, 285-291.
- [162] Grung, M., Metzger, P., and Liaaen-Jensen, S. (1989) Primary and secondary carotenoids in two races of the green alga *Botryococcus braunii*., *Biochem. Syst. Ecol.* 17, 263-269.
- [163] 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.
- [164] Kitazato, H., Asaoka, S., and Iwamoto, H. (1989) Catalytic cracking of hydrocarbons from microalgae, *Sekiyu Gakkaishi* 32, 28-34.
- [165] Tran, N. H., Bartlett, J. R., Kannangara, G. S. K., Milev, A. S., Volk, H., and Wilson, M. A. (2010) Catalytic upgrading of biorefinery oil from micro-algae, *Fuel* 89, 265-274.
- [166] Tatli, M., Naik, M. T., Okada, S., Dangott, L. J., and Devarenne, T. P. (2017) Isolation and characterization of cyclic C₃₃ botryococcenes and a trimethylsqualene isomer from *Botryococcus braunii* Race B, *Journal of natural products*, 80, 953-958.
- [167] Chun, H. J., Weiss, T. L., Devarenne, T. P., and Laane, J. (2013) Vibrational spectra and DFT calculations of squalene, *J. Mol. Struc.* 1032, 203-206.

- [168] Chu, S. P. (1942) The influence of the mineral composition of the medium on the growth of planktonic algae Part 1. Methods and culture media, *J Ecol* 30, 284-325.
- [169] Okada, S., Murakami, M., and Yamaguchi, K. (1995) Hydrocarbon composition of newly isolated strains of the green microalga *Botryococcus braunii*, *J. Appl. Phycol.* 7, 555-559.
- [170] Camp, C. H., Jr., Lee, Y. J., Heddlestone, J. M., Hartshorn, C. M., Hight Walker, A. R., Rich, J. N., Lathia, J. D., and Cicerone, M. T. (2014) High-speed coherent raman fingerprint imaging of biological tissuesw, *Nature photonics* 8, 627-634.
- [171] Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Mennucci, B., Petersson, G. A., Nakatsuji, H., Caricato, M., Li, X., Hratchian, H. P., Izmaylov, A. F., Bloino, J., Zheng, G., Sonnenberg, J. L., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Montgomery, J., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M., Heyd, J. J., Brothers, E., Kudin, K. N., Staroverov, V. N., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Rega, N., Millam, J. M., Klene, M., Knox, J. E., Cross, J. B., Bakken, V., Adamo, C., Jaramillo, J., Gomperts, R., Stratmann, R. E., Yazyev, O., Austin, A. J., Cammi, R., Pomelli, C., Ochterski, J. W., Martin, R. L., Morokuma, K., Zakrzewski, V. G., Voth, G. A., Salvador, P., Dannenberg, J. J., Dapprich, S., Daniels, A. D., Farkas, Ö., Foresman, J. B., Ortiz, J. V., Cioslowski, J., and Fox, D. J. (2009) Gaussian 09, Revision A.02, *Gaussian, Inc., Wallingford CT*.

- [172] Becke, A. D. (1993) Density-Functional Thermochemistry .3. The Role of Exact Exchange, *J Chem Phys* 98, 5648-5652.
- [173] Lee, C. T., Yang, W. T., and Parr, R. G. (1988) Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density, *Phys Rev B* 37, 785-789.
- [174] AGUI program from Semichem, I. S., KS 66216. (2016) www.semichem.com.
- [175] Chun, H. J., Meinander, N., Villarreal, J. R., and Laane, J. (2015) Vibrational spectra, theoretical calculations, and two-dimensional potential energy surface for the ring-puckering vibrations of 2,4,7-trioxa[3.3.0]octane, *J Phys Chem A* 119, 410-417.
- [176] Brownlee, C. (2002) Role of the extracellular matrix in cell-cell signalling: paracrine paradigms, *Current opinion in plant biology* 5, 396-401.
- [177] Michel, G., Tonon, T., Scornet, D., Cock, J. M., and Kloareg, B. (2010) The cell wall polysaccharide metabolism of the brown alga *Ectocarpus siliculosus*. Insights into the evolution of extracellular matrix polysaccharides in Eukaryotes, *The New phytologist* 188, 82-97.
- [178] Zhong, B., Liu, L., Yan, Z., and Penny, D. (2013) Origin of land plants using the multispecies coalescent model, *Trends in plant science* 18, 492-495.
- [179] Zhong, B., Xi, Z., Goremykin, V. V., Fong, R., McLenachan, P. A., Novis, P. M., Davis, C. C., and Penny, D. (2014) Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes, *Molecular biology and evolution* 31, 177-183.

- [180] Coleman, A. W. (2012) A comparative analysis of the Volvocaceae (Chlorophyta), *Journal of phycology* 48, 491-513.
- [181] Metzger, P., Berkaloff, C., Casadevall, E., and Coute, A. (1985) Alkadiene-producing and botryococcene-producing races of wild strains of *Botryococcus braunii*, *Phytochem.* 24, 2305-2312.
- [182] Metzger, P., Templier, J., Largeau, C., and Casadevall, E. (1986) An *n*-alkatriene and some *n*-alkadienes from the A race of the green alga *Botryococcus braunii*, *Phytochem.* 25, 1869-1872.
- [183] Templier, J., Largeau, C., and Casadevall, E. (1991) Biosynthesis of *n*-Alkatrienes in *Botryococcus braunii*, *Phytochemistry* 30, 2209-2215.
- [184] Metzger, P., Casadevall, E., and Couté, A. (1988) Botryococcene distribution in strains of the green alga *Botryococcus braunii*, *Phytochem.* 27, 1383-1388.
- [185] 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, *Phytochem.* 26, 129-134.
- [186] Metzger, P., Rager, M. N., and Fosse, C. (2008) Braunicetals: Acetals from condensation of macrocyclic aldehydes and terpene diols in *Botryococcus braunii*, *Phytochem.* 69, 2380-2386.
- [187] Berkaloff, C., Rousseau, B., Coute, A., Casadevall, E., Metzger, P., and Chirac, C. (1984) Variability of cell-wall structure and hydrocarbon type in different strains of *Botryococcus braunii*, *J. Phycol.* 20, 377-389.

- [188] Wolf, F. R., and Cox, E. R. (1981) Ultrastructure of active and resting colonies of *Botryococcus braunii* (Chlorophyceae), *J. Phycol.* 17, 395-405.
- [189] Uno, Y., Nishii, I., Kagiwada, S., and Noguchi, T. (2015) Colony sheath formation is accompanied by shell formation and release in the green alga *Botryococcus braunii* (race B), *Algal Res* 8, 214-223.
- [190] Molnar, I., Lopez, D., Wisecaver, J. H., Devarenne, T. P., Weiss, T. L., Pellegrini, M., and Hackett, J. D. (2012) Bio-crude transcriptomics: Gene discovery and metabolic network reconstruction for the biosynthesis of the terpenome of the hydrocarbon oil-producing green alga, *Botryococcus braunii* race B (Showa), *BMC genomics* 13, 576.
- [191] Ghahremani, M., Stigter, K. A., and Plaxton, W. (2016) Extraction and characterization of extracellular proteins and their post-translational modifications from *Arabidopsis thaliana* suspension cell cultures and seedlings: A critical review, *Proteomes* 4, 25.
- [192] Moreno, N., Bougourd, S., Haseloff, J., and Feijó, J. A. (2006) Imaging Plant Cells, In *Handbook of Biological Confocal Microscopy* (Pawley, J. B., Ed.) 3 ed., pp 769-787, Springer Science and Business Media, New York.
- [193] Devarenne, T. P., Ekengren, S. K., Pedley, K. F., and Martin, G. B. (2006) Adi3 is a Pdk1-interacting AGC kinase that negatively regulates plant cell death, *EMBO J* 25, 255-265.

- [194] Mellquist, J. L., Kasturi, L., Spitalnik, S. L., and Shakin-Eshleman, S. H. (1998) The amino acid following an Asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency, *Biochemistry* 37, 6833-6837.
- [195] Aoki, K., Perlman, M., Lim, J. M., Cantu, R., Wells, L., and Tiemeyer, M. (2007) Dynamic developmental elaboration of N-linked glycan complexity in the *Drosophila melanogaster* embryo, *The Journal of biological chemistry* 282, 9127-9142.
- [196] Nguema-Ona, E., Vicre-Gibouin, M., Gotte, M., Plancot, B., Lerouge, P., Bardor, M., and Driouich, A. (2014) Cell wall O-glycoproteins and N-glycoproteins: aspects of biosynthesis and function, *Frontiers in plant science* 5, 499.
- [197] Briesemeister, S., Rahnenfuhrer, J., and Kohlbacher, O. (2010) Going from where to why--interpretable prediction of protein subcellular localization, *Bioinformatics* 26, 1232-1238.
- [198] Briesemeister, S., Rahnenfuhrer, J., and Kohlbacher, O. (2010) YLoc--an interpretable web server for predicting subcellular localization, *Nucleic acids research* 38, W497-502.
- [199] Naidu, R. A., Ingle, C. J., Deom, C. M., and Sherwood, J. L. (2004) The two envelope membrane glycoproteins of *Tomato spotted wilt virus* show differences in lectin-binding properties and sensitivities to glycosidases, *Virology* 319, 107-117.

- [200] Kawashima, H., Sueyoshi, S., Li, H., Yamamoto, K., and Osawa, T. (1990) Carbohydrate binding specificities of several poly-*N*-acetyllactosamine-binding lectins, *Glycoconjugate journal* 7, 323-334.
- [201] Allard, B., and Casadevall, E. (1990) Carbohydrate-composition and characterization of sugars from the green microalga *Botryococcus braunii*, *Phytochemistry* 29, 1875-1878.
- [202] Fernandes, H. L., Tome, M. M., Lupi, F. M., Fialho, A. M., Sá-Correia, I., and Novais, J. M. (1989) Biosynthesis of high concentrations of an exopolysaccharide during the cultivation of the microalga *Botryococcus braunii*, *Biotech. Lett.* 11, 433-436.
- [203] Lupi, F. M., Fernandes, H. L., Sá-Correia, I., and Novais, J. M. (1991) Temperature profiles of cellular growth and exopolysaccharide synthesis by *Botryococcus braunii* Kütz. UC 58, *J. Appl. Phycol.* 3, 35-42.
- [204] Lupi, F. M., Fernandes, H. L., Tome, M. M., Sá-Correia, I., and Novais, J. M. (1994) Influence of nitrogen source and photoperiod on exopolysaccharide synthesis by the microalga *Botryococcus braunii* UC 58, *Enzyme Microb Tech* 16, 546-550.
- [205] Johnson, K. L., Cassin, A. M., Lonsdale, A., Bacic, A., Doblin, M. S., and Schultz, C. J. (2017) A motif and amino acid bias bioinformatics pipeline to identify hydroxyproline-rich glycoproteins, *Plant Physiol* 174, 886-903.
- [206] Johnson, K. L., Cassin, A. M., Lonsdale, A., Wong, G. K., Soltis, D. E., Miles, N. W., Melkonian, M., Melkonian, B., Deyholos, M. K., Leebens-Mack, J., Rothfels, C. J., Stevenson, D. W., Graham, S. W., Wang, X., Wu, S., Pires, J. C., Edger, P.

- P., Carpenter, E. J., Bacic, A., Doblin, M. S., and Schultz, C. J. (2017) Insights into the evolution of hydroxyproline-rich glycoproteins from 1000 plant transcriptomes, *Plant Physiol* 174, 904-921.
- [207] Kavi Kishor, P. B., Hima Kumari, P., Sunita, M. S., and Sreenivasulu, N. (2015) Role of proline in cell wall synthesis and plant development and its implications in plant ontogeny, *Frontiers in plant science* 6, 544.
- [208] Hijazi, M., Roujol, D., Nguyen-Kim, H., Del Rocio Cisneros Castillo, L., Saland, E., Jamet, E., and Albenne, C. (2014) Arabinogalactan protein 31 (AGP31), a putative network-forming protein in *Arabidopsis thaliana* cell walls?, *Annals of botany* 114, 1087-1097.
- [209] Bollig, K., Lamshoft, M., Schweimer, K., Marner, F. J., Budzikiewicz, H., and Waffenschmidt, S. (2007) Structural analysis of linear hydroxyproline-bound *O*-glycans of *Chlamydomonas reinhardtii*--conservation of the inner core in *Chlamydomonas* and land plants, *Carbohydrate research* 342, 2557-2566.
- [210] Schiller, B., Hykollari, A., Yan, S., Paschinger, K., and Wilson, I. B. (2012) Complicated *N*-linked glycans in simple organisms, *Biological chemistry* 393, 661-673.
- [211] Takahashi, M., Kuroki, Y., Ohtsubo, K., and Taniguchi, N. (2009) Core fucose and bisecting GlcNAc, the direct modifiers of the *N*-glycan core: their functions and target proteins, *Carbohydrate research* 344, 1387-1390.
- [212] Strasser, R. (2016) Plant protein glycosylation, *Glycobiology* 26, 926-939.

- [213] Takeuchi, T., Hayama, K., Hirabayashi, J., and Kasai, K. (2008) *Caenorhabditis elegans* *N*-glycans containing a Gal-Fuc disaccharide unit linked to the innermost GlcNAc residue are recognized by *C. elegans* galectin LEC-6, *Glycobiology* 18, 882-890.
- [214] Paschinger, K., Razzazi-Fazeli, E., Furukawa, K., and Wilson, I. B. (2011) Presence of galactosylated core fucose on *N*-glycans in the planaria *Dugesia japonica*, *Journal of mass spectrometry : JMS* 46, 561-567.
- [215] Takahashi, N., Masuda, K., Hiraki, K., Yoshihara, K., Huang, H. H., Khoo, K. H., and Kato, K. (2003) *N*-Glycan structures of squid rhodopsin, *European journal of biochemistry* 270, 2627-2632.
- [216] Utille, J. P., and Priem, B. (2000) Synthesis of allyl 2-*O*-(α -L-arabinofuranosyl)-6-*O*-(α -D-mannopyranosyl)- β -D-mannopyranoside, a unique plant *N*-glycan motif containing arabinose, *Carbohydrate research* 329, 431-439.
- [217] Sturm, A. (1992) A wound-inducible glycine-rich protein from *Daucus carota* with homology to single-stranded nucleic acid-binding proteins, *Plant Physiol* 99, 1689-1692.
- [218] Priem, B., Gitti, R., Bush, C. A., and Gross, K. C. (1993) Structure of ten free *N*-glycans in ripening tomato fruit. Arabinose is a constituent of a plant *N*-glycan, *Plant Physiol* 102, 445-458.
- [219] Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions, *Genome Biology* 14, R36.

- [220] Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010) Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms, *Nature biotechnology* 28, 511-515.
- [221] Pagni, M., Ioannidis, V., Cerutti, L., Zahn-Zabal, M., Jongeneel, C. V., Hau, J., Martin, O., Kuznetsov, D., and Falquet, L. (2007) MyHits: improvements to an interactive resource for analyzing protein sequences, *Nucleic acids research* 35, W433-437.
- [222] Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T., and Brunak, S. (2005) Prediction of twin-arginine signal peptides, *BMC bioinformatics* 6, 167.
- [223] Kall, L., Krogh, A., and Sonnhammer, E. L. (2004) A combined transmembrane topology and signal peptide prediction method, *Journal of molecular biology* 338, 1027-1036.
- [224] Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nature methods* 8, 785-786.
- [225] Chill, L., Trinh, L., Azadi, P., Ishihara, M., Sonon, R., Karnaukhova, E., Ophir, Y., Golding, B., and Shiloach, J. (2009) Production, purification, and characterization of human $\alpha 1$ proteinase inhibitor from *Aspergillus niger*, *Biotechnology and bioengineering* 102, 828-844.

- [226] Angel, P. M., Lim, J. M., Wells, L., Bergmann, C., and Orlando, R. (2007) A potential pitfall in ^{18}O -based *N*-linked glycosylation site mapping, *Rapid communications in mass spectrometry : RCM* 21, 674-682.
- [227] Ma, L., Chen, Z., Huang da, W., Kutty, G., Ishihara, M., Wang, H., Abouelleil, A., Bishop, L., Davey, E., Deng, R., Deng, X., Fan, L., Fantoni, G., Fitzgerald, M., Gogineni, E., Goldberg, J. M., Handley, G., Hu, X., Huber, C., Jiao, X., Jones, K., Levin, J. Z., Liu, Y., Macdonald, P., Melnikov, A., Raley, C., Sassi, M., Sherman, B. T., Song, X., Sykes, S., Tran, B., Walsh, L., Xia, Y., Yang, J., Young, S., Zeng, Q., Zheng, X., Stephens, R., Nusbaum, C., Birren, B. W., Azadi, P., Lempicki, R. A., Cuomo, C. A., and Kovacs, J. A. (2016) Genome analysis of three *Pneumocystis* species reveals adaptation mechanisms to life exclusively in mammalian hosts, *Nat Commun* 7, 10740.
- [228] Anumula, K. R., and Taylor, P. B. (1992) A comprehensive procedure for preparation of partially methylated alditol acetates from glycoprotein carbohydrates, *Analytical biochemistry* 203, 101-108.
- [229] Mercado, C. P., Quintero, M. V., Li, Y., Singh, P., Byrd, A. K., Talabnin, K., Ishihara, M., Azadi, P., Rusch, N. J., Kuberan, B., Maroteaux, L., and Kilic, F. (2013) A serotonin-induced *N*-glycan switch regulates platelet aggregation, *Scientific reports* 3, 2795.
- [230] Woo, C. M., Felix, A., Byrd, W. E., Zuegel, D. K., Ishihara, M., Azadi, P., Iavarone, A. T., Pitteri, S. J., and Bertozzi, C. R. (2017) Development of IsoTaG, a chemical

glycoproteomics technique for profiling intact *N*- and *O*-glycopeptides from whole cell proteomes, *Journal of proteome research* 16, 1706-1718.

APPENDIX A

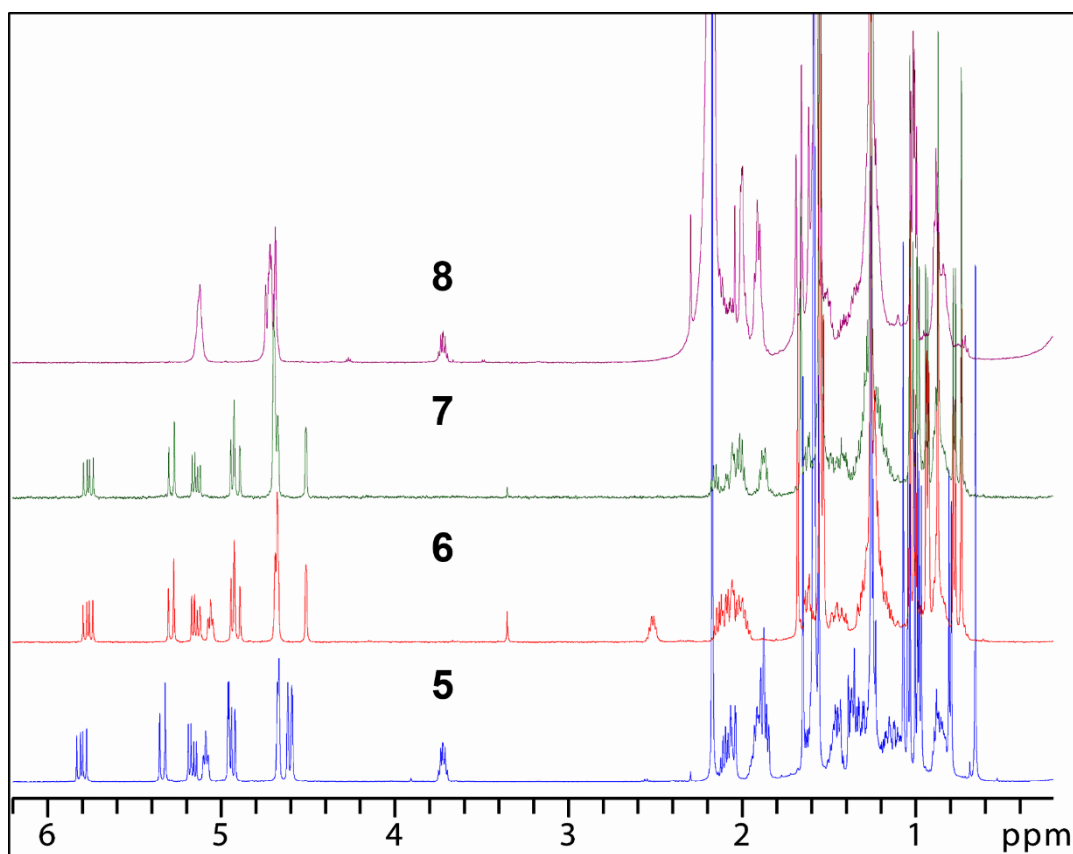


Figure A-1 Overlay of 1D ¹H NMR spectra for C₃₃-2 trimethylsqualene isomer 8 (top, magenta), cyclic C₃₃-3 botryococcene 7 (green), cyclic C₃₃-2 botryococcene 6 (red), and cyclic C₃₃-1 botryococcene 5 (bottom, blue) is shown here.

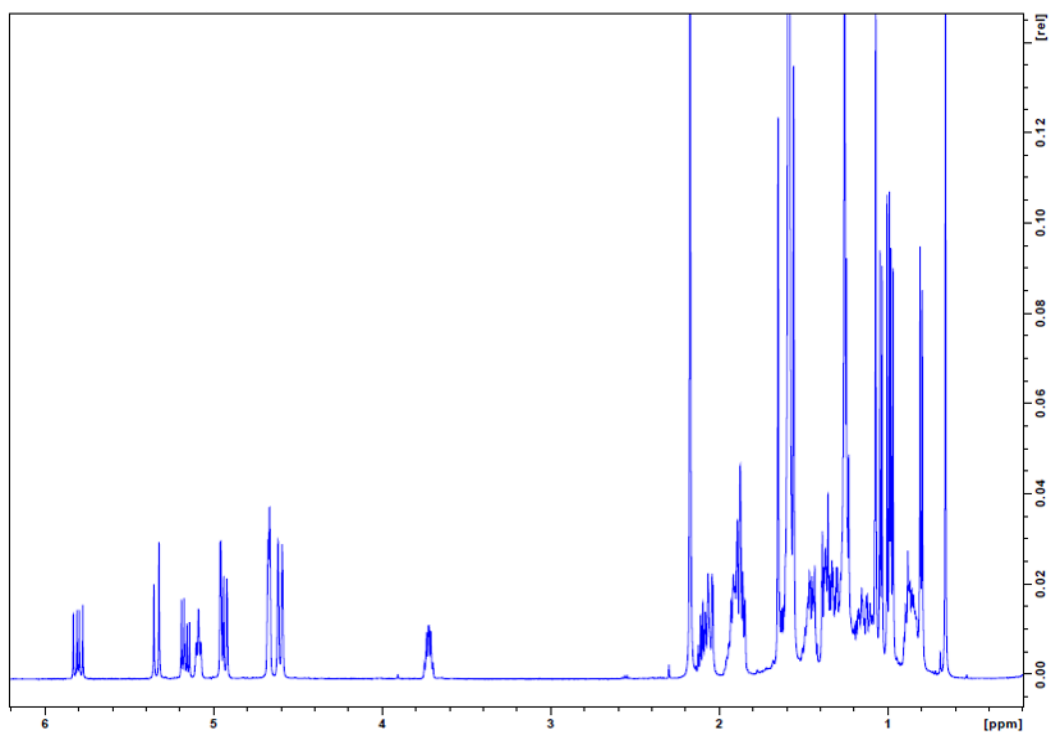


Figure A-2 1D ^1H NMR spectra for cyclic $\text{C}_{33}\text{-1}$ botryococcene 5.

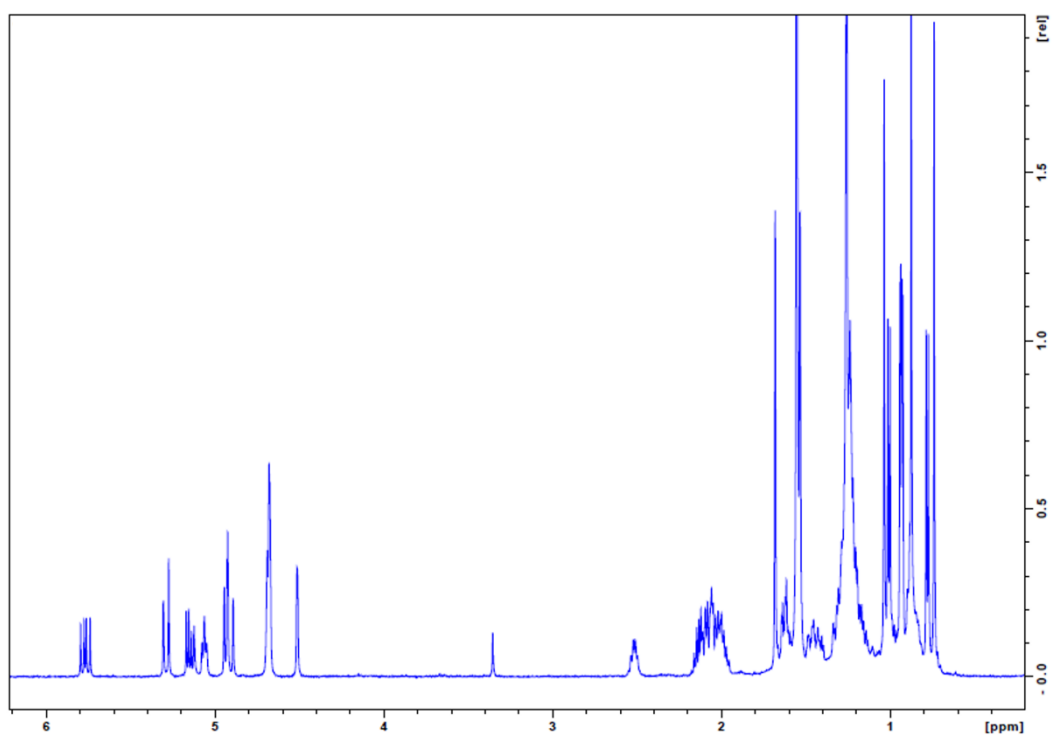


Figure A-3 1D ^1H NMR spectra for cyclic $\text{C}_{33}\text{-2}$ botryococcene 6.

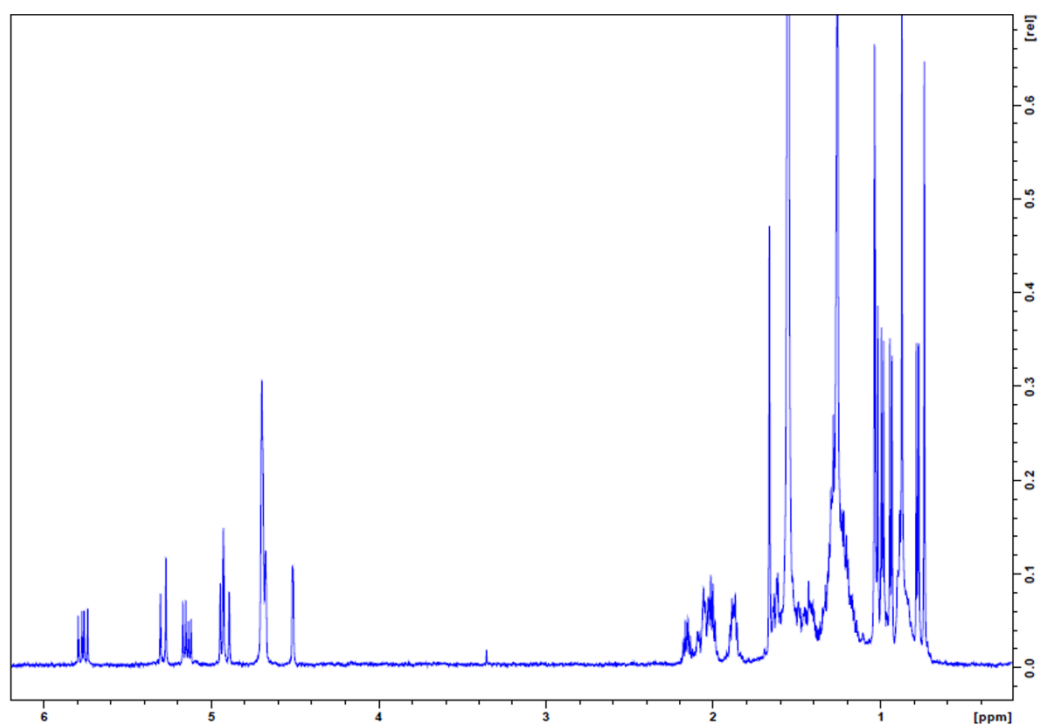


Figure A-4 1D ^1H NMR spectra for cyclic $\text{C}_{33}\text{-3}$ botryococcene 7.

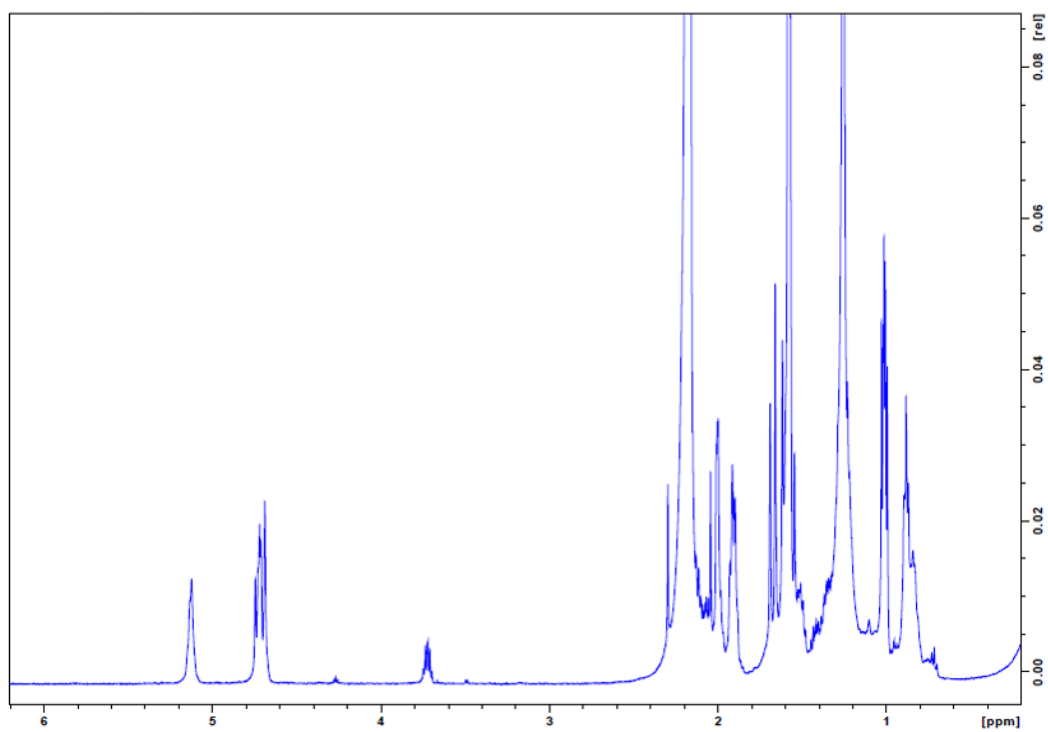


Figure A-5 1D ^1H NMR spectra for C₃₃-2 trimethylsqualene isomer 8.

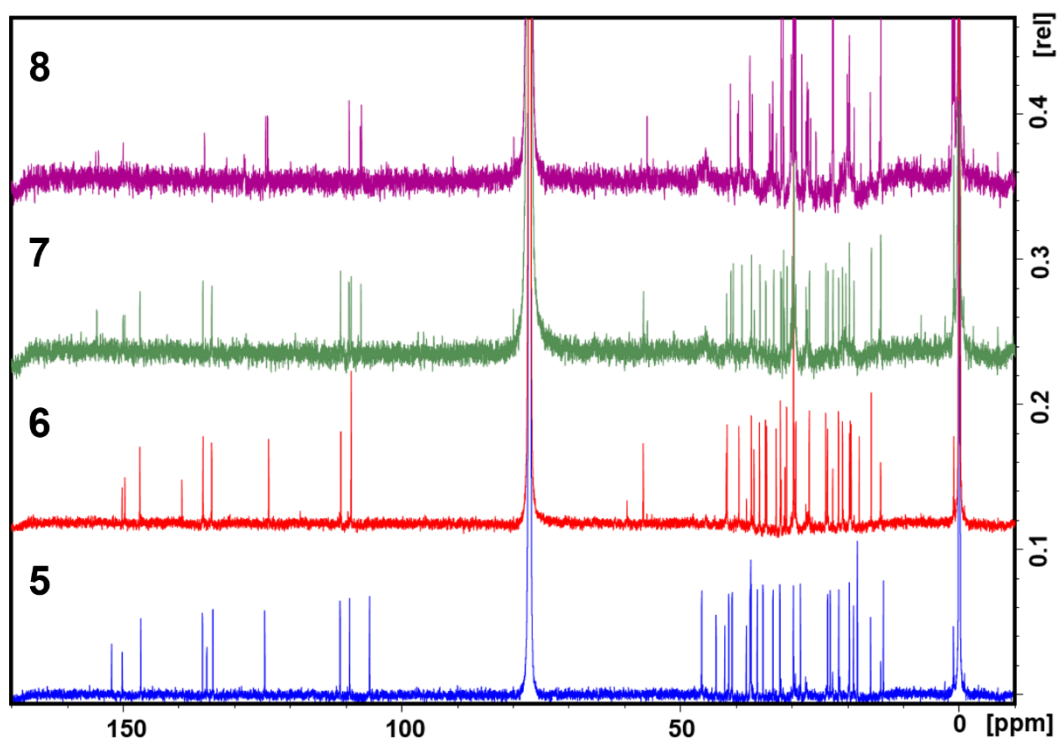


Figure A-6 Overlay of 1D ¹³C NMR spectra for C₃₃-2 trimethylsqualene isomer 8 (top, magenta), cyclic C₃₃-3 botryococcene 7 (green), cyclic C₃₃-2 botryococcene 6 (red), and cyclic C₃₃-1 botryococcene 5 (bottom, blue) is shown here.

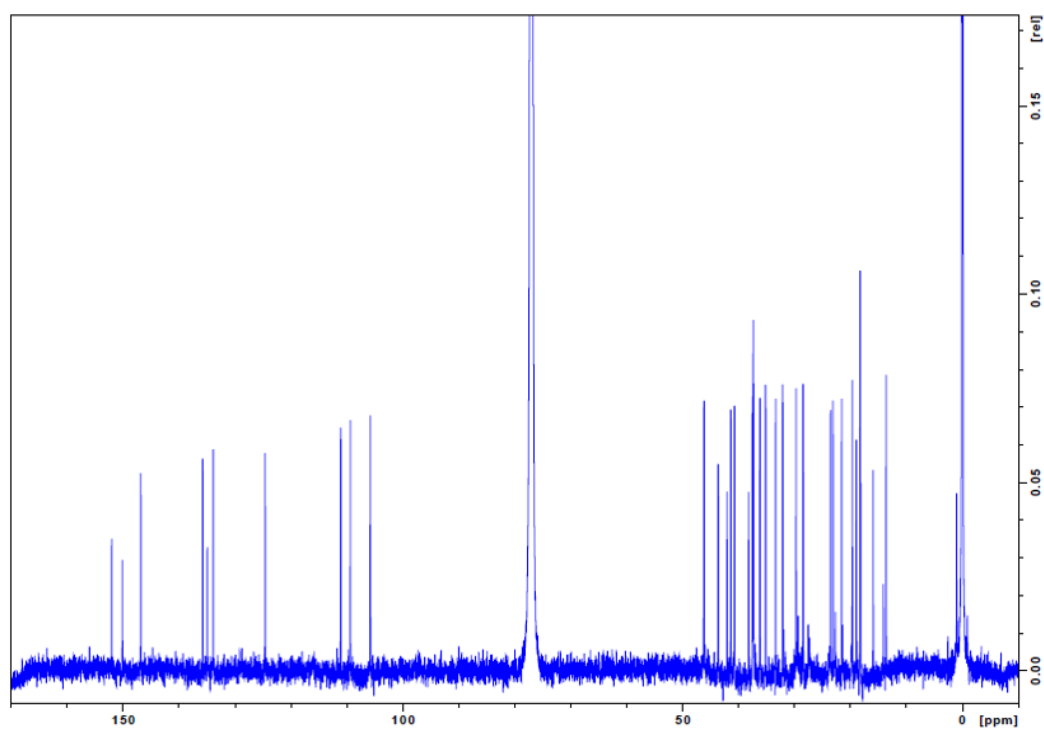


Figure A-7 1D ^{13}C NMR spectra for cyclic C₃₃-1 botryococcene 5.

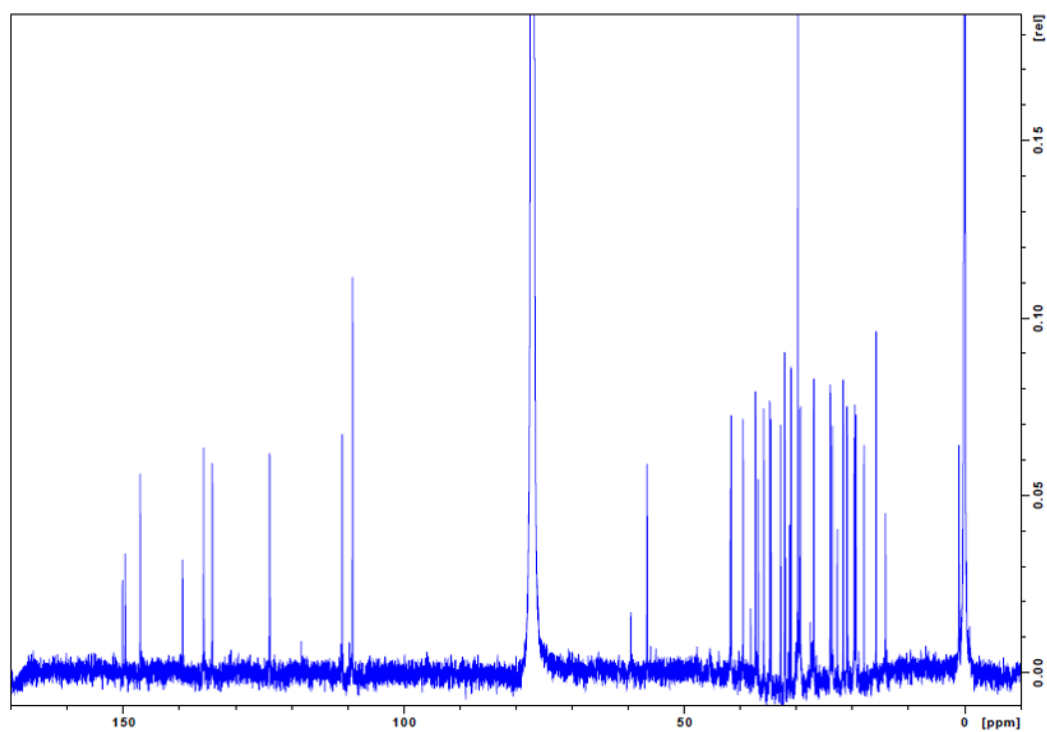


Figure A-8 1D ^{13}C NMR spectra for cyclic C_{33-2} botryococcene 6.

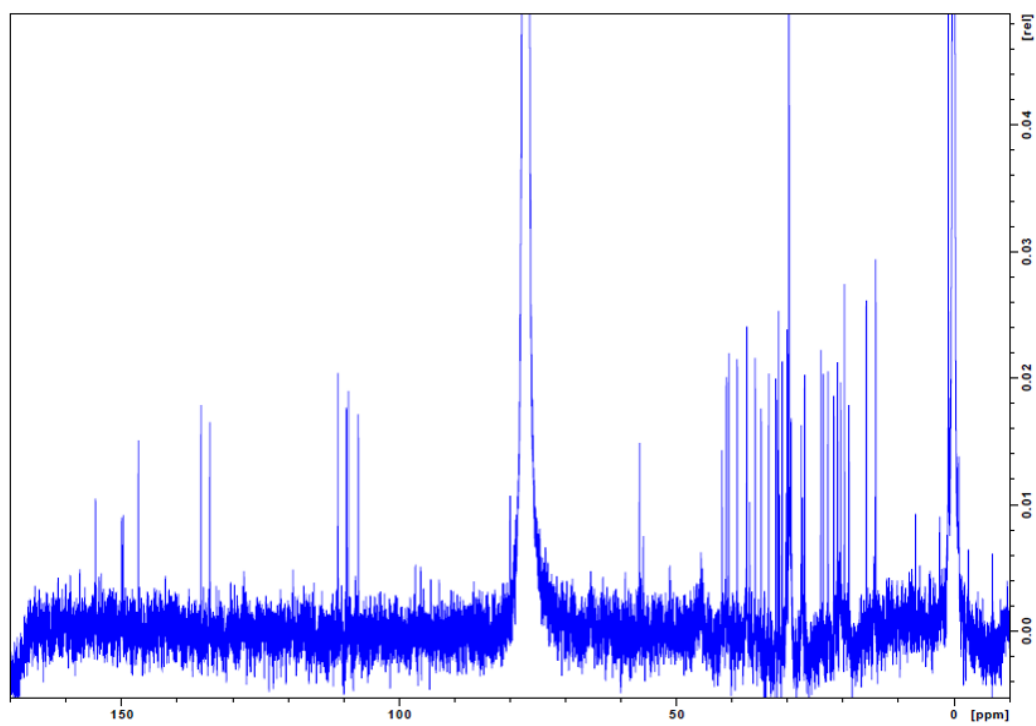


Figure A-9 1D ^{13}C NMR spectra for cyclic $\text{C}_{33}\text{-3}$ botryococcene **7**.

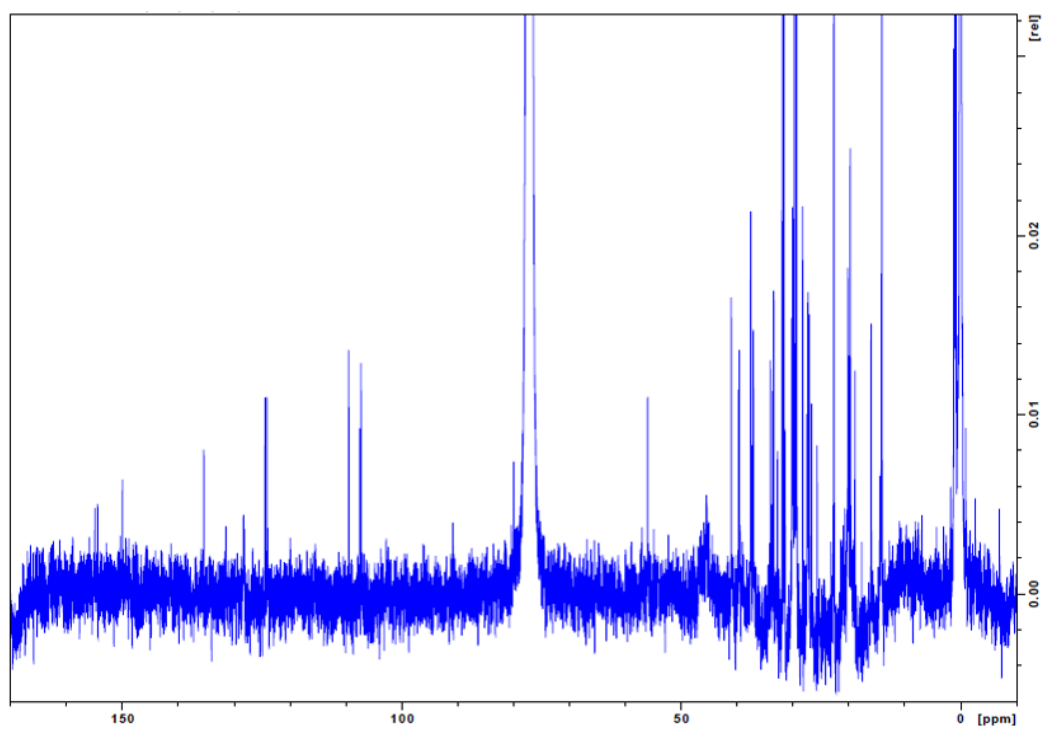


Figure A-10 1D ^{13}C NMR spectra for C₃₃-2 trimethylsqualene isomer 8.