



University of Kentucky  
UKnowledge

---

Theses and Dissertations--Plant and Soil  
Sciences

Plant and Soil Sciences

---

2014

## UNDERSTANDING THE CHEMICAL GYMNASTICS OF ENZYME- CATALYZED 1'-1 AND 1'-3 TRITERPENE LINKAGES

Stephen A. Bell

University of Kentucky, sbell8653@gmail.com

[Right click to open a feedback form in a new tab to let us know how this document benefits you.](#)

---

### Recommended Citation

Bell, Stephen A., "UNDERSTANDING THE CHEMICAL GYMNASTICS OF ENZYME-CATALYZED 1'-1 AND 1'-3 TRITERPENE LINKAGES" (2014). *Theses and Dissertations--Plant and Soil Sciences*. 43.  
[https://uknowledge.uky.edu/pss\\_etds/43](https://uknowledge.uky.edu/pss_etds/43)

This Doctoral Dissertation is brought to you for free and open access by the Plant and Soil Sciences at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Plant and Soil Sciences by an authorized administrator of UKnowledge. For more information, please contact [UKnowledge@lsv.uky.edu](mailto:UKnowledge@lsv.uky.edu).

## **STUDENT AGREEMENT:**

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

## **REVIEW, APPROVAL AND ACCEPTANCE**

The document mentioned above has been reviewed and accepted by the student's advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student's thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Stephen A. Bell, Student

Dr. Joseph Chappell, Major Professor

Dr. Arthur Hunt, Director of Graduate Studies

UNDERSTANDING THE CHEMICAL GYMNASTICS OF  
ENZYME-CATALYZED 1'-1 AND 1'-3 TRITERPENE LINKAGES

---

DISSERTATION

---

A dissertation submitted in partial fulfillment of the  
requirements for the degree of Doctor of Philosophy in the  
College of Agriculture, Food, and Environment  
at the University of Kentucky

By  
Stephen A. Bell

Lexington, Kentucky

Director: Dr. Joseph Chappell, Professor of Plant Physiology

Lexington, Kentucky

2014

Copyright © Stephen A. Bell 2014

## ABSTRACT OF DISSERTATION

### UNDERSTANDING THE CHEMICAL GYMNASTICS OF ENZYME-CATALYZED 1'-1 AND 1'-3 TRITERPENE LINKAGES

Squalene synthase (SS) is an essential enzyme in eukaryotic systems responsible for an important branch point in isoprenoid metabolism that leads to sterol formation. The mechanistic complexity of SS has made it a difficult enzyme to study. The green alga *Botryococcus braunii* race B possesses several squalene synthase-like (SSL) enzymes that afford a unique opportunity to study the complex mechanism of triterpene biosynthesis. SSL-1 catalyzes presqualene diphosphate (PSP) formation, which can either be converted to squalene by SSL-2 or botryococcene by SSL-3. A rationally designed mutant study of *B. braunii* squalene synthase (BbSS) and SSL-3 was conducted to understand structure-function relations among these enzymes. These studies revealed two amino acid positions in SSL-3 (N171, G207) that appeared to control 1'-3 versus 1'-1 linkages. The reciprocal mutations in the corresponding positions of BbSS did not convert this enzyme into a botryococcene synthase.

Next, a genetic selection was developed to evolve SSL enzymes towards a fully functional SS. Previous studies have shown that *Saccharomyces cerevisiae* squalene synthase (ScSS) can be knocked out and although lethal, growth can be restored by providing an exogenous source of ergosterol. Additional studies have shown that successful complementation of the ScSS knockout with a non-fungal SS is possible but requires a fungal SS carboxy-terminus region. Given these observations, proof-of-principle experiments were conducted to demonstrate that SSL-SSL fusion enzymes could complement the ScSS knockout followed by construction of a mutant SSL-SSL fusion enzyme library that was screened in the ScSS knockout yeast line. From this library, mutant SSL-SSL fusion enzymes were identified that were able to complement, which demonstrated the feasibility of this approach as a genetic selection for mutant SSL enzymes.

Squalene and botryococcene have valuable industrial applications in vaccine adjuvant formations, cosmetic products, and renewable energy feedstock material. Limitations in natural sources of these molecules have made heterologous production of them an important research target. Algae represent a

desirable group of organisms that could be engineered to produce these metabolites because they are photosynthetic and capable of using non-arable farmland. The feasibility, approach, and progress for engineering green algae to produce squalene and botryococcene are discussed.

KEYWORDS: triterpene synthase, structure-function map, directed evolution, metabolic engineering, green algae

Stephen A. Bell  
\_\_\_\_\_  
Student's Signature

4-20-2014  
\_\_\_\_\_  
Date

UNDERSTANDING THE CHEMICAL GYMNASTICS OF  
ENZYME-CATALYZED 1'-1 AND 1'-3 TRITERPENE LINKAGES

By

Stephen A. Bell

Dr. Joseph Chappell

---

Director of Dissertation

Dr. Arthur Hunt

---

Director of Graduate Studies

4-20-2014

---

## ACKNOWLEDGEMENTS

The scientific work presented in this dissertation reflects not just the growth and maturity of my scientific mind, but also the guidance and inspiration of many key people throughout the past ten years. First and foremost, I have to thank my dissertation mentor, Dr. Joseph Chappell. His continual support, guidance, and confidence in me as a person and scientist were crucial throughout this work. Without him, I would not be the person I am today. Next, I must thank my committee members Dr. Arthur Hunt, Dr. David Watt, and Dr. Seth DeBolt as well as my outside examiner Dr. Eric Oldfield. Their challenging questions and example setting ways encouraged me to strive for excellence. A special thanks to Dr. Arthur Hunt who graciously allowed me to work in his laboratory near the end of my undergraduate studies, which ultimately led to my dissertation work in the Chappell laboratory.

To all the previous members of the Chappell laboratory, thank you for allowing me to stand on your shoulders. Without each of you, the scientific frontier that I was so fortunate to have the pleasure of venturing through would not have been possible. Special thanks also to all the members of the Chappell laboratory and other nearby laboratories that I was lucky enough to work alongside. Collectively, your hard work and smiling faces aided me in all of my accomplishments and for that, I am forever grateful.

Finally, I have to acknowledge all the people who, although not directly involved in my dissertation work, provided critical advice and support throughout the years. Andrew Berry, who gave me my first opportunity to perform a structured research project, has been a great voice of reason during my college years and more importantly, an incredible friend. To all the other great friends (you know who you are), thank you for continual encouragement throughout the course of my dissertation work. And last but in no way least, a huge thank you to my immediate family for always believing in me, giving me the confidence that anything can be achieved with hard work and persistence, and most importantly your everlasting, unconditional support of my free will.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS .....	iii
LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
Chapter One: Background and Introduction	
Part One: Triterpene Synthases .....	1
Part Two: Triterpene Metabolic Engineering .....	9
Aims of Dissertation .....	13
Chapter Two: Understanding the Mechanistic Specificity Between 1'-1 and 1'-3 Triterpene Linkages Catalyzed by Squalene Synthase <i>versus</i> Botryococcene Synthase	
Background and Introduction .....	14
Results .....	18
Identification of candidate residues for mutation .....	18
Rationale for <i>in vitro</i> activity screening of wild-type and mutant triterpene synthases .....	18
Overexpression of key MEP pathway biosynthetic genes in <i>E. coli</i> .....	20
Triterpene synthase permutations .....	22
M1 series mutants .....	22
M3 series mutants .....	24
M4 series mutants .....	26
M5 series mutants .....	28
M7 series mutants .....	30
M8 series mutants .....	32
M11 series mutants .....	34
Additional verification for mutants of interest from previous mutant series .....	38
GC-MS verification of <i>in vitro</i> and <i>in vivo</i> enzyme products .....	44
Crude catalytic efficiency assessment of mutants relative to wild type .....	46
Discussion .....	49
Materials and Methods .....	53
3D modeling .....	53
PSPD docking .....	53
Wild-type <i>B. braunii</i> triterpene synthase cloning .....	53
PCR construction of mutant genes .....	56



Construction of a methyl erythritol phosphate (MEP) pathway overexpression cassette .....	56
Protein expression .....	57
Hot <sup>3</sup> H-FPP <i>in vitro</i> enzyme assays .....	57
Cold FPP <i>in vitro</i> enzyme assays .....	57
<i>In vivo</i> chemical profiling .....	58
Radiolabeled crude relative efficiency assays .....	58

### Chapter Three: Development of a Selection Platform for the Directed Evolution of Triterpene Synthases

Background and Introduction .....	60
Results .....	66
Establishing the yeast selection platform .....	66
Validation of inducible gene expression system and complementation testing in ZX178-08 .....	66
Complementation testing of SSL-1 and SSL-2 single-gene and co-expressed constructs in ZX178-08 .....	69
Chemical profiling and growth assessment of SSL-1 and SSL-2 single-gene and co-expressed constructs in ZXD .....	73
Complementation testing of SSL1-SSL2 fusion constructs in ZX178-08 .....	73
Chemical profiling and growth assessment of SSL1-SSL2 fusion constructs in ZXD .....	75
Complementation testing of SSL-1, SSL-3, and mtSSL-3 co-expressed constructs in ZX178-08 .....	75
Chemical profiling and growth assessment of SSL-1, SSL-3, and mtSSL-3 co-expressed constructs in ZXD .....	78
Complementation testing of SSL1-SSL3 and SSL1-mtSSL3 fusion constructs in ZX178-08 .....	78
Chemical profiling and growth assessment of SSL1-SSL3 and SSL1-mtSSL3 fusion constructs in ZXD .....	81
Complementation and growth analysis of SSL-1, SSL-3, and mtSSL-3 single-gene and fusion constructs with normalized cell concentrations .....	81
Sterol profiling of the squalene synthase knockout lines complemented with mutant squalene synthase-like genes .....	83
Library construction and validation .....	89
Library screening in ZX178-08 for <i>erg9Δ</i> complementation .....	89
Discussion .....	94
Materials and Methods .....	98
Plasmids and yeast strains .....	98

Maintenance of ZX178-08 and ZXD yeast lines, transformation, and media .....	102
Complementation and growth testing .....	102
Chemical profiling of hydrocarbon products .....	102
Sterol profiling for complementing constructs .....	103
Growth analysis for complementing constructs .....	103
Mutant library construction .....	103
Mutant library selection in ZX178-08 .....	104
Chapter Four: Metabolic Engineering for Triterpene Accumulation in the Green Alga <i>Chlamydomonas reinhardtii</i>	
Background and Introduction .....	106
Results .....	112
First-generation nuclear transformation vectors .....	112
Efficiency of nuclear transformation with pHRC1 vectors .....	112
Verification of transgene insertion and farnesol accumulation measurements .....	116
Second-generation nuclear transformation vectors .....	116
Third-generation nuclear transformation vectors .....	117
Discussion .....	120
Materials and Methods .....	122
Construction of first-generation vectors .....	122
Construction of second-generation vectors .....	122
Construction of third-generation vectors .....	125
<i>C. reinhardtii</i> and <i>D. tertiolecta</i> maintenance and media .....	126
<i>C.reinhardtii</i> electroporation protocol for nuclear transformation .....	126
PCR screening for acFPS transgene insertion .....	127
Chemical profiling of hydrocarbon products in transgenic lines .....	127
Chapter Five: Concluding Remarks .....	128
References .....	130
Vita .....	136

## LIST OF TABLES

Table 2.1 Primers used for cloning and screening enzymology constructs .....	54
Table 3.1 Primers used for cloning and screening yeast constructs .....	99
Table 4.1 Primers used for cloning and screening algae constructs .....	123

## LIST OF FIGURES

Figure 1.1 Terpenoid diversity that arises from the five-carbon isoprenoid building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), generated by various modifications such as successive condensation of the allylic diphosphates themselves, dephosphorylation, cyclization, desaturation, reduction, oxidation, and hydroxylation .....	2
Figure 1.2 Biosynthesis of triterpenes from farnesyl diphosphate (FPP) .....	3
Figure 1.3 ClustalW multiple sequence alignment (gap penalty: 22 and gap length penalty: 0.75) of triterpene synthases .....	6
Figure 1.4 Similarities between CrtM and HsSS crystal structures solved with the unreactive substrate analogue farnesyl thiodiphosphate (FSPP) and the stable intermediate PSPP .....	7
Figure 1.5 World total energy consumption by fuel type according to EIA as of 2011 .....	10
Figure 1.6 Generic depiction of plant and green algae cells showing locales of the MVA and MEP pathways and products that typically arise from them .....	12
Figure 2.1 Proposed reaction cascade for enzyme-catalyzed squalene and botryococcene formation .....	15
Figure 2.2 Strategy for identification of residues controlling first- and second-step reaction activity and product specificity in triterpene synthases .....	17
Figure 2.3 Coupled assay approach to screen mutant triterpene synthases for the first, second, and both steps of the squalene/botryococcene synthase reactions .....	19
Figure 2.4 <i>In vivo</i> determinations for which of four enzymes affects carbon flux into squalene in <i>E. coli</i> BL21 (DE3) .....	21
Figure 2.5 M1 series mutants .....	23
Figure 2.6 M3 series mutants .....	25
Figure 2.7 M4 series mutants .....	27
Figure 2.8 M5 series mutants .....	29
Figure 2.9 M7 series mutants .....	31
Figure 2.10 M8 series mutants .....	33
Figure 2.11 BbSS M11 series mutants .....	35
Figure 2.12 SSL-1 M11 series mutants .....	36
Figure 2.13 SSL-3 M11 series mutants .....	37
Figure 2.14 M1, M3, and M4 series mutants follow-up <i>in vitro</i> assays .....	39
Figure 2.15 M1, M3, and M4 series mutants Western blots .....	40
Figure 2.16 M1, M3, and M4 series mutants follow-up <i>in vivo</i> assays .....	41
Figure 2.17 M8 and M11 series mutants <i>in vivo</i> chemical profiling to assess	

contribution of highly conserved residues to first, second, and both steps of squalene and botryococcene biosynthesis .....	42
Figure 2.18 GC-MS analysis of wild-type and mutant triterpene synthase product profiles .....	45
Figure 2.19 Purified wild-type and mutant triterpene synthase quality control assays .....	47
Figure 2.20 Relative efficiency assays for squalene (red line) and botryococcene (blue line) formation by M3-11 and M4-10 mutants .....	48
Figure 2.21 3D models of BbSS and SSL-3 with PSPP docked into active site showing residues known to affect product specificity as spheres .....	51
Figure 3.1 Two-step reaction mechanism for the biosynthesis of squalene and botryococcene and possible directed evolutions for SSL enzymes .....	61
Figure 3.2 Using <i>S. cerevisiae</i> as a selection platform for directed evolution of SSL enzymes .....	64
Figure 3.3 Architecture of squalene synthase in the eukaryotic cell .....	65
Figure 3.4 Complementation and growth testing of ScSS and BbSS full-length, truncated, and chimeric constructs in ZX178-08 on four different media types .....	68
Figure 3.5A Complementation, growth, and chemical profile analysis of SSL-1 and SSL-2 single-gene constructs in ZX178-08 and ZXD yeast lines .....	71
Figure 3.5B Complementation, growth, and chemical profile analysis of SSL-1 and SSL-2 co-expressed constructs in ZX178-08 and ZXD yeast lines .....	72
Figure 3.5C Complementation, growth, and chemical profile analysis of SSL-1 and SSL-2 gene-fusion constructs in ZX178-08 and ZXD yeast lines .....	74
Figure 3.6A Complementation, growth, and chemical profile analysis of SSL-1, SSL-3, and mtSSL-3 co-expressed constructs in ZX178-08 and ZXD yeast lines .....	76
Figure 3.6B SSL-1, SSL-3 and mtSSL-3 co-expressed constructs in ZX178-08 after fifteen days of growth .....	77
Figure 3.6C Complementation, growth, and chemical profile analysis of SSL1-SSL3 and SSL1-mtSSL3 fusion constructs in ZX178-08 and ZXD yeast lines .....	79
Figure 3.6D SSL1-SSL3 and SSL1-mtSSL3 fusion constructs in ZX178-08 after fifteen days of growth .....	80
Figure 3.7 Quantitative comparisons of yeast lines ZX178-08 and ZXD transformed with the indicated constructs for their ability to complement the <i>erg9</i> deletion .....	82
Figure 3.8A Proposed cyclization of botryococcene after epoxidation by Erg1p	

.....	85
Figure 3.8B GC-MS sterol profiling for ZX178-08 and ZXD with complementing constructs .....	86
Figure 3.8C Comparison of growth rates for the ZX178-08 yeast line expressing ScSS, SSL1-SSL3, SSL1-SSL3Yct92, SSL1-mtSSL3, and SSL1-mtSSL3Yct92 in SCglu media .....	87
Figure 3.8D Quantified GC-MS results for sterol profiling of the ZX178-08 ( <i>erg9Δ</i> ) yeast line complemented with ScSS, SSL1-SSL3, SSL1-SSL3Yct92, SSL1-mtSSL3, and SSL1-mtSSL3Yct92 .....	87
Figure 3.9 Construction of the mutant enzyme library to test the ZX178-08 selection methodology .....	91
Figure 3.10 Mutant library selection in ZX178-08 .....	92
Figure 3.10C Complementation, growth, and chemical profile analyses for mutants isolated from ZX178-08 selection of mutant library .....	93
Figure 4.1 Assessment of the strategy for squalene metabolic engineering in tobacco where enzymes were directed to the cytosol or chloroplast compartments to affect carbon flux through the MVA (cytosol) and MEP (chloroplast) pathways .....	110
Figure 4.2 Proposed metabolic engineering scheme to engineer production of botryococcene in the green alga <i>Chlamydomonas reinhardtii</i> .....	111
Figure 4.3 First-generation <i>C. reinhardtii</i> nuclear transformation vectors .....	113
Figure 4.4 Nuclear transformation efficiency in <i>C. reinhardtii</i> .....	115
Figure 4.5 Second-generation <i>C. reinhardtii</i> nuclear transformation vectors .....	117
Figure 4.6 Third-generation <i>C. reinhardtii</i> nuclear transformation vectors .....	119

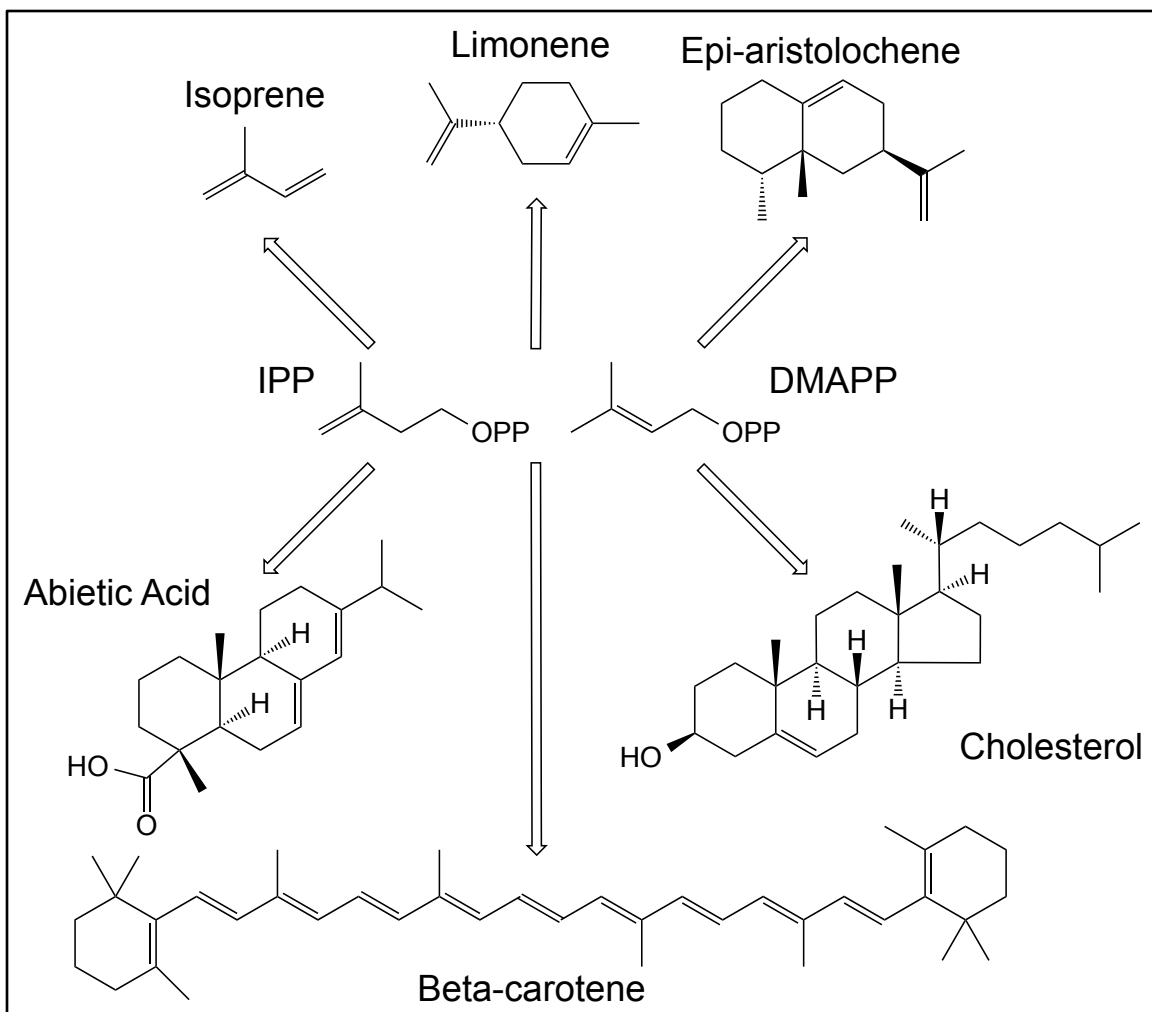
## Chapter One: Background and Introduction

### Part One: Triterpene Synthases

Terpenoids represent a diverse class of naturally occurring molecules that have potential for applications as nutraceuticals,<sup>1</sup> pharmaceuticals,<sup>2</sup> and renewable fuels.<sup>3</sup> These hydrocarbon molecules are found naturally intertwined with all aspects of life from essential primary metabolic functions in all eukaryotic organisms,<sup>4,5</sup> to more specialized roles specific to only certain species (**Figure 1.1**).<sup>6</sup> Triterpenes are a distinct subclass composed of six isoprene units (30 carbons) that arise from the joining of two molecules of farnesyl diphosphate (FPP); some very fundamental triterpenes are squalene, dehydrosqualene and botryococcene. The enzymes that catalyze the formation of these triterpene molecules are intriguing purely for the biochemical feats they catalyze and the evolutionary paths that gave rise to their development. Because these enzymes hold potential pharmacological applications (squalene and dehydrosqualene synthases) and their reaction products have a multitude of industrial uses (biofuels, cosmetics), squalene synthases and squalene synthase-like enzymes are attractive research targets.

Squalene and dehydrosqualene are similarly linked 1'-1 molecules differing only in the presence of a double bond at the 1'-1 linkage in dehydrosqualene (**Figure 1.2**). They are investigated extensively because of their medical relevance to cholesterol formation in humans<sup>7</sup> and virulence factor formation in *Staphylococcus aureus*,<sup>8</sup> respectively. Botryococcene possesses a more unusual 1'-3 linkage between the two farnesyl chains<sup>9</sup> making it more branched than squalene and dehydrosqualene. Botryococcene and its methylated derivatives are made efficiently by the extant green alga *Botryococcus braunii* race B and hold great potential as a renewable energy feedstock.<sup>9</sup> These molecules are known to have contributed to current oil and coal shale deposits around the globe.<sup>10-12</sup> Oil produced and extracted from *B. braunii* can be used to produce liquid fuels that are compatible with modern combustion engines and the currently available infrastructure for refining, storage, and transport.<sup>13,14</sup> Squalene-containing oil produces high-value fuels from hydrocracking and distillation as well.<sup>15</sup> Given the structural similarities of dehydrosqualene with squalene and botryococcene, it too would appear to be suitable for these purposes.

Formation of these molecules has been studied synthetically<sup>16,17</sup> and biosynthetically, with greater attention to the latter because of the chemical complexity associated with the biosynthetic reaction. Squalene biosynthesis has been shown to proceed through a two-step reaction mechanism in which the first step consists of a head-to-head condensation of two farnesyl diphosphate (FPP) molecules to form the stable cyclopropyl intermediate, presqualene diphosphate (PSPD) (**Figure 1.2**).<sup>18</sup> In the second step, PSPD undergoes a reductive rearrangement in the presence of NADPH to yield squalene that possesses a 1'-1 linkage between the two farnesyl moieties.<sup>19</sup> A divalent cation such as Mg<sup>2+</sup> is required for this mechanism, which appears to bridge the allylic diphosphate



**Figure 1.1** Terpenoid diversity that arises from the five-carbon isoprenoid building blocks, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP), generated by various modifications such as successive condensation of the allylic diphosphates themselves, dephosphorylation, cyclization, desaturation, reduction, oxidation, and hydroxylation.<sup>20,21</sup>





with an aspartate-rich region near the opening of the active site in crystal structures.<sup>25,26</sup> Other metal cofactors such as  $Mn^{2+}$  have been reported to substitute for  $Mg^{2+}$  but are less efficient ( $k_{cat} = 3.3 \text{ s}^{-1}$  with  $Mg^{2+}$  vs.  $0.4 \text{ s}^{-1}$  with  $Mn^{2+}$ ).<sup>27</sup> Formation of dehydrosqualene occurs through a similar mechanism where the first reaction step proceeds *via* PSPP that is subsequently rearranged in the second step to yield an unsaturated 1'-1 linkage.<sup>23,24</sup> Huang and Poulter (1989) suggested that because of its structural similarity to squalene, botryococcene biosynthesis could also occur *via* an analogous reaction mechanism with the initial reaction proceeding through PSPP, followed by a reductive cleavage between the 1'-2 bond of the cyclopropyl ring to yield the ethyl branch in the 1'-3 linked botryococcene product.<sup>28</sup>

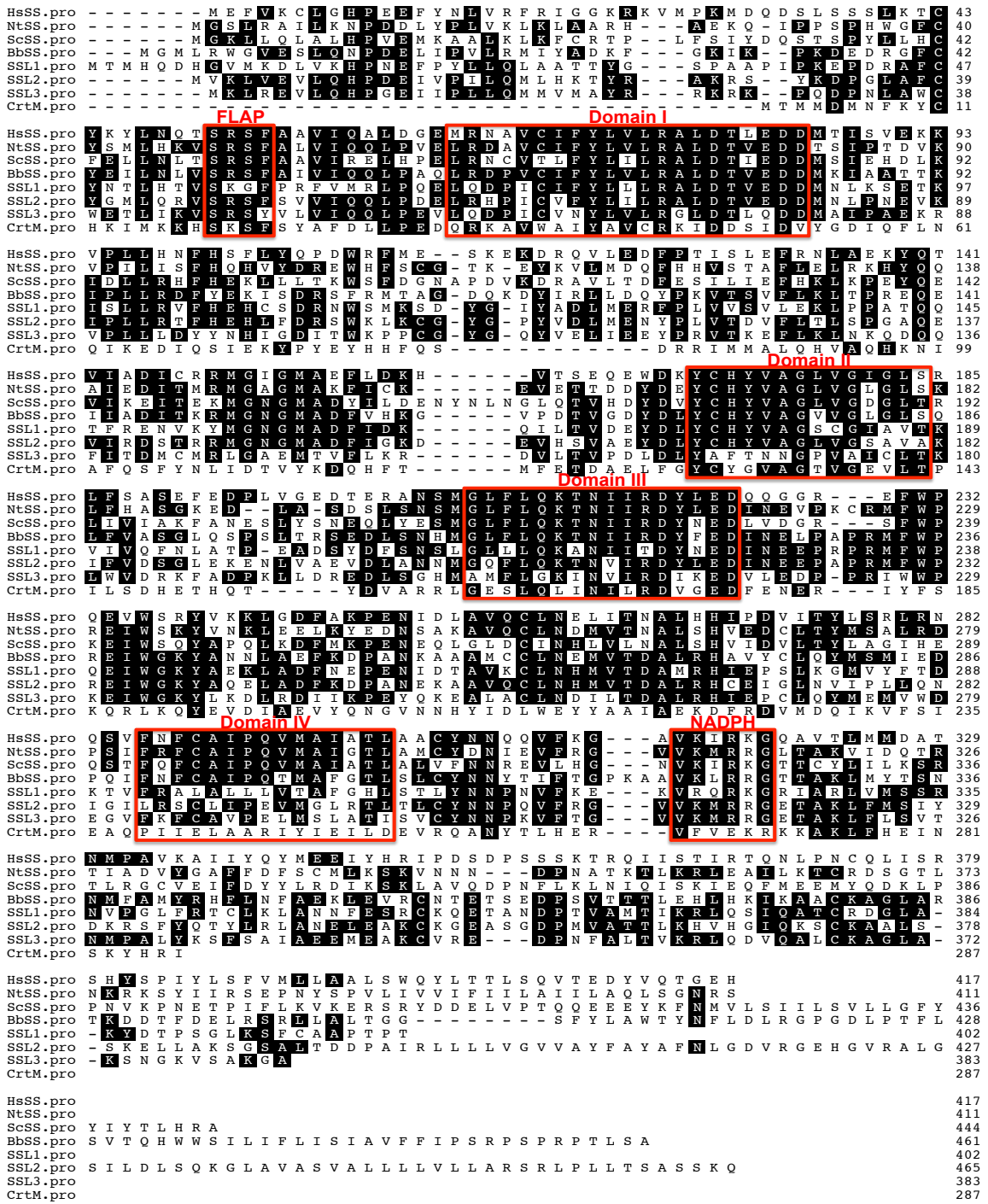
Squalene synthase genes have been isolated from a wide range of eukaryotic organisms including mammals, plants, and fungi. Interestingly, several amino acid sequences within functionally characterized squalene synthases are highly conserved across genera and even kingdoms, and ultimately have been correlated with particular steps in the squalene synthase reaction mechanism. For instance, Robinson *et al.* (1993) identified six domains based on sequence comparisons between fungal and animal squalene synthases and phytoene synthases from plants.<sup>29</sup> Phytoene synthases catalyze the condensation of two C20 prenyl diphosphate intermediates, geranylgeranyl diphosphate (GGPP) to the C40 product phytoene, which like squalene, dehydrosqualene and botryococcene occurs *via* a cyclopropyl intermediate. The importance of these conserved domains was revealed when they were shown to line the interior surface of the putative active site pocket in the human squalene synthase (HsSS) crystal structure solved by Pandit *et al.* (2000).<sup>30</sup> Gu *et al.* (1998) functionally defined the analogous domains to be associated with either the first and/or second step reactions based on site-directed mutations that altered or eliminated catalytic activities in the rat enzyme.<sup>31</sup>

Enzymes responsible for dehydrosqualene and botryococcene formation have also been identified and characterized. The Götz laboratory has reported the successful identification and characterization of the enzymes contributing to staphyloxanthin formation in *S. aureus*.<sup>32,33</sup> The *S. aureus* dehydrosqualene synthase (CrtM) was originally identified by screening a chromosomal DNA library in *S. camosus*, an unpigmented strain, for clones that caused yellow/orange coloration. Restriction mapping and sequencing of positive clones led to the deduced coding sequence for CrtM that was verified for activity and dehydrosqualene production with an *in vitro* enzyme assay.<sup>33</sup>

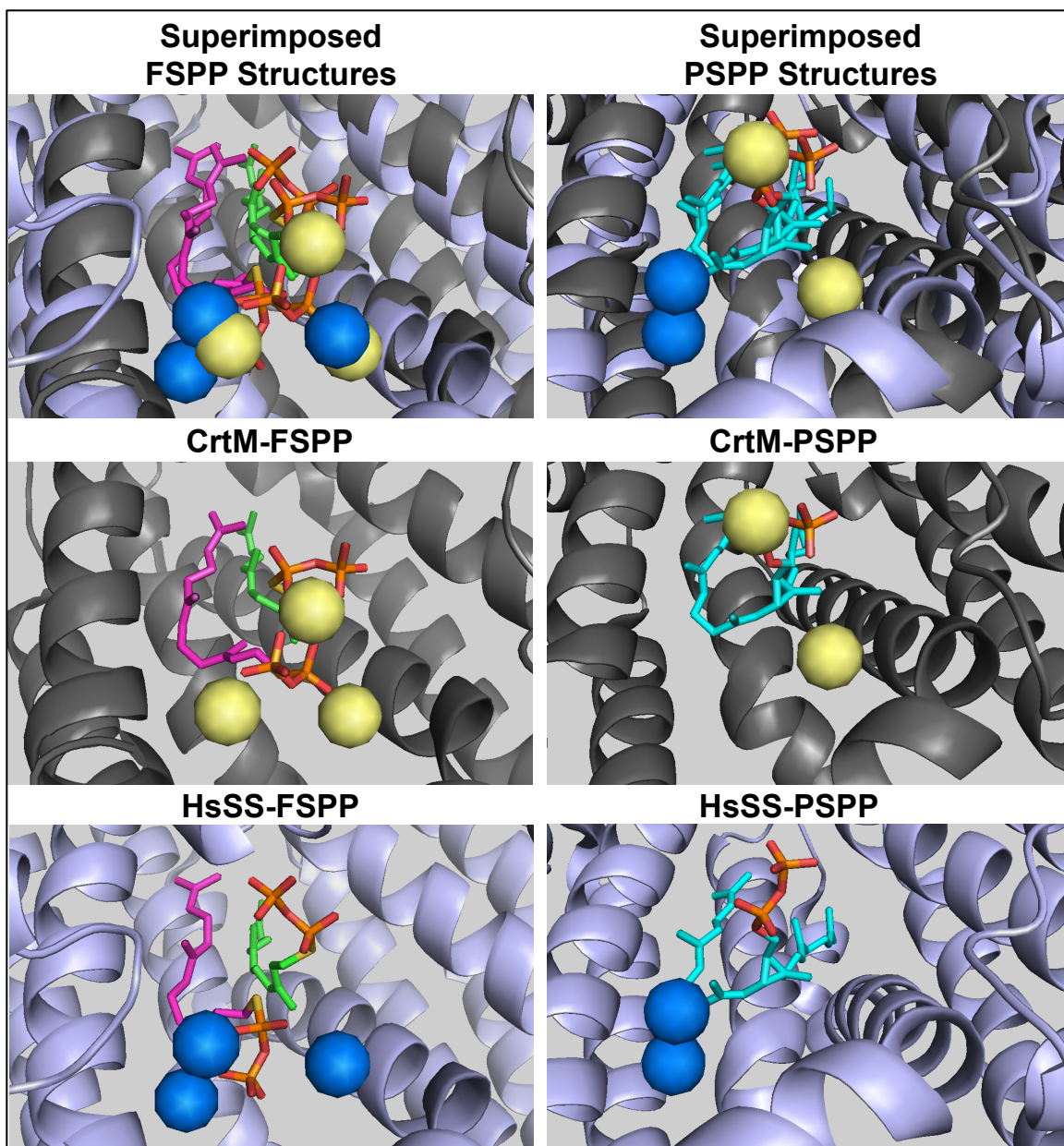
Efforts in the Chappell laboratory were directed towards the isolation of genes coding for botryococcene synthase enzyme activity by focusing on the isolation of squalene synthase and squalene synthase-like genes from *B. braunii*, followed by biochemical characterization using *in vitro* and *in vivo* assessments. Various techniques were used to obtain candidate genes for *B. braunii* squalene synthase (BbSS) and the squalene synthase-like (SSL) enzymes including degenerate oligonucleotides, radiolabeled cDNA probes, and transcriptome sequencing.<sup>34-36</sup> Collectively, this work led to the isolation and characterization of BbSS and three SSL enzymes (SSL-1, SSL-2, and SSL-3). Unexpectedly,

Niehaus *et al.* (2011) found that the SSL enzymes accomplish squalene and botryococcene biosynthesis through the activity of two pairs of these three SSL enzymes – SSL-1 and SSL-2 catalyze squalene biosynthesis while SSL-1 and SSL-3 form botryococcene.<sup>36</sup> This was surprising because up to that time, no enzymes were known that catalyzed only one step of triterpene biosynthesis. Protein sequence alignments of squalene synthases from different kingdoms, the *B. braunii* triterpene synthases, and *S. aureus* CrtM are shown in **Figure 1.3** with important regions denoted.

Recent crystallographic studies of CrtM and HsSS solved with various substrates and inhibitors revealed the structure-function relationship in these enzymes. CrtM structures solved by Lin *et al.* (2010) with an unreactive thio-FPP analog and the stable intermediate PSPP have demonstrated that the <sup>48</sup>DDxxD<sup>52</sup> motif (S1 site) acts as the prenyl donor site where diphosphate ionization occurs while the <sup>172</sup>DxxED<sup>176</sup> motif (S2 site) acts as the prenyl acceptor (**Figure 1.4**).<sup>26</sup> CrtM structures solved with PSPP in the active site showed the diphosphate moiety bound to the S1 site. This result suggested that after PSPP formation, the diphosphate relocated from the S2 to the S1 site where the second ionization could occur to initiate the second-step reaction of dehydrosqualene formation. More recent crystallographic studies with HsSS by Liu *et al.* (2014) corroborated the contention that the S1 and S2 sites were the prenyl donor and acceptor sites, respectively.<sup>25</sup> Furthermore, Liu *et al.* (2014) also proposed a substrate-binding order based on their crystallographic studies and observations made from previous kinetic studies – the S2 site binds the first FPP molecule followed by binding of the second FPP in the S1 site.<sup>25,37</sup> An important difference in the CrtM and HsSS structures obtained was noted wherein after PSPP formation, the diphosphate of PSPP does not appear to relocate to the S1 site in HsSS. Only when HsSS Y73A was solved with Mn<sup>2+</sup> and PSPP was it observed that the diphosphate of PSPP shifted from the S2 to S1 site. Liu *et al.* (2014) suggested this was a result of the additional flexibility of the active site pocket near the S1 substrate-binding site provided by the Y73A mutation. They hypothesized that in the wild-type HsSS enzyme, the S2 to S1 shift occurs but only when NADPH is bound to cause a conformational change. HsSS structures with NADP<sup>+</sup> or its analogues were unable to be obtained as part of these studies. Interesting to note is an observation made where *in vitro* incubations of squalene synthase with Mn<sup>2+</sup> in the absence of NADPH resulted in dehydrosqualene as the dominant reaction product.<sup>24</sup>



**Figure 1.3** ClustalW multiple sequence alignment (gap penalty: 22 and gap length penalty: 0.75) of triterpene synthases. Areas boxed in red represent conserved regions of the triterpene synthases thought to line the active site pocket and/or contribute to enzymatic activity. *Homo sapiens* squalene synthase (HsSS), *Nicotiana tabacum* squalene synthase (NtSS), *Saccharomyces cerevisiae* squalene synthase (ScSS), *Botryococcus braunii* squalene synthase (BbSS), *Botryococcus braunii* squalene synthase-like (SSL-1, SSL-2, SSL-3), *Staphylococcus aureus* dehydrosqualene synthase (CrtM).



**Figure 1.4** Similarities between CrtM and HsSS crystal structures solved with the unreactive substrate analogue farnesyl thiodiphosphate (FSPF) and the stable intermediate PSPF.<sup>25,26</sup> FSPF substrate colored in pink resides in the S1 site; green FSPF is located in the S2 site. Prenyl substrates are positioned generally similar in both active sites; note the differences between locations of Mg<sup>2+</sup> ions (CrtM – yellow spheres, HsSS – blue spheres) suggested to be a result of differences in the second-step reaction for these enzymes.

Understanding triterpene synthases at the molecular level has important implications for the development of small molecules that inhibit enzyme activities. The Oldfield laboratory has investigated and been successful in using inhibitors that target both dehydrosqualene synthase and squalene synthase (dual-target treatment) for controlling *S. aureus* infection by 1) inhibiting dehydrosqualene synthase, an important enzyme in the staphyloxanthin biosynthetic pathway, and 2) increasing host innate immunity by inhibition of squalene synthase which promotes formation of neutrophil extra-cellular traps (NETs).<sup>38</sup> Knowledge of the crystal structures for these enzymes coupled with *in silico* screening of small molecule libraries allowed identification of candidate inhibitors that were rationally modified with the goal of improving their therapeutic index, a measure that takes into account how effective a given inhibitor is as a dual-target treatment. Empirical tests of the modified inhibitors demonstrated the success of this approach with a ~100-fold increase in the so-called therapeutic index (TI). With increasing incidence of *S. aureus* related deaths and antibiotic resistance on the rise, such knowledge will continue to pay dividends.

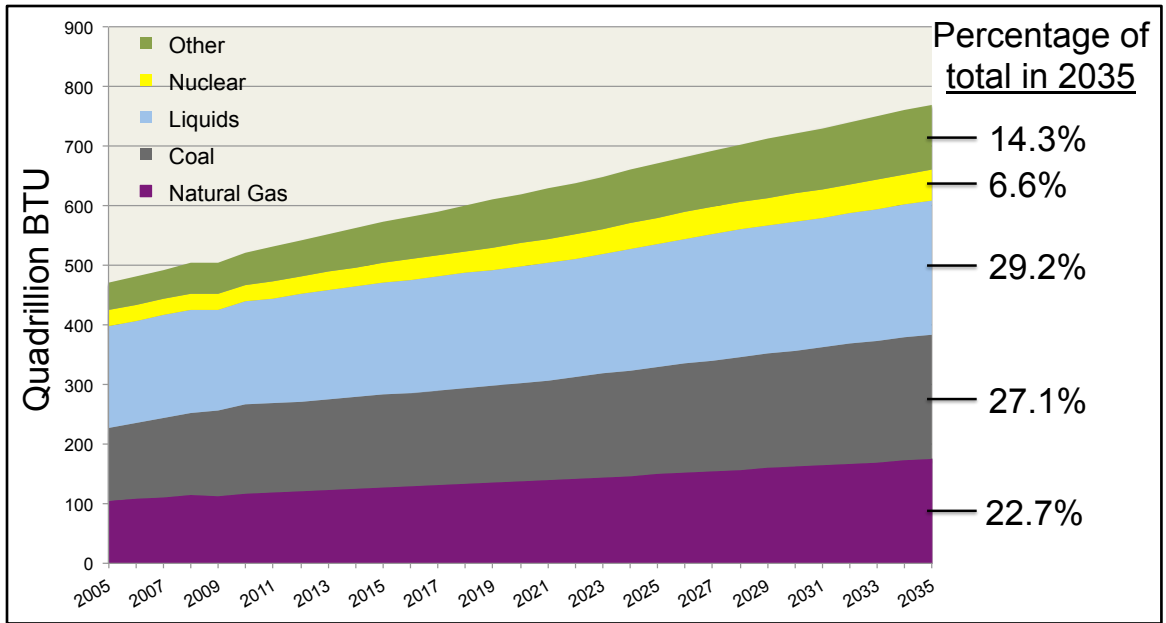
Alternatively, a molecular understanding of these enzymes might enable mutant enzymes to be created with improvements in catalytic efficiency or with novel activities. While mutant enzymes often have lower catalytic efficiencies than their wild-type counterparts, creating mutant enzymes that produce alternative products is not uncommon.<sup>39</sup> For example, studies with the sesquiterpene cyclase, epi-isozizaene synthase (EIZS) defined a single amino acid mutation that resulted in a 100-fold decrease in catalytic efficiency compared to wild type, but yielded a new product,  $\beta$ -curcumene which can be hydrogenated to produce bisabolane, a potential biofuel. While a bisabolene synthase is known, the mutant EIZS catalyzes the bisabolane precursor 68-fold more efficiently.<sup>39</sup> Subjecting the mutant EIZS to subsequent mutations could improve  $\beta$ -curcumene production by either promoting higher product specificity for  $\beta$ -curcumene or increasing catalytic efficiency. Enzymatic improvements such as these could be important for metabolic engineering efforts aimed at increasing product accumulation levels of these desirable compounds in heterologous hosts.

## Part Two: Triterpene Metabolic Engineering

World population is projected to reach 9.6 billion people by 2050.<sup>40</sup> Natural resource consumption increases are sure to follow. Among these resources, liquid fuels rank at the top of the list in relation to coal and natural gas usage (**Figure 1.5**) (EIA, 2011). Speculatively, demand for these liquid fuels is unlikely to ever wane. Large construction projects will always be dependent upon diesel-powered, heavy machinery as just one example. Moreover, the massive transportation fuel infrastructure currently available is designed around hydrocarbon-based liquid fuels. Rather than develop and implement new fuel sources such as hydrogen gas and the necessary infrastructure to support it, a sustainable means to produce renewable hydrocarbon fuels would be highly desirable.

The triterpenoids biosynthesized by the extant green alga *Botryococcus braunii* race B are suggested to have contributed to current day coal<sup>10</sup> and oil<sup>11</sup> shales. Additionally, it has been shown that they are an excellent feedstock for the hydrocracking process already used to produce liquid transportation fuels.<sup>14</sup> However, the slow growth rate of *B. braunii* has hampered the mass culture of this alga for production of large amounts of the triterpenoids. To circumvent this problem, the genetic blueprint has been investigated with the intention to mobilize and deploy it in alternative hosts for robust production of these valuable products. As already discussed, genes that catalyze botryococcene have been characterized and even shown to be functional in bacteria, fungi, and higher plant hosts for producing botryococcene. While these demonstrations are all critically important in the overall goal of renewable energy production platforms, other systems such as algae need to be evaluated.

Algae are a polyphyletic group of water dwelling, eukaryotic single and multicellular organisms with diverse morphological, physiological, and biochemical features. Similar to plants, algae fix carbon dioxide through photosynthesis that can ultimately be directed to the biosynthesis of numerous down-stream products such as triacylglycerides, starch, and terpenoids. Mass algae cultures have been proposed as a means for production of large amounts of terpenoids and other high value chemical targets that are useful in nutraceutical,<sup>1</sup> pharmaceutical,<sup>2</sup> and renewable fuel<sup>3</sup> applications without competing for arable farmland used in food production. Algae can achieve higher cellular densities than plants enabling them to produce more compound per given unit of land area.<sup>41</sup> Several species are already cultivated on a large scale for the production of tetraterpene carotenoids with useful antioxidant and pigmentation properties, and others such as *Botryococcus braunii* race B are under consideration for renewable fuel production.<sup>42,43</sup> Higher titers for these terpenoids and other products are always desirable, and metabolic engineering offers one route to accomplish this goal.<sup>41,44</sup> Currently, no single alga can be labeled as superior for metabolic engineering in general or for terpenoid engineering in particular. However, from a biotechnological point of view *Chlamydomonas reinhardtii* presents the best opportunities because; 1) a large



**Figure 1.5** World total energy consumption by fuel type according to EIA as of 2011.

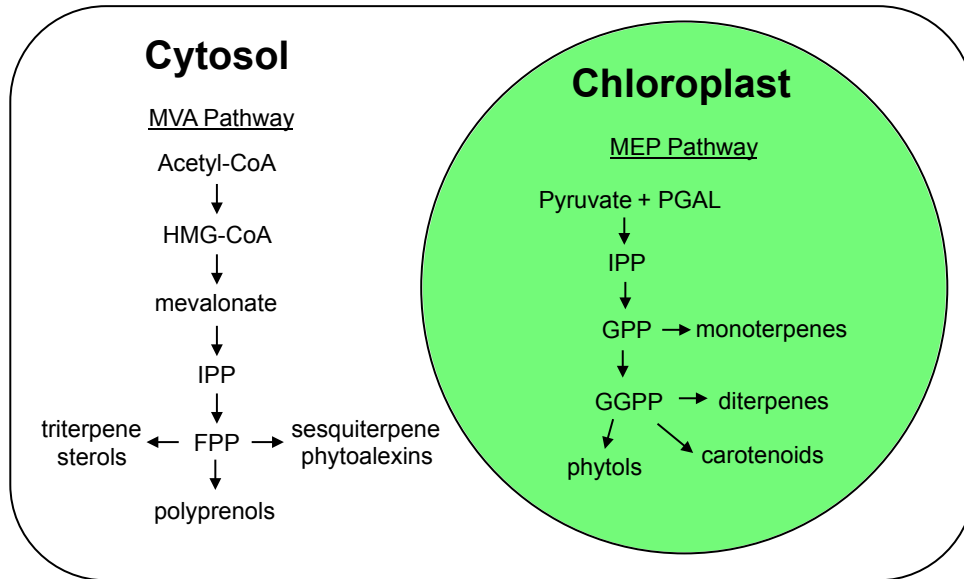


number of engineering tools have been developed for this alga; 2) it is the best studied alga 'omically,' physiologically, and biochemically; and 3) anecdotally, it is the alga where the most effort to engineer terpenoid biosynthesis has occurred.

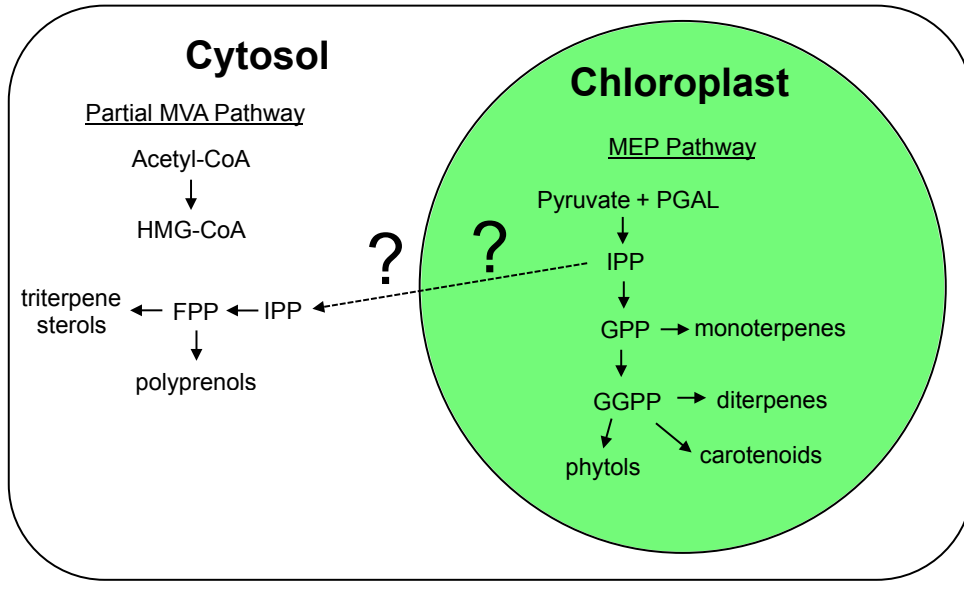
Two pathways produce the C5 isoprenoid building block isopentenyl diphosphate (IPP): the mevalonate (MVA) pathway and methyl erythritol phosphate (MEP) pathway (**Figure 1.6**). In the green algal lineage (Chlorophyta), *C. reinhardtii* and other members are known to possess only the MEP pathway based on biochemical studies where labeling patterns of terpenoid products were examined and shown to be specific for only MEP-derived isoprenoids.<sup>45</sup> Other studies that computationally analyzed green algal genomes for the presence of MVA pathway related genes demonstrated their absence.<sup>46,47</sup> Higher plants however, are well documented to possess both pathways.<sup>48,49</sup> The MVA pathway operates in the cytoplasm and is dedicated to terpenoids derived from the C15 intermediate FPP. These include sterols, dolichols (for glycosylation reactions) and substrates for protein prenylation. The MEP pathway operates exclusively in the chloroplast and is solely responsible for monoterpene, diterpene and carotenoid biosynthetic precursors. Work in the Chappell laboratory has assessed the efficiency of novel terpene biosynthesis and accumulation in transgenic plants *via* these two pathways by directing the relevant enzymes to the cytoplasm *versus* the chloroplast compartments.<sup>50,51</sup> This was accomplished by introducing the respective genes into the nuclear genome with and without chloroplast targeting signal sequences appended to the 5' ends of the corresponding genes and then chemically profiling the resulting transgenic lines. Building a new metabolic channel for diverting IPP from the MEP pathway in the chloroplast into sesquiterpenoids and triterpenoids has generally been the most successful approach in higher plants.

A similar comparison in algae should be possible despite known differences in terpene metabolism between green algae and higher plants. Information regarding the flux of carbon from this pathway into the cytosolic compartment is scant. Whether the algal MEP pathway has the capacity to support additional isoprenoid biosynthesis, either in the cytoplasm or in the chloroplast, remains to be determined. Thus, a study in green algae whereby triterpene synthases are targeted to either the cytosol or chloroplast compartments would be informative not only for the ability of these algae to produce hydrocarbon molecules that would be useful as renewable energy sources, but also for information as to how these algae may be regulating carbon flux through the MEP pathway.

# Higher Plant Cell



# Green Algae Cell



**Figure 1.6** Generic depiction of plant and green algae cells showing locales of the MVA and MEP pathways and products that typically arise from them.

## Aims of Dissertation

Two over-arching objectives of the current work were to 1) unravel the biochemical features dictating the catalytic outcomes of squalene synthase *versus* botryococcene synthase and 2) engineer novel biosynthetic capacity for triterpene production into *Chlamydomonas reinhardtii*. Accomplishing these aims utilized various molecular manipulation techniques and the incredible body of knowledge generated over the past five decades.

Chapter 2 focuses on enzymological studies of *B. braunii* squalene synthase and squalene synthase-like enzymes in order to dissect the molecular features that promote their various catalytic specificities. Techniques used to accomplish this work include rational design of mutants guided by protein sequence alignments and three-dimensional homology models. Individual mutants were made and tested for changes in substrate and product specificities with an *in vitro* assay (bacterial expressed protein), followed by *in vivo* chemical profiling (expression in *E. coli* line that makes additional FPP) for mutants of interest. A mutant SSL-3 enzyme is described that has been altered in its product specificity.

Building on efforts described in Chapter 2 for rationally designed mutagenesis, Chapter 3 describes a more unbiased approach to understanding the structure-function maps of SS and SSL enzymes. In this work, *Saccharomyces cerevisiae* that contains a squalene synthase deletion was used to demonstrate how the SSL enzymes are capable of complementing. A mutant library was subsequently constructed and selected for mutants that were able to complement. Proof-of-principle was described for future studies aimed at evolving single SSL enzymes toward becoming a fully functional squalene synthase.

Production of squalene and botryococcene has been implicated as one avenue to supply renewable liquid transportation fuels. Chapter 4 will focus on metabolic engineering efforts in *Chlamydomonas reinhardtii* to determine the most efficient means for engineering triterpene metabolism in algae. Nuclear transgene expression in algae is neither well established nor have any appreciable reports been published for algae metabolically engineered for high-level accumulation of triterpenoids. This chapter will discuss these issues and lay out an approach for how triterpene metabolism might be manipulated in *C. reinhardtii*.

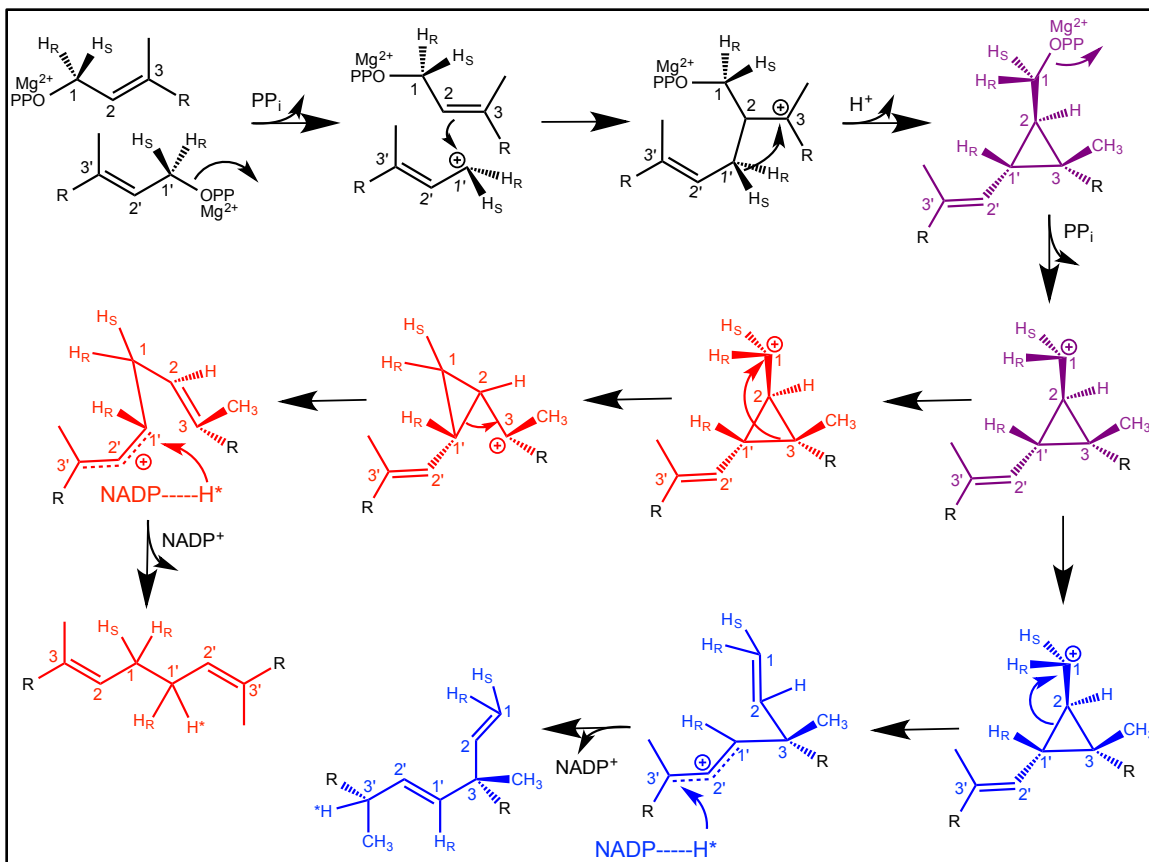
## Chapter Two: Understanding the Mechanistic Specificity Between 1'-1 and 1'-3 Triterpene Linkages Catalyzed by Squalene Synthase *versus* Botryococcene Synthase

### Background and Introduction

Squalene is a thirty carbon, acyclic, branched chain composed of six isoprene units. It is an important precursor of sterols in eukaryotes and hopanoids in prokaryotes that play critical roles in a variety of cellular functions as well as membrane physical properties.<sup>52,53</sup> Enzyme mediated catalysis of squalene occurs *via* a two-step reaction by an enzyme known as squalene synthase. Multiple groups have described the same mechanism for squalene synthase isolated from yeast, rat, and human.<sup>23,30,31,54</sup> Briefly, the proposed mechanism proceeds as two steps within the same active site where two molecules of farnesyl diphosphate (FPP) are joined in a head-to-head coupling (**Figure 2.1**). Presqualene diphosphate (PSPP) is generated during the first reaction step as a stable, cyclopropyl intermediate in an NADPH independent fashion.<sup>18,55</sup> It remains in the active site where, in the presence of NADPH, it is reductively rearranged to give the final 1'-1 linked, symmetrical squalene molecule.<sup>19,23,24</sup> Normally, squalene is the sole product observed when squalene synthase is incubated with NADPH and FPP. In the absence of NADPH with prolonged incubation times, however, squalene synthase is known to generate various 1'-1 and 1'-3 linked solvolytic products.<sup>24,56</sup>

Kinetic analyses performed with *Saccharomyces cerevisiae* squalene synthase (ScSS)<sup>19,23,24,54</sup> and detailed crystallographic studies of *Homo sapiens* squalene synthase (HsSS)<sup>30,57</sup> and *Staphylococcus aureus* dehydrosqualene synthase (CrtM)<sup>26</sup> provide insights into how FPP is bound to this enzyme and subsequently catalyzed to squalene. However, much less is known about residues that are catalytically important beyond the aspartate-rich (DxxED) motifs.<sup>31</sup> Robinson *et al.* (1993) predicted that the five conserved domains among squalene synthases are likely helices that line the active-site pocket,<sup>29</sup> this claim has since been confirmed by structural evidence.<sup>30,57</sup> Defined roles of these domains with respect to their involvement in diphosphate ionization, NADPH binding, substrate orientation and product specificity has not been described empirically. A detailed map of the residue(s) important for the first, second, or both reaction steps that correlates with kinetic and structural data is still needed. Therefore, we set out to map those residues important for one discrete catalytic function with these synthases using an activity interconversion strategy exploiting uniquely characterized triterpene synthases from an alga.

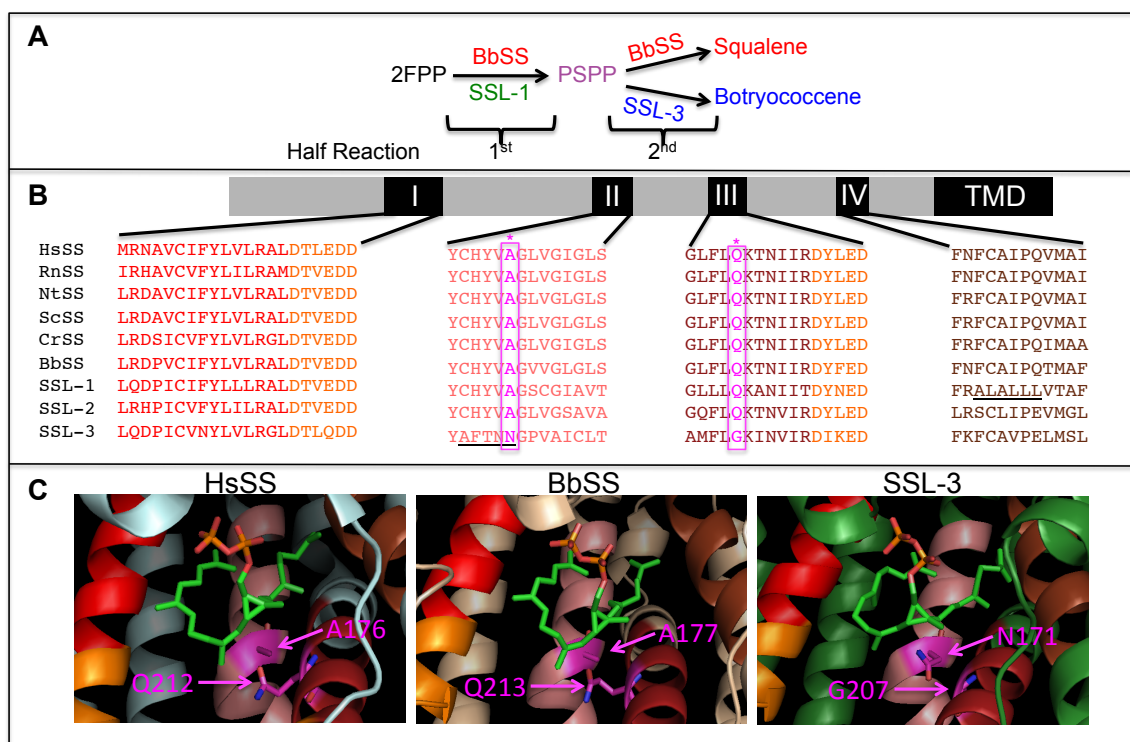
Botryococcene, a structural analog of squalene, is produced in copious amounts by the green alga *Botryococcus braunii* race B.<sup>9</sup> Due to structural similarities between botryococcene and squalene, botryococcene was predicted to arise via a squalene synthase-like reaction that proceeds through PSPP.<sup>58</sup> However, instead of the final 1'-1 linkage for squalene, a more branched, 1'-3 asymmetrical linkage between the two prenyl chains is observed for botryococcene (**Figure 2.1**).<sup>28,59,60</sup>



**Figure 2.1** Proposed reaction cascade for enzyme-catalyzed squalene and botryococcene formation. First-step reaction occurs where two molecules of FPP (black) are joined to form the stable intermediate PSPP (purple). Subsequently, in the presence of NADPH, a reductive rearrangement occurs during the second-step reaction to yield either squalene (red) or botryococcene (blue). Proton positions are reported as described by Huang and Poulter (1989).<sup>28</sup>

Recently, the biochemical pathway responsible for the formation of botryococcene was elucidated by Niehaus *et al.* (2011). Two distinct squalene synthase-like enzymes (SSL-1 and SSL-3) were found to be responsible for the two reaction steps of botryococcene biosynthesis, as opposed to the single polypeptide that was originally hypothesized.<sup>36,59,35</sup> This report capitalized on previous efforts to identify and characterize enzymes<sup>35,61</sup> that contribute to the ability of *B. braunii* to accumulate ~30% of its biomass in hydrocarbon oils.<sup>62</sup>

The biochemistry in *B. braunii* as described by Niehaus *et al.* (2011) has not been definitively demonstrated, as key experiments like SSL gene knockouts needed for proof have been hampered by a lack in methodology for genetic transformation. Nonetheless, the SSL enzymes present an interesting opportunity to aid in the identification of residues responsible for the first- and second-step reaction activities of squalene and botryococcene biosynthesis. By exploiting these enzymes, we proposed to either introduce reaction activities (*i.e.*, introduce second-step reaction capability into SSL-1) or interconvert product specificities between enzymes (*i.e.*, convert BbSS product specificity to botryococcene) in order to map functional residues and/or domains. To identify catalytically important residues, protein sequence alignments were examined in concert with three-dimensional models generated by threading the *B. braunii* enzymes onto previously determined crystal structures for squalene synthase<sup>30,57</sup> and dehydrosqualene synthase<sup>26</sup> (**Figure 2.2**). Residues differing in the highly conserved domains were substituted to match the enzyme in which we wanted to introduce a new activity. *In vitro* assays were subsequently used to screen for mutant enzyme activities, followed by *in vivo* characterization achieved by heterologous expression and chemical profiling of the triterpene products accumulating in bacterial cultures. Of the mutants that were rationally designed, constructed and tested, residues N171 and G207 of SSL-3 were identified from this approach and appeared to play a role in controlling product specificity of this enzyme whereas the corresponding residues in BbSS, A177 and Q213, were probably necessary but not sufficient for squalene product specificity.



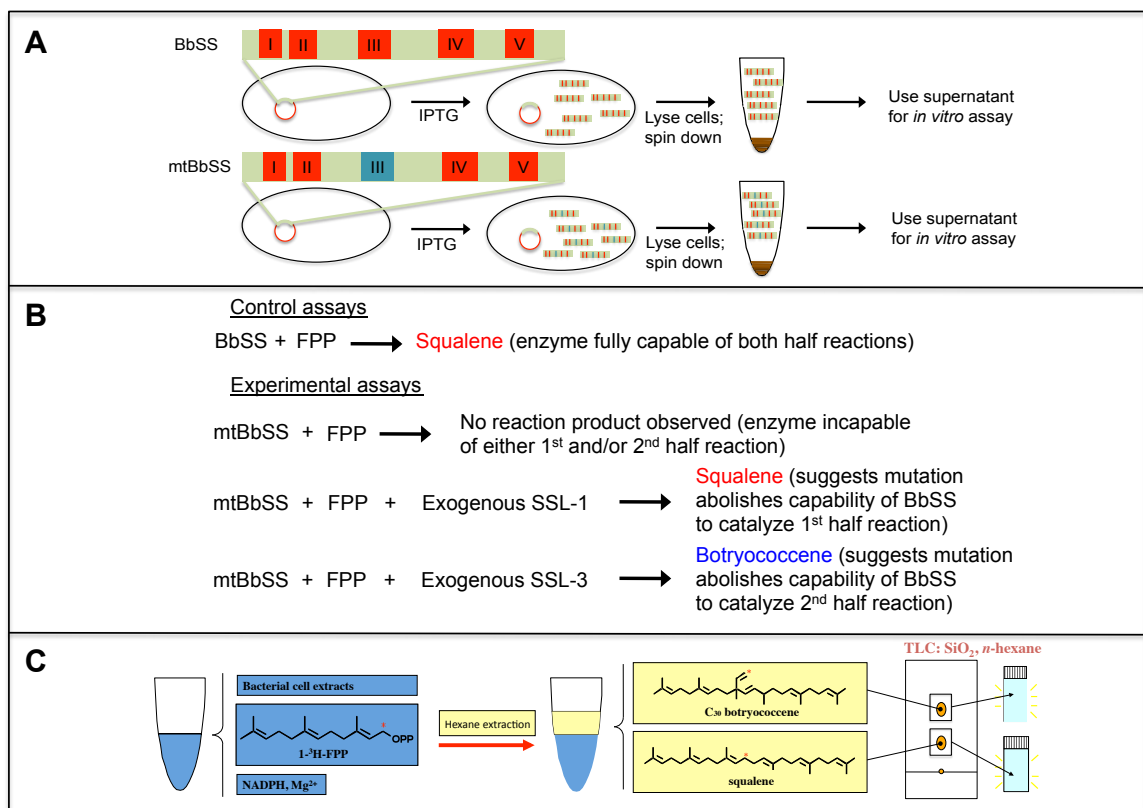
**Figure 2.2** Strategy for identification of residues controlling first- and second-step reaction activity and product specificity in triterpene synthases. Simplified biosynthetic pathway for squalene and botryococcene is depicted for the first- and second-step reactions and the enzymes responsible (A). Map of domain organization for a triterpene synthase and protein sequence alignment of the corresponding regions I-IV. Underlined positions denote regions of SSL-3 and SSL-1 that were significantly different than the squalene synthase enzymes and hypothesized to contribute to their lack of first- or second-step reaction specificities, respectively (B). Crystal structure of HsSS (PDB 3VJ8) and 3D models of BbSS and SSL-3 with PSPP docked in the active site. Residues highlighted in pink represented positions that were interchanged between BbSS and SSL-3 that control product specificity in SSL-3 (C). Abbreviations: farnesyl diphosphate (FPP); presqualene diphosphate (PSPP); transmembrane domain (TMD); *Homo sapiens* squalene synthase (HsSS); *Rattus norvegicus* squalene synthase (RnSS); *Nicotiana tabacum* squalene synthase (NtSS); *Saccharomyces cerevisiae* squalene synthase (ScSS); *Chlamydomonas reinhardtii* squalene synthase (CrSS); *Botryococcus braunii* squalene synthase (BbSS); squalene synthase-like (SSL).

## Results

*Identification of candidate residues for mutation.* Amino acid positions chosen for mutagenesis were based on several fundamental characteristics. First, the residues needed to be present within domains I-IV of the protein sequence (**Figure 2.2B**) with their R-group positioned into the active site where interaction with the substrate could occur (**Figure 2.2C**). ClustalW sequence alignments of squalene synthases from various kingdoms along with the *B. braunii* triterpene synthases were analyzed in the first step to identify candidate residues. Positions differing in these regions were analyzed in a second step using 3D models generated with MODELLER 9v8.<sup>63</sup> Originally, our *B. braunii* triterpene synthases were threaded onto the HsSS structure (PDB 1EZf), which was solved at a resolution of 2.15Å.<sup>30</sup> Other higher resolution structures have since been published for HsSS (PDB 3VJ8) as well as structures for CrtM (PDB 3ADZ, 2ZCO).<sup>26,57</sup> We used these structures to generate additional threaded models of the *B. braunii* triterpene synthases to obtain the highest quality models possible with no significant differences observed between any of the models.

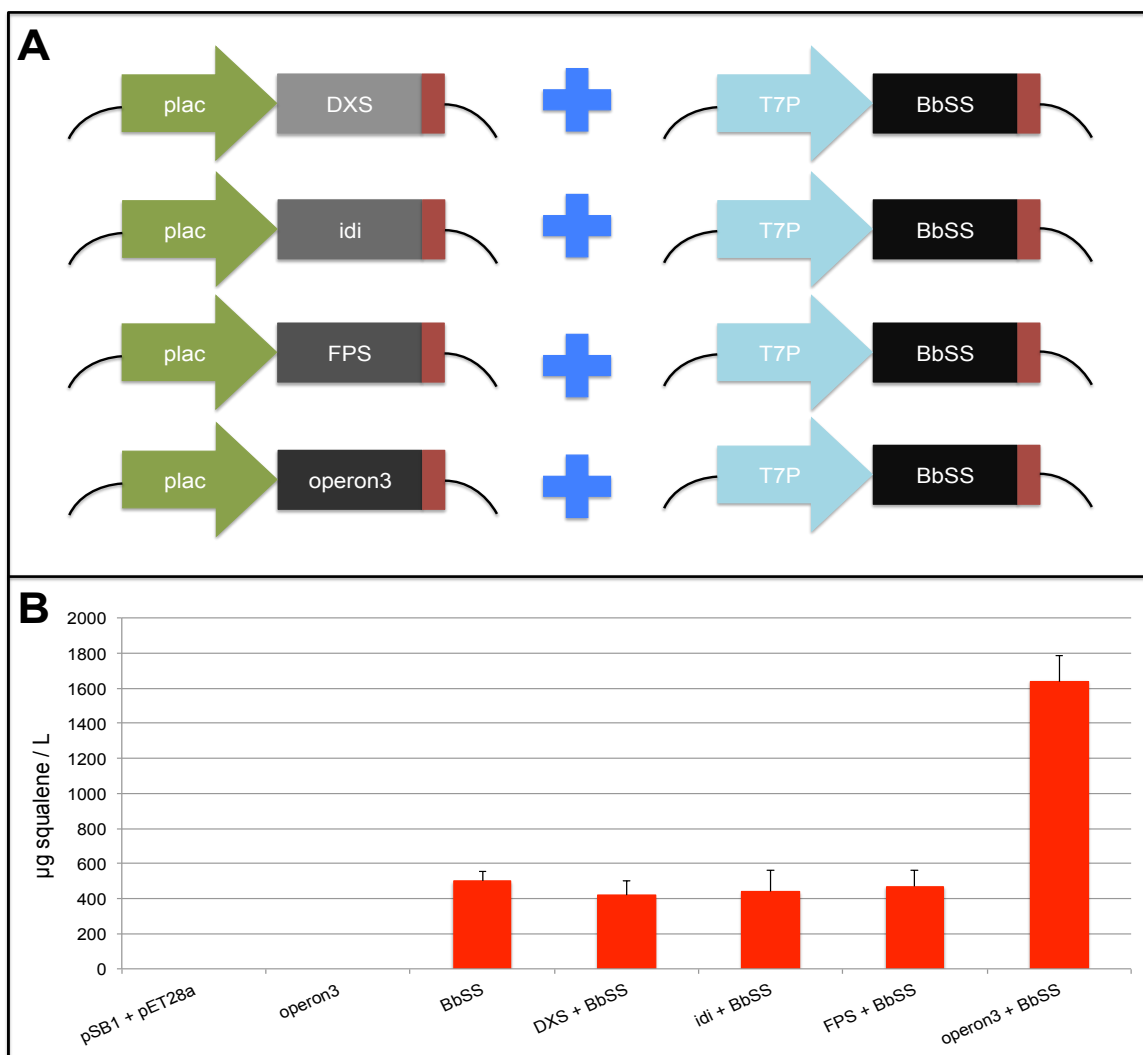
*Rationale for in vitro activity screening of wild-type and mutant triterpene synthases.* Because BbSS is a bifunctional enzyme that can catalyze predominantly PSPP or squalene given the appropriate assay conditions, we desired to assay both reaction step capabilities for all mutants because of the possibility one reaction step or the other may have been introduced or altered. However, our assay only detects hexane extractable products from the second-step reaction of these enzymes so a mutant BbSS enzyme compromised in catalyzing PSPP but still capable of second-step reaction activity would be overlooked because we provide only FPP as a substrate. Another possibility exists where a mutant BbSS has lost second-step reaction capability but still produces PSPP not detectable by our assay causing more valuable information to go unseen. To overcome these technicalities, we chose to perform co-incubations with either SSL-1 (to provide PSPP) or SSL-3 (to convert PSPP to botryococcene) rather than attempting arduous synthetic means to produce enough PSPP for use as substrate in second-step reaction assays or labor intensive, time consuming extractions of PSPP followed by dephosphorylation to assay for first-step reaction capability. Although all of these techniques have been documented,<sup>36,64</sup> we felt the most direct means to screen mutants for both reaction step activities would be a co-incubation or coupled assay approach (**Figure 2.3**). An important caveat of these assays is the use of bacterially expressed proteins. Despite standardization of expression technique, variation in assay results was always possible between experiments. Therefore, control constructs were always expressed and assayed in parallel with mutants to allow for normalization across experiments.





**Figure 2.3.** Coupled assay approach to screen mutant triterpene synthases for the first, second, and both steps of the squalene/botryococcene synthase reactions. Wild-type and mutant enzymes are over-expressed in bacteria followed by cell lysis *via* sonication and centrifugation to clear lysate (A). The crude extract is subsequently used in a series of assays as illustrated in (B) to assay for the given enzyme's ability to catalyze either the first, second or both steps of the reaction. Schematic depiction of radiolabeled substrate based activity assay for triterpene synthases.  $^3\text{H}$ -FPP is incubated with enzyme(s) followed by hexane extraction. TLC is used to separate extractable products and the zones corresponding to botryococcene and squalene (boxed areas) are scraped from TLC plate and analyzed by scintillation spectroscopy (C).

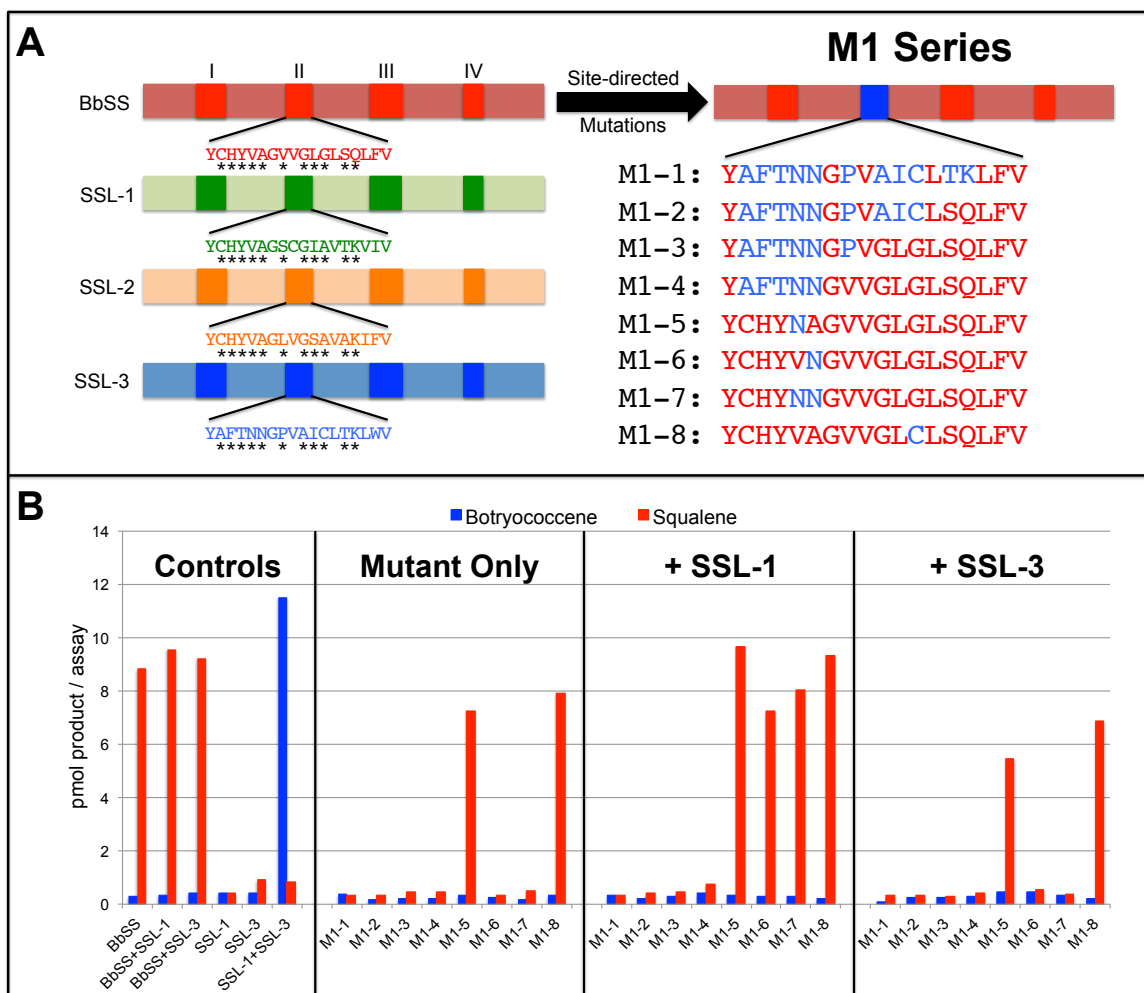
*Overexpression of key MEP pathway biosynthetic genes in E. coli.* For verification of *in vitro* assay results obtained for mutant triterpene synthases in this study, an *in vivo* system in *E. coli* was developed because it would be compatible with pET28a expression vectors that already contained mutant enzyme sequences. The approach was to overexpress key MEP biosynthetic genes responsible for supplying the five-carbon building blocks of isoprenoid biosynthesis: IPP and DMAPP. pSB1:operon3 was constructed to encode a synthetic operon containing 1-deoxy-D-xylulose-5-phosphate synthase (DXS), isopentenyl diphosphate isomerase (*idi*), and farnesyl diphosphate synthase (FPS) under control of a lactose inducible promoter (*plac*). To prove that the system was functional as well as examine if DXS, *idi*, or FPS contributed more carbon flux into squalene, each enzyme of the operon was expressed individually with BbSS (**Figure 2.4**). No background levels of squalene were observed with empty vectors (pSB1 + pET28a) or pSB1 expressing operon 3; BbSS expressed alone, however, did accumulate squalene. BbSS expressed individually with DXS, *idi*, or FPS did not accumulate additional squalene in comparison to BbSS only. Squalene titers were highest only when the entire operon 3 was expressed with BbSS.



**Figure 2.4** *In vivo* determinations for which of four enzymes affects carbon flux into squalene in *E. coli* BL21 (DE3). Constructs used for study: plac, lac promoter in pSB1 vector; DXS, 1-deoxy-D-xylulose 5-phosphate synthase; idi, isopentenyl diphosphate isomerase; FPS, farnesyl diphosphate synthase; operon 3, operon containing DXS, idi, and FPS; T7P, T7 promoter in pET28a vector; BbSS, *Botryococcus braunii* squalene synthase (A). Constructs described in (A) were expressed in BL21 for 24 hours followed by lysis with acetone, hexane extraction, and GC-MS product profiling. Results are representative of three biological replicates from three independent experiments (n=3) (B).

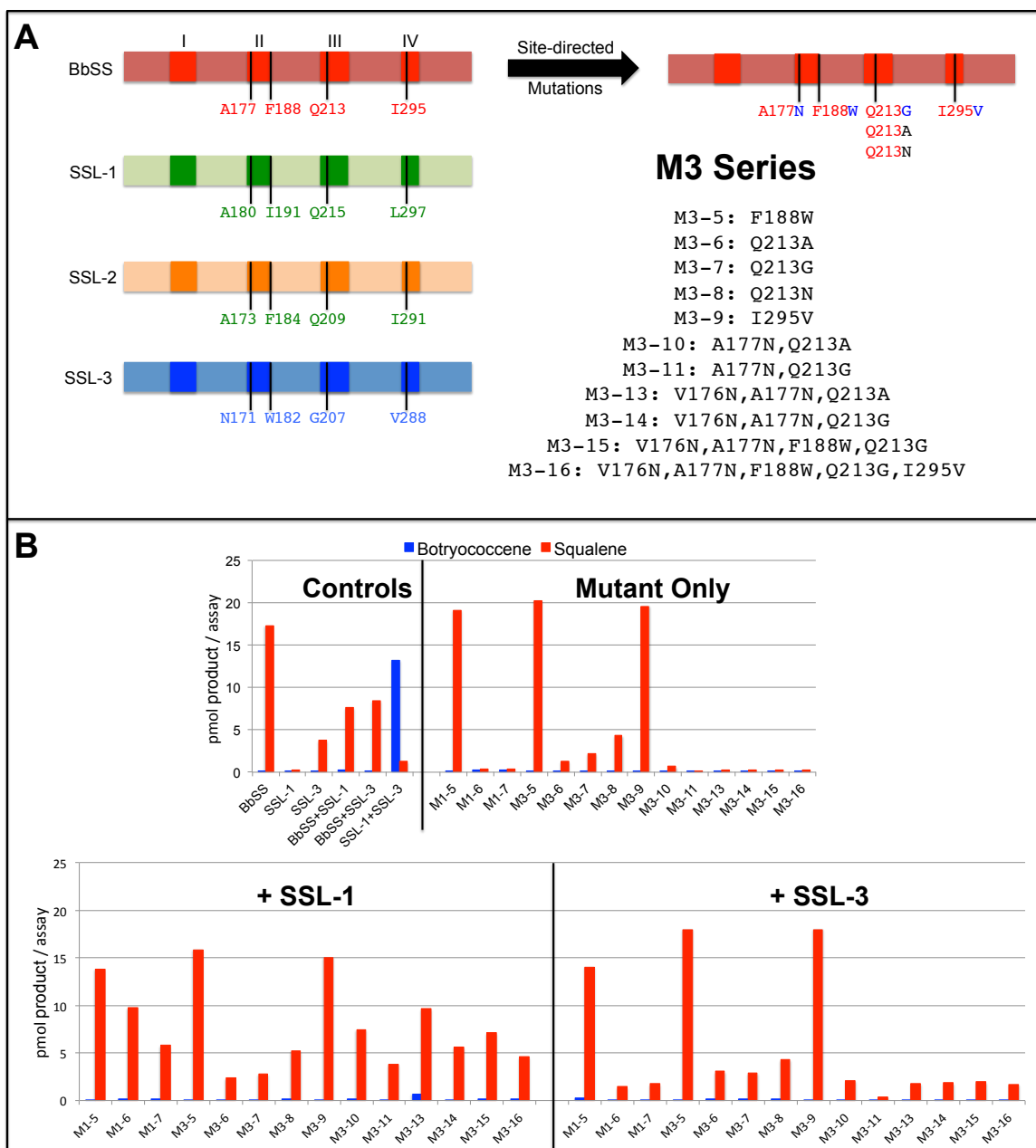
*Triterpene synthase permutations.* Efforts were undertaken throughout the course of this work to swap all combinations of domains between BbSS and SSL-1 (very early effort; M1-M16 mutants), BbSS and SSL-3 (M2 series), and SSL-1 and SSL-3 (M6 series) based on the conserved domains as defined in **Figure 2.2B**. These three series of permutations containing 16 mutants each (4 domains, 2 enzymes =  $4^2 = 16$  mutants) yielded enzymes that were not catalytically active and expressed poorly in *E. coli* (data not shown). The poor results led us to focus on the site-directed mutations described in the sections to follow.

*M1 series mutants.* Initial mutants were constructed based on differences in domain II of BbSS and SSL-3. This region exhibits high sequence homology between all squalene synthases, including SSL-1 and SSL-2, but harbors considerable differences in the corresponding area of SSL-3 (**Figure 2.2B**, underlined sequence). Therefore, this region was proposed to be responsible for mediating product specificity for squalene or botryococcene in BbSS and SSL-3, respectively. The M1 series mutants were constructed in domain II of the BbSS backbone because these mutations would interconvert product specificity from squalene to botryococcene. Eight mutants were made containing various mutation(s) based on corresponding differences in the same position for SSL-3 (**Figure 2.5**). *In vitro* assay results suggested that BbSS A177 was critical in the first-step reaction activity based on results for M1-6 – no squalene synthase activity was detected in the mutant-only assays, but was recovered when co-incubated with SSL-1, which provides PSPP. Botryococcene was not detected in any mutant assessments suggesting none of the mutations made in the BbSS background caused second-step reaction loss (+ SSL-3 assays) or product specificity interconversion (mutant only, + SSL-1 assays). Important to note here is the possibility that BbSS second-step reaction capability may have been lost in some mutants while maintaining first-step reaction activity, but because squalene synthases are known to retain the intermediate PSPP during catalysis,<sup>19</sup> any PSPP made by a mutant might not have been released for SSL-3 to convert to botryococcene.



**Figure 2.5** M1 series mutants. Conserved positions in domain II of BbSS were assessed for their role in catalytic activity by introducing reciprocal mutations to match the corresponding positions in SSL-3 (A). Each mutant was tested *in vitro* for the ability to perform the first-step reaction (+ SSL-3), the second-step reaction (+ SSL-1), or both steps of the reaction (mutant only). Results represent one independent experiment (B).

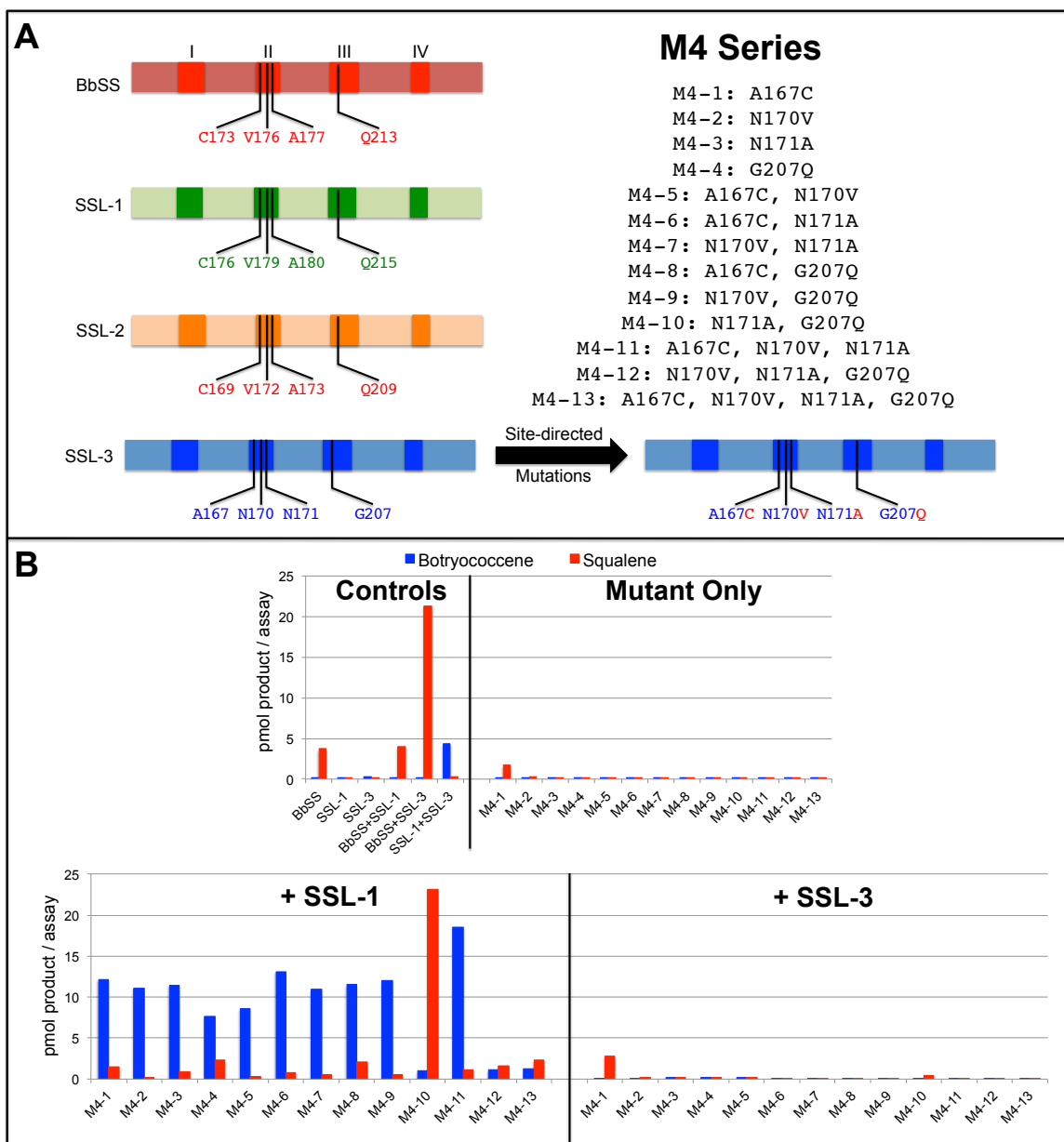
*M3 series mutants.* Based on results from M1 series mutants, residues occupying space in close proximity ( $\leq 5\text{\AA}$ ) to the cyclopropyl ring might affect catalytic activity as evidenced by the BbSS A177N mutant (M1-6). Thus, additional positions in BbSS and SSL-3 that occupy such spaces with residues that were drastically different between the two enzymes were sought out. Further analysis of sequence alignments and 3D models led to the identification of position Q213 in BbSS, which is highly conserved in domain III of all squalene synthases. In SSL-3, the corresponding position (207) is occupied by glycine. We hypothesized that positions A177 and Q213 of BbSS might together be responsible for product specificity between these two enzymes. Furthermore, positions further from where bond formations take place might favor slight changes in substrate orientation and thus alter product specificity. With this in mind, BbSS F188 and I295 were identified; the corresponding positions in SSL-3 were W182 and V288, respectively. The M3 mutant series was constructed to assess these residues' contribution to catalytic activity by mutating positions BbSS Q213, F188, and I295 to match the corresponding position in SSL-3. Single, double, and multiple mutations were made to determine if one, two or multiple changes were necessary for activity interconversion (**Figure 2.6**). Activity screening of the M3 series revealed position Q213 of BbSS to be critically important for both steps of the reaction for squalene biosynthesis. Mutation of this residue to glycine to match the position in SSL-3 resulted in considerable loss of activity (M3-7, mutant only). Even when PSPP was provided to this mutant by inclusion of SSL-1 in the assay (M3-7, + SSL-1), the apparent activity was still greatly diminished. Co-incubation of M3-7 with SSL-3 indicated that this mutant does not make PSPP or it is made but not released which would be necessary for SSL-3 to convert it to botryococcene. Mutation of BbSS Q213 to alanine (M3-6) and asparagine (M3-8) resulted in similar effects. The BbSS double mutant A177N, Q213G (M3-11) retained some second-step reaction activity but appeared completely devoid of the first. Contrary to the hypothesis that residues A177 and Q213 of BbSS control product specificity for squalene in BbSS, instead it seems that these residues play an important role in the ability to perform the first-step reaction. F188W and I295V mutations did not seem to alter product specificities either by themselves (M3-5 and M3-9) or in combination with other mutations (M3-15 and M3-16). M1-5, M1-6, and M1-7 mutants were simultaneously assayed in this experiment to allow for cross comparison between experiments. Although the activities measured in Figure 2.6 were approximately double the activities determined in Figure 2.5, the product profiles were identical, which attested to the reproducibility of this approach.



**Figure 2.6** M3 series mutants. Conserved positions in domains II, III, and IV of BbSS were assessed for their role in catalytic activity by introducing reciprocal mutations to match the corresponding positions in SSL-3 (A). Each mutant was tested *in vitro* for the ability to perform the first-step reaction (+ SSL-3), the second-step reaction (+ SSL-1), or both steps of the reaction (mutant only). Results represent one independent experiment (B).

*M4 series mutants.* Results from the M1 and M3 mutant series suggested BbSS positions A177 and Q213 play key roles in the catalytic activity for both steps of the reaction leading to squalene, especially the first-step reaction. Given this, we next asked what the effect would be if the corresponding positions in SSL-3 were interconverted to match BbSS. We hypothesized that interconverting the corresponding SSL-3 positions (N171, G207) to match those of BbSS (A177, Q213) would confer first-step reaction capability to SSL-3, in effect creating a bifunctional, botryococcene synthase. Additional residues investigated in SSL-3 include A167 and N170, which were mutated to match the corresponding positions in BbSS, C173 and V176, respectively. As before, a series of mutants were made to assess the effects of individual and multiple mutations on SSL-3 enzyme activity (**Figure 2.7**). Surprisingly, the double mutations N171A and G207Q in SSL-3 (M4-10) made to match the corresponding residues in BbSS (A177 and Q213) did not confer first-step reaction capability to SSL-3 but instead converted its product specificity from botryococcene to squalene (M4-10, + SSL-1). Neither single mutation (N171A or G207Q) in SSL-3 was by itself enough to introduce first-step reaction capability or convert product specificity from botryococcene to squalene. The A167C did not have an effect on SSL-3 by itself or when in combination with other mutations. N170V had differing effects on botryococcene activity – by itself, the mutation did not alter activity but in combination with N171A and G207Q, activity was abolished.

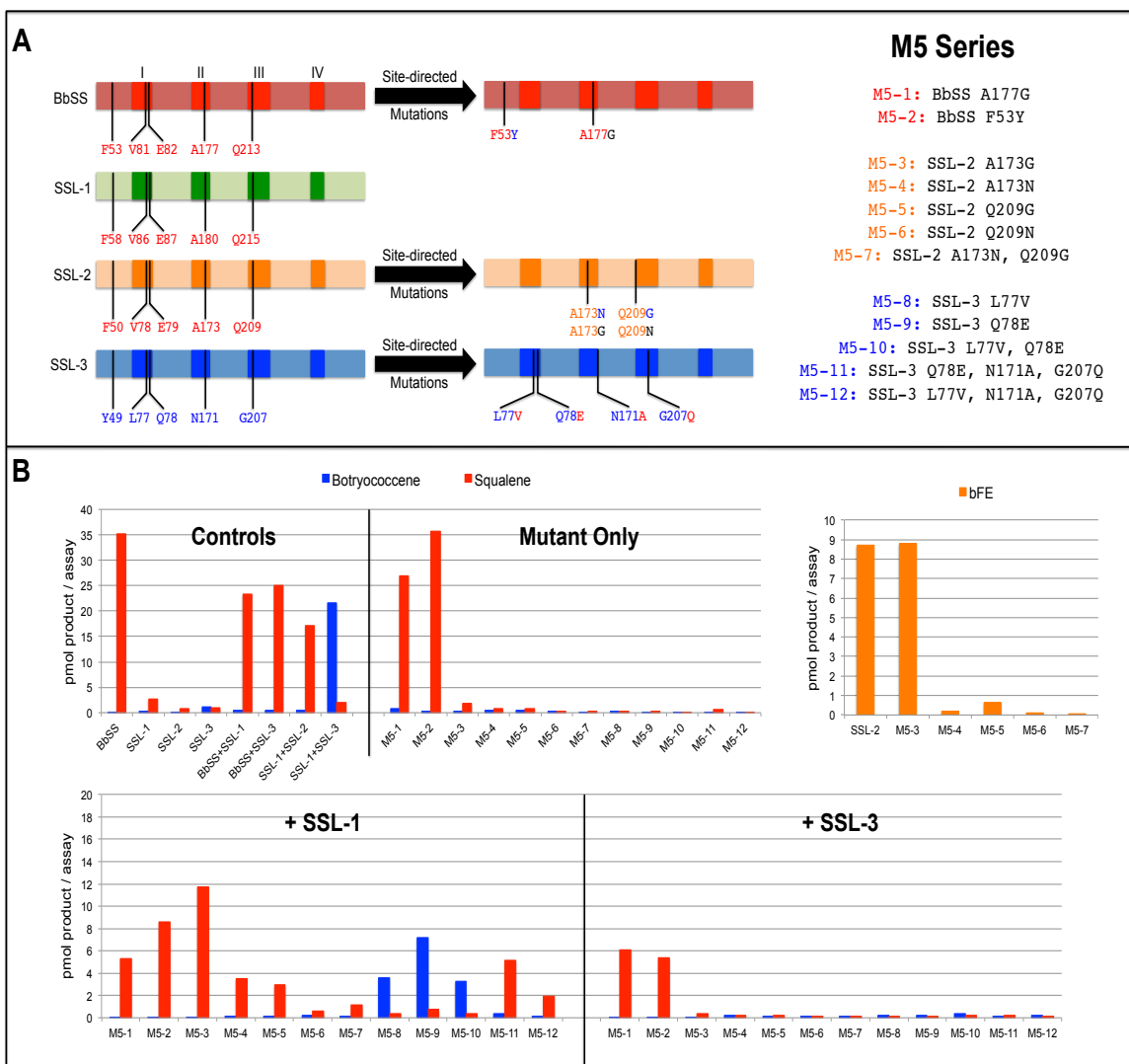




**Figure 2.7** M4 series mutants. Conserved positions in domains II and III of SSL-3 were assessed for their role in catalytic activity by introducing reciprocal mutations to match the corresponding positions in BbSS (A). Each mutant was tested *in vitro* for the ability to perform the first-step reaction (+ SSL-3), the second-step reaction (+ SSL-1), or both steps of the reaction (mutant only). Results represent one independent experiment (B).

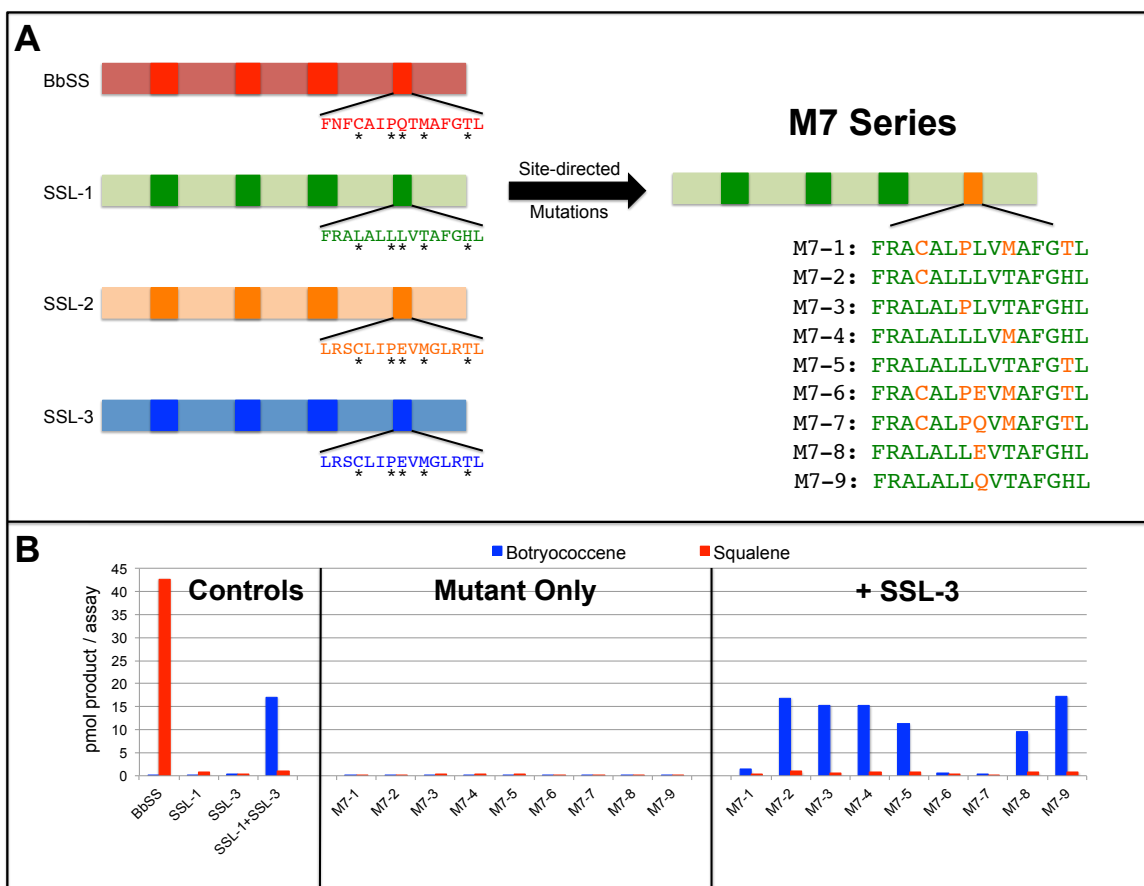
*M5 series mutants.* Because SSL-2 produces squalene when supplied with PSPP, the contribution of the two positions in SSL-2 (A173 and Q209) that correspond to BbSS A177, Q213 and SSL-3 N171, G207 were assessed. To understand the role of positions A173 and Q209 in SSL-2 enzyme activity, mutations were introduced to match the corresponding positions in SSL-3 followed by enzyme activity measurements for squalene/botryococcene. Mutations with similar amino acids were also introduced in SSL-2 including A173G and Q209N. Additionally, SSL-2 can solely catalyze formation of bis-farnesyl ether (bFE) from two molecules of FPP, so this enzyme activity was also tested for the mutants. To follow up on the observation that the BbSS A177N mutant loses first-step reaction capability, an A177G mutation was made to verify the role of this position in proper shaping of the active site topography. Also, a F53Y mutant was made in BbSS because this position is present in the highly conserved 'Flap Region,' and the corresponding position in SSL-3 is occupied by tyrosine. Finally, a few SSL-3 mutants were made to examine key difference between it and the S1 DxxED site of squalene synthases; these mutations include L77V and Q78E. As for previous mutant series, individual mutations were examined by themselves and in combination with others. **(Figure 2.8)**

BbSS mutants A177G (M5-1) and F53Y (M5-2) did not appear to greatly effect activity for squalene biosynthesis; the decrease in activity observed for + SSL-1 and + SSL-3 assays might be due to the additional SSL-1, SSL-3, or milieu of proteins in the crude lysate 'soaking up' substrate thereby making it unavailable for the mutant enzyme. Alternatively, the mutant enzymes present in the crude bacterial lysate could have lost activity over the course of the experiment. SSL-2 A173G (M5-3) did not greatly affect squalene formation, as was the case for BbSS A177G (M5-1). SSL-2 mutations A173N, Q209G, and Q209N did appear to have a detrimental effect on this enzyme's catalysis for squalene, similar to mutations of the corresponding sites within the BbSS enzyme. bis-farnesyl ether activity of SSL-2 was also affected by these mutations. The SSL-2 A173N, Q209G (M5-7) double mutant also compromised activities for both squalene and bFE formation. SSL-3 mutations L77V and Q78E did not appear to significantly alter activity when assessed individually (M5-8 and M5-9), as a double mutant (M5-10), or in combination with N171A and G207Q mutations (M5-11 and M5-12).



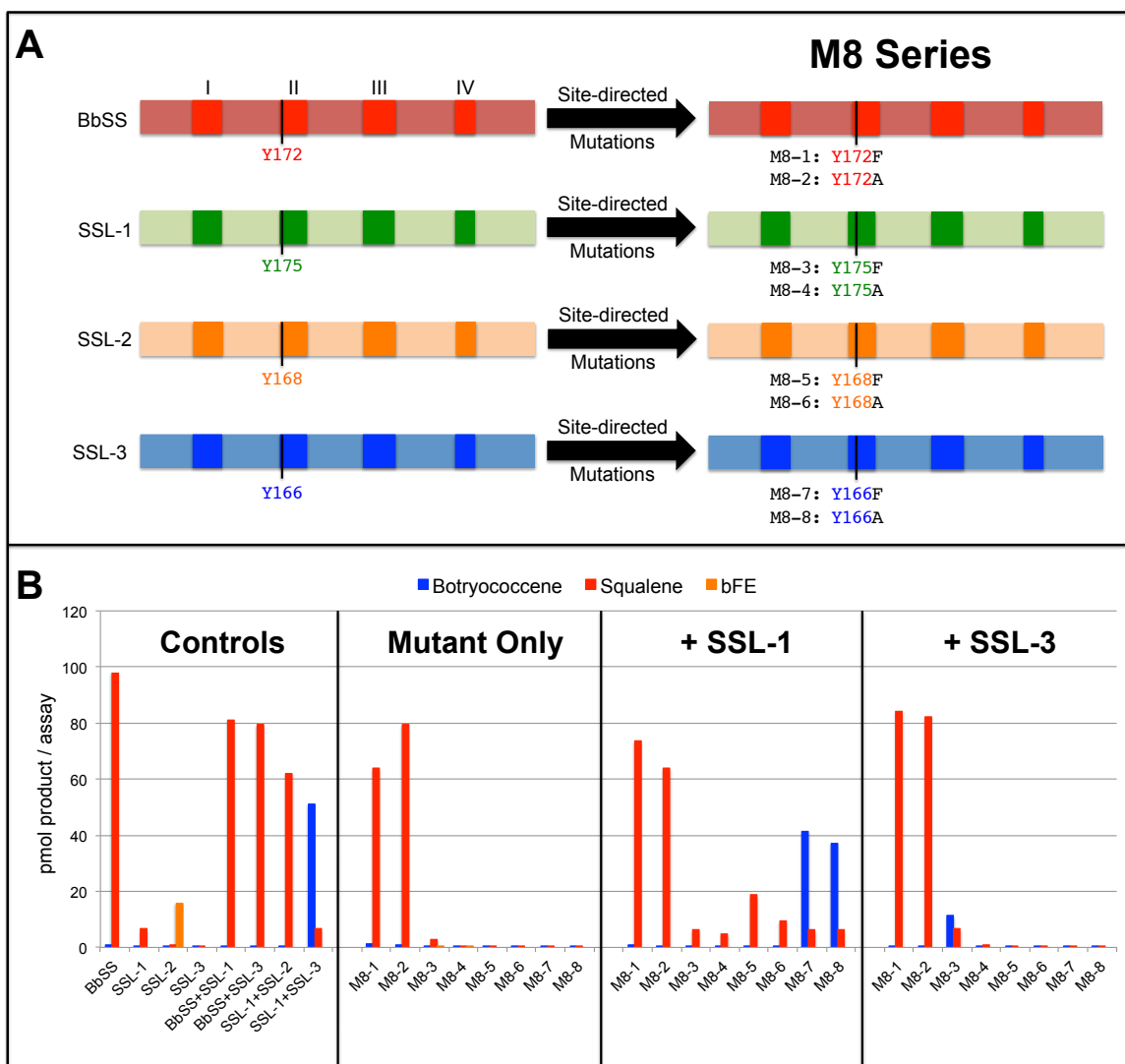
**Figure 2.8** M5 series mutants. Conserved positions in BbSS, SSL-2, and SSL-3 were assessed for their role in catalytic activity by introducing reciprocal mutations to match corresponding positions in the other *B. braunii* triterpene synthases (A). Each mutant was tested *in vitro* for the ability to perform the first-step reaction (+ SSL-3), the second-step reaction (+ SSL-1), or both steps of the reaction (mutant only). Results represent one independent experiment (B).

*M7 series mutants.* SSL-1 can perform only the first-step reaction in botryococcene and squalene formation – the biosynthesis of PSPP. Several reasons for this outcome derive from the mechanism for both steps of the reaction. These include 1) release of PSPP before diphosphate ionization can occur, 2) inability to ionize PSPP due to lack of the appropriate residues needed for Mg<sup>2+</sup> binding/orientation, 3) inability to ionize PSPP due to lack of the appropriate residues in the active site, 4) inability to rearrange the cyclopropyl intermediate after PSPP ionization, 5) inability to bind NADPH, or 6) some combination of these possibilities. From sequence alignments of the *B. braunii* triterpene synthases, it was observed that four positions in domain IV are conserved among BbSS, SSL-2, and SSL-3. Because this domain seems to be specific to squalene synthases, it has been postulated as being important for completion of the second-step reaction.<sup>29,30</sup> SSL-1 cannot perform the second-step reaction and displays considerable sequence divergence in these corresponding positions of domain IV. To address the possibility that these changes in SSL-1 domain IV might be contributing to its inability to carry out the second-step reaction, they were mutated to match the consensus sequence of BbSS, SSL-2, and SSL-3 in this region. A fifth residue was also mutated which is a glutamine in squalene synthases and glutamate in SSL-2 and SSL-3. Six SSL-1 single mutants were made at each of the five positions, and three other mutants that contained either four or five of the mutations were also made – these mutants comprise the M7 series (**Figure 2.9**). Mutant enzymes that were assayed containing only one amino acid substitutions (M7-2, M7-3, M7-4, M7-5, M7-8, M7-9) did not display drastic alterations in catalytic ability in the coupled assays with SSL-3 where botryococcene was measured. Coupled assays for mutants containing four (M7-1, + SSL-3) or five (M7-6 and M7-7, + SSL-3) of the substitutions did not produce significant levels of botryococcene, which suggested that these multiple SSL-1 mutations negatively affected SSL-1, either in its ability to properly fold, catalyze PSPP formation, or some combination of both.



**Figure 2.9** M7 series mutants. Divergent positions in domain IV of SSL-1 were assessed for their role in catalytic activity by introducing reciprocal mutations to match the corresponding positions in BbSS, SSL-2, and SSL-3 (A). Each mutant was tested *in vitro* for the ability to perform the first-step reaction (+ SSL-3) or both steps of the reaction (mutant only). Results represent one independent experiment (B).

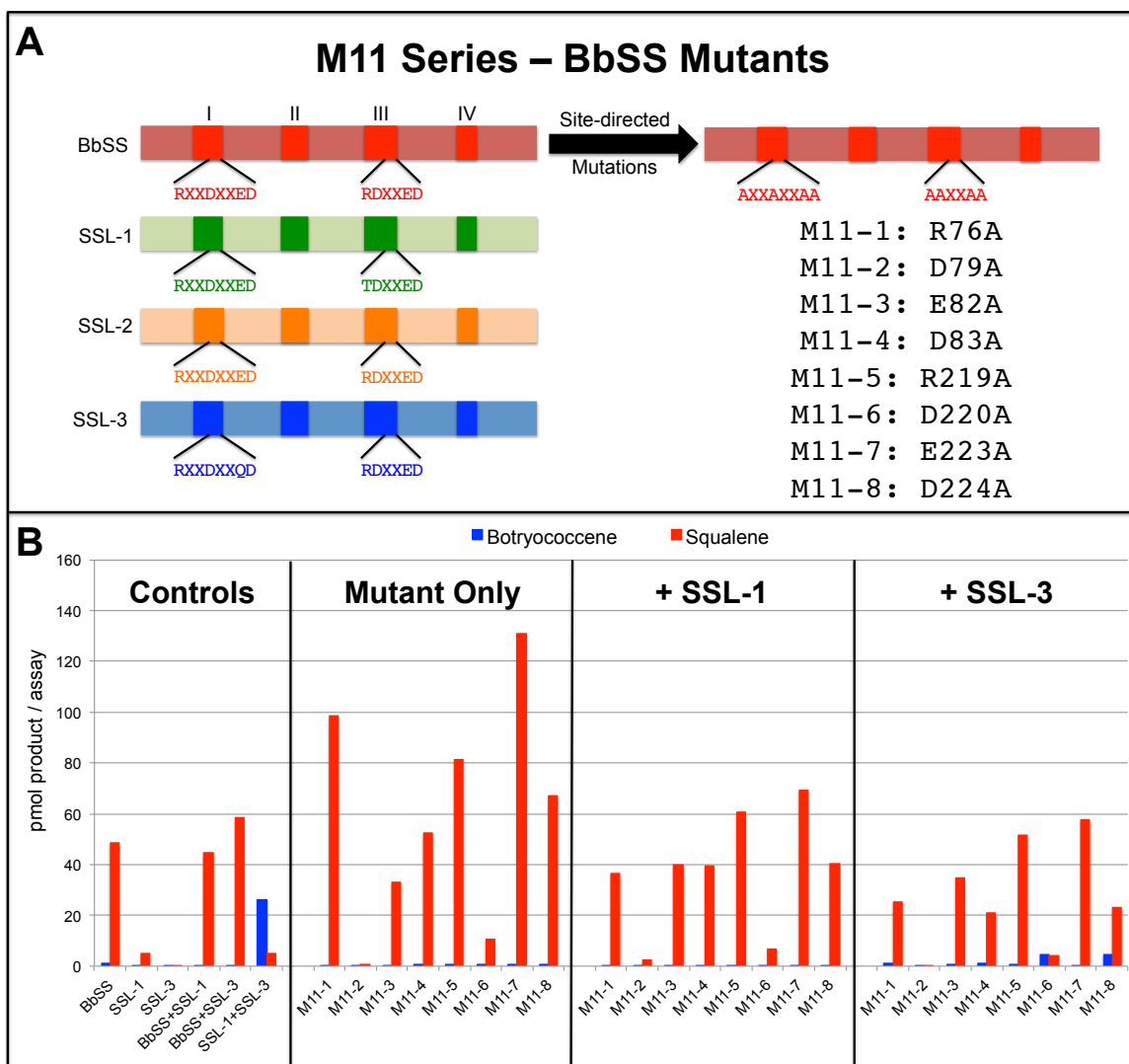
*M8 series mutants.* According to a structure-function study of rat squalene synthase performed by Schechter *et al.* (1998), Y171S, Y171F, and Y171W mutations resulted in the loss of both first- and second-step reaction activities, which suggested that this residue, specifically the tyrosine hydroxyl, was critically involved in catalytic activity.<sup>31</sup> To assess the contribution of this highly conserved position to squalene and botryococcene catalysis in the *B. braunii* triterpene synthases, it was mutated to alanine and phenylalanine in BbSS, SSL-1, SSL-2, and SSL-3 (**Figure 2.10**). Mutants were assessed as previously described where the first, second and both steps of the reaction were tested. Contrary to results published by Schechter *et al.* (1998), it does not appear that the highly conserved Tyr residue plays such a critical role in BbSS and SSL-3 whereas SSL-1 and SSL-2 are affected considerably. All mutants do however seem to be more effected by the tyrosine to alanine mutation.



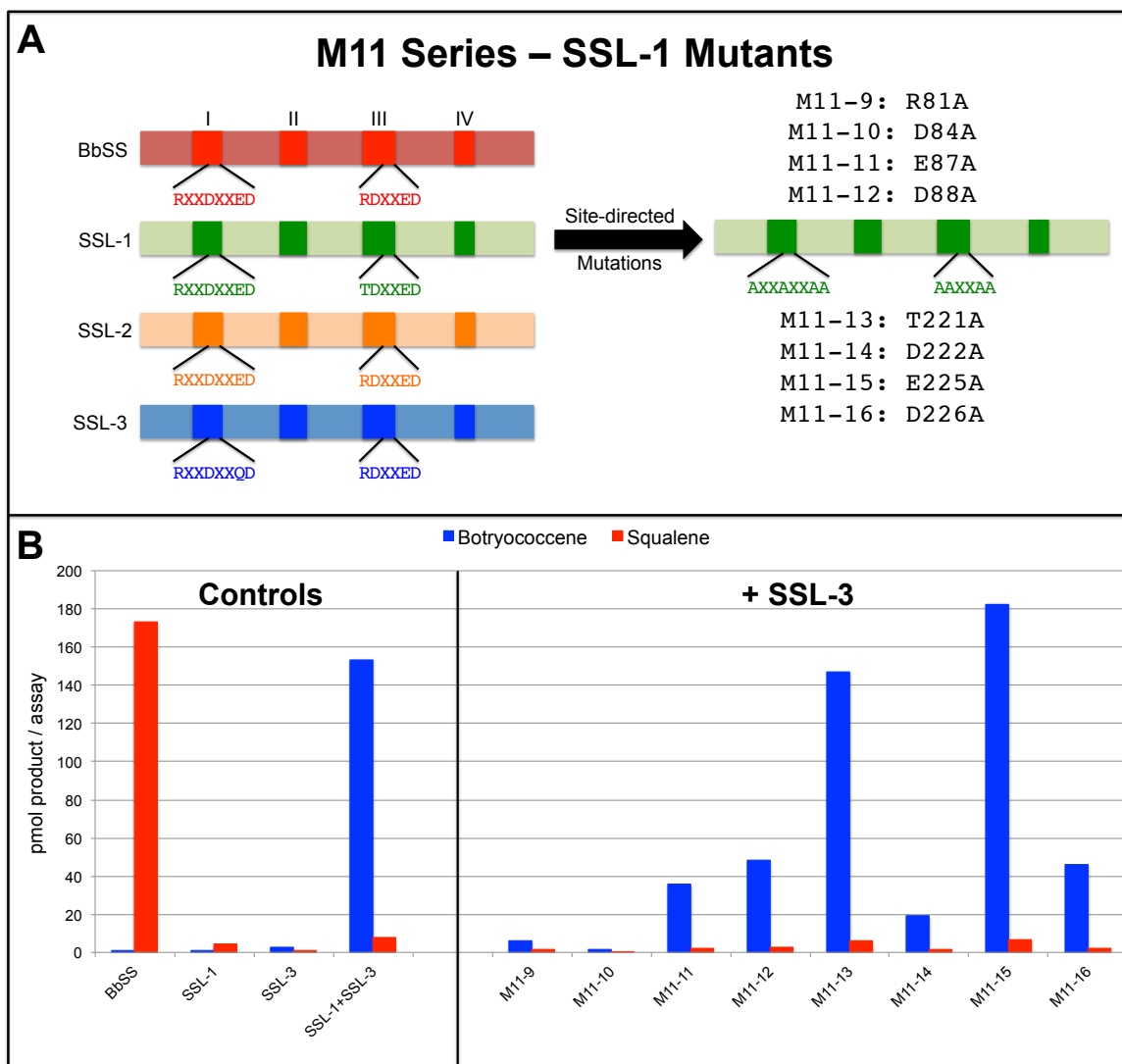
**Figure 2.10** M8 series mutants. Assessment of a highly conserved tyrosine residue in domain II of BbSS, SSL-1, SSL-2, and SSL-3 for its role in catalytic activity by mutating it to phenylalanine and alanine (A). Each mutant was tested *in vitro* for the ability to perform the first-step reaction (+ SSL-3), the second-step reaction (+ SSL-1), or both steps of the reaction (mutant only). Results represent one independent experiment (B).

*M11 series mutants.* A hallmark feature of class I terpene synthase enzymes are the aspartate/glutamate-rich domain(s) thought to be responsible for coordination and ionization of the diphosphate moiety of the prenyl substrate(s). In squalene synthases, these are designated as the S1 and S2 sites depending on whether they are proximal to the N- or C-terminus, respectively;<sup>25,65</sup> the consensus sequence for the S1 and S2 sites are RxxDxxED and RDxxED, respectively. The conserved arginine residues included in the consensus sequences have been implicated in diphosphate anion removal from active site to prevent bond reformation with the primary carbocation while the conserved aspartate/glutamate residues appeared to be important for organizing the overall architecture of the Mg<sup>2+</sup>-diphosphate complexes near the active site entry.<sup>25,26,29</sup> Lin *et al.* (2011) suggested that the S1 site in CrtM is the location where ionization occurs for FPP and PSPP in the first- and second-step reactions.<sup>26</sup> To assess the possibility that the S1 or S2 site function independent or together as the site of ionization in the *B. braunii* triterpene synthases, each position of the consensus in BbSS (**Figure 2.11**), SSL-1 (**Figure 2.12**), and SSL-3 (**Figure 2.14**) was individually mutated to alanine then tested for both reaction-step capabilities. BbSS single mutations D79A (M11-2) and D220A (M11-6) in the S1 and S2 sites, respectively, resulted in an apparent complete loss of activity for both reaction steps. Squalene was not detected for M11-2 or M11-6 in mutant only assessments or when co-incubated with SSL-1 nor was botryococcene detected in co-incubations with SSL-3. The other alanine single mutations had less drastic effects, possibly even increasing activity for BbSS E223A (M11-7). Effects of the single alanine mutations in SSL-1 were similar to BbSS; D84A (M11-10) and D222A (M11-14) resulted in an apparent inability to catalyze PSPP formation based on co-incubation with SSL-3. Additionally, the R81A mutation in SSL-1 also had near complete loss of activity, an outcome that was not observed for BbSS. Other single alanine mutations appeared to reduce activity in the SSL-1 mutants. SSL-3 was affected similarly to BbSS. The D75A (M11-18) and D213A (M11-22) mutants were nearly inactive; the other single alanine mutants were less affected.

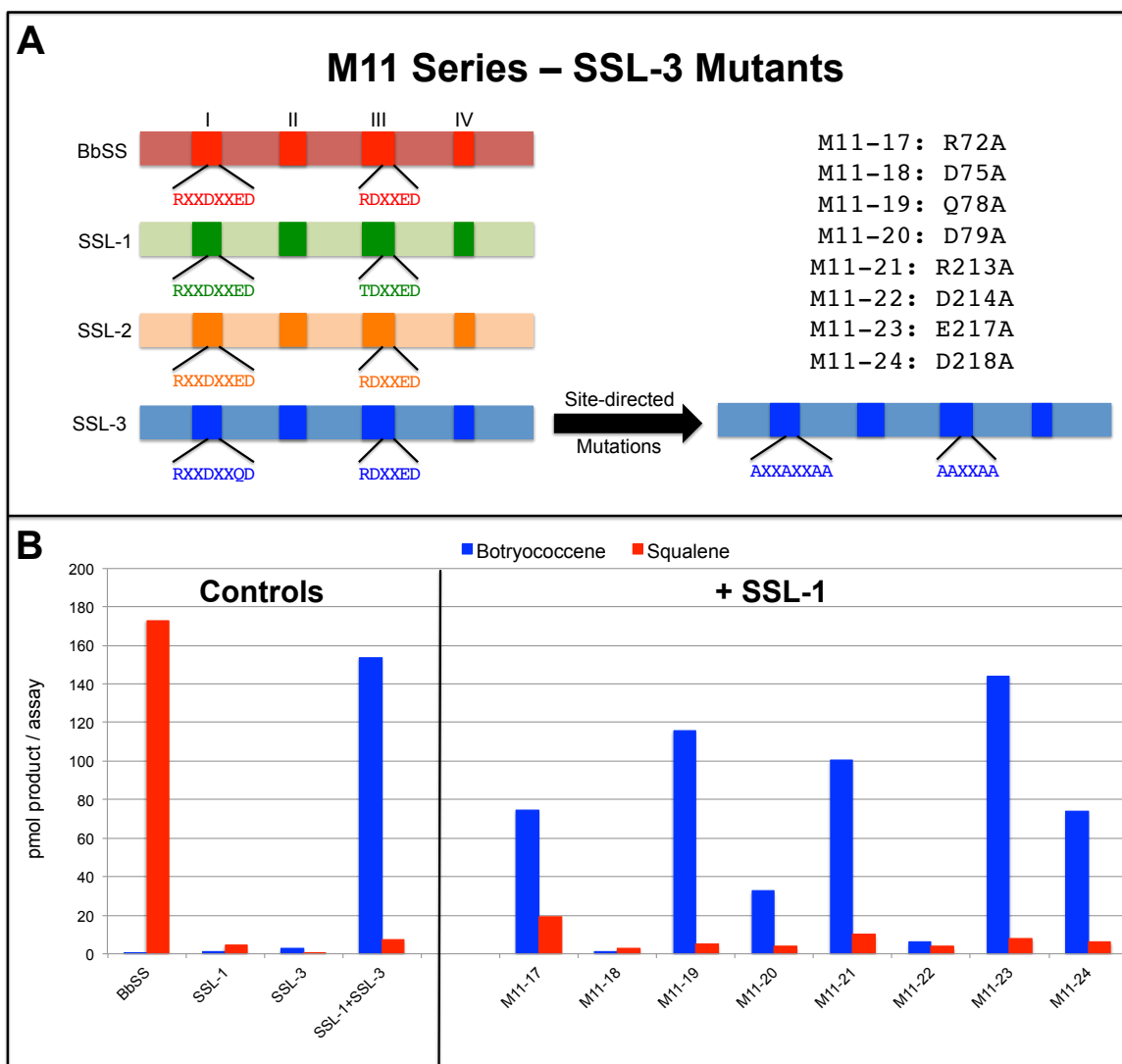




**Figure 2.11** BbSS M11 series mutants. Conserved positions in BbSS aspartate-rich motifs were assessed for their role in catalytic activity by introducing single alanine mutations (A). Each mutant was tested *in vitro* for the ability to perform the first-step reaction (+ SSL-3), the second-step reaction (+ SSL-1), or both steps of the reaction (mutant only). Results represent one independent experiment (B).

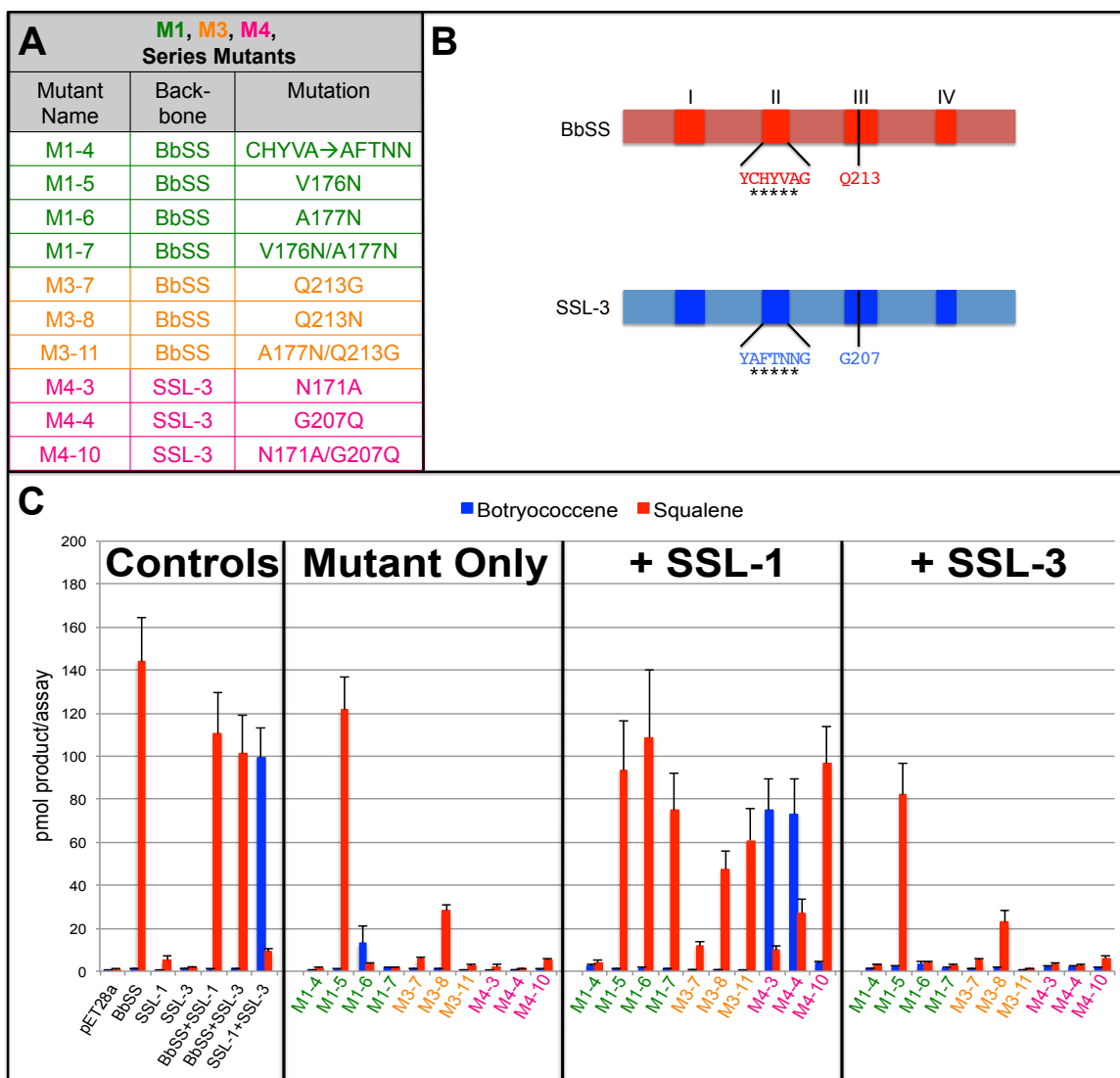


**Figure 2.12** SSL-1 M11 series mutants. Conserved positions in SSL-1 aspartate-rich motifs were assessed for their role in catalytic activity by introducing single alanine mutations (A). Each mutant was tested *in vitro* for the ability to perform the first-step reaction (+ SSL-3). Results represent one independent experiment (B).

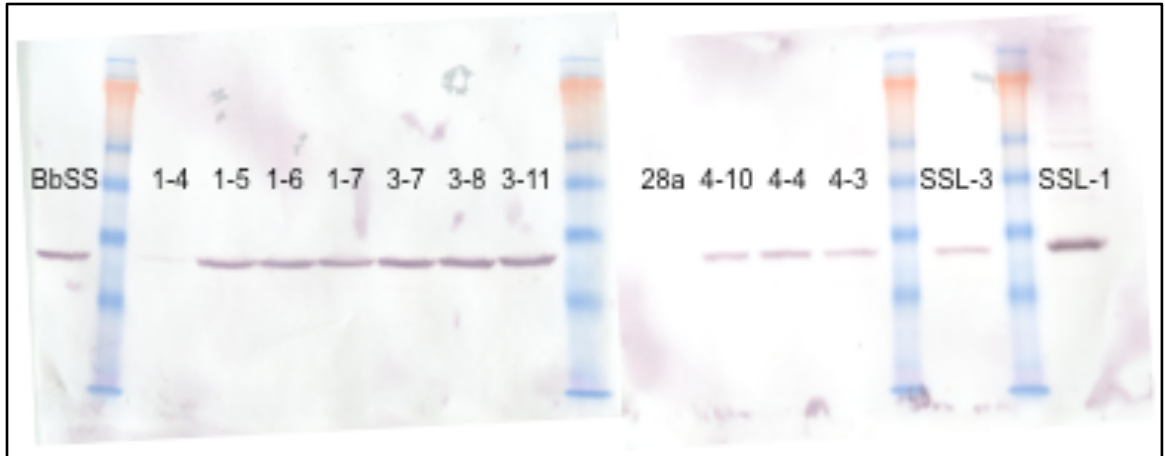


**Figure 2.13** SSL-3 M11 series mutants. Conserved positions in SSL-3 aspartate-rich motifs were assessed for their role in catalytic activity by introducing single alanine mutations (A). Each mutant was tested *in vitro* for the ability to perform the second-step reaction (+ SSL-1). Results represent one independent experiment (B).

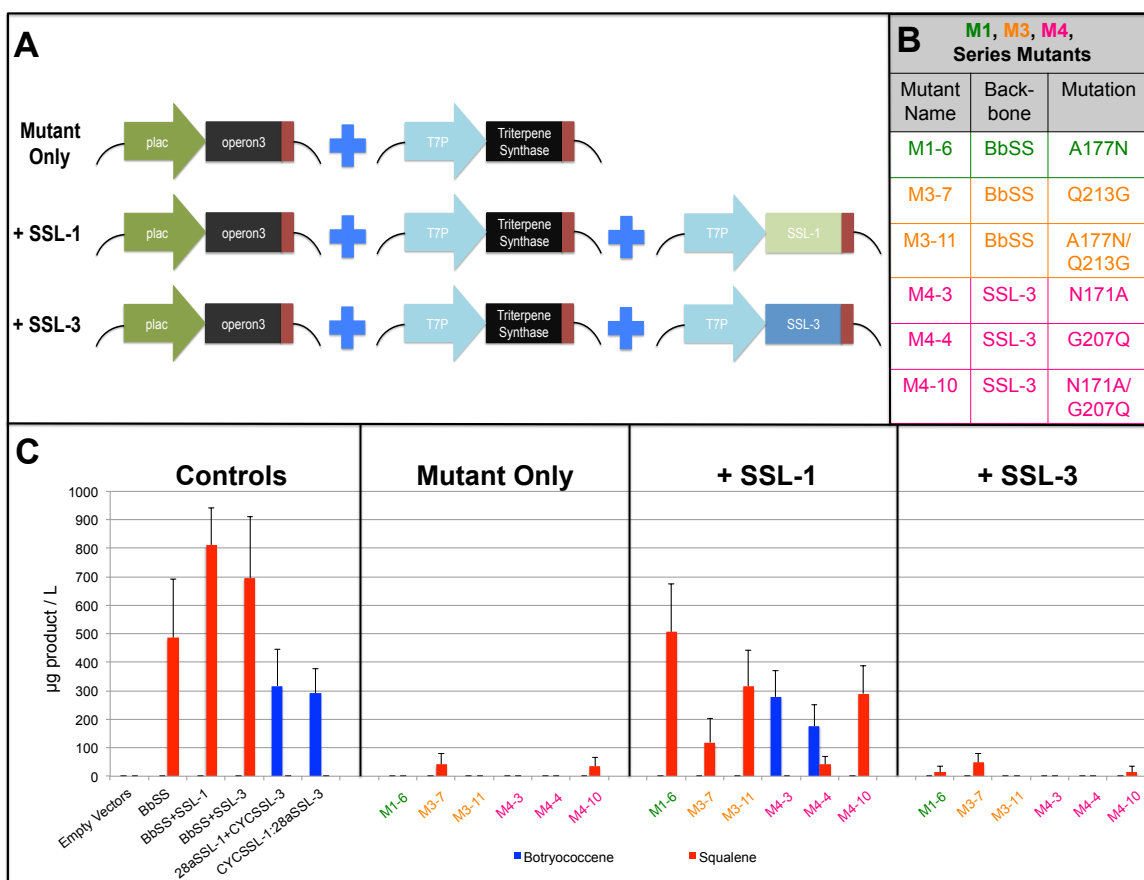
*Additional verification for mutants of interest from previous mutant series.* Because the previously described work was only an initial screen to check enzymatic activities for respective mutants, further testing was needed with the appropriate replications for *in vitro* assays to support these initial observations. **Figure 2.14A** lists the M1, M3, and M4 series mutants that were chosen for subsequent *in vitro* analyses with their corresponding mutations and the enzyme backbone from which they were constructed. *In vitro* assays using crude lysate were performed in replicate for these mutants as previously described; protein levels in crude lysate total protein were assessed *via* Western blotting (**Figure 2.15**). Activity levels for the mutant enzymes, although somewhat variable, were repeatable and consistent with previously made measurements. Protein expression levels were relatively similar for mutants compared to wild type. In addition to the *in vitro* assays, the bacterial line expressing operon 3 to produce additional FPP was used to chemically profile the M1-6, M3-7, M3-11, M4-3, M4-4, and M4-10 mutants to obtain an *in vivo* enzyme assessment. *In vivo* analyses of mutant enzymes were conducted by expressing mutants only or co-expressing mutants with either SSL-1 or SSL-3 such that both steps of the reaction could be assessed, much the same way they were for *in vitro* assays (**Figure 2.16**). These analyses recapitulated observations made for *in vitro* activity screening where residues A177 and Q213 of BbSS appear to play roles in both first- and second-step reactions without affecting product specificity whereas N171 and G207 appear responsible for squalene and botryococcene product specificity in SSL-3. Further assessment of the M8 and M11 series mutants was also carried out in replicate in the bacterial line that produces additional FPP to verify the mutant enzyme activities (**Figure 2.17**).



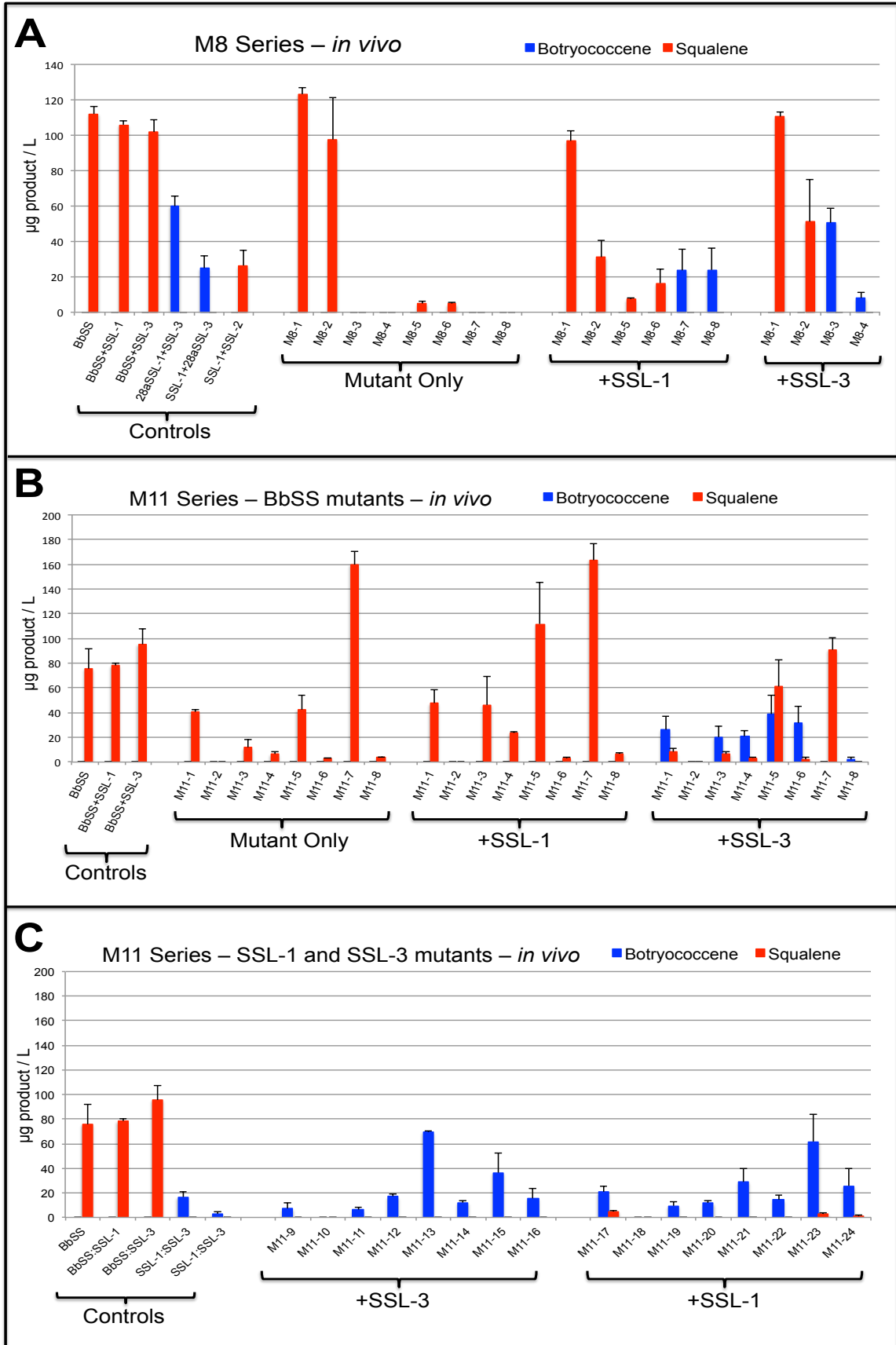
**Figure 2.14** M1, M3, and M4 series mutants follow-up *in vitro* assays. M1, M3, and M4 series mutants used for this analysis are listed along with the backbone from which they are derived and their respective mutation(s) (A). Generic depiction of BbSS and SSL-3 enzymes showing where the positions mutated reside in relation to the conserved domains (B). Crude extracts of bacterially expressed enzymes were used as described in **Figure 2.3** to assay for the given enzyme's ability to perform the first, second, or both steps of squalene and botryococcene biosynthesis (C). Results are representative of three independent replicates (two technical replicates each; (n=3).



**Figure 2.15** M1, M3, and M4 series mutants Western blots. Anti-His Western blot assessment of N-terminal hexahistidine-tagged proteins in bacterial crude lysate used for *in vitro* activity measurement in **Figure 2.14C**. Fifteen  $\mu\text{L}$  of each crude extract was loaded per lane and resolved on a 9.5% acrylamide gel followed by transfer to nitrocellulose and blotting with monoclonal Anti-His-alkaline phosphatase conjugate antibody.



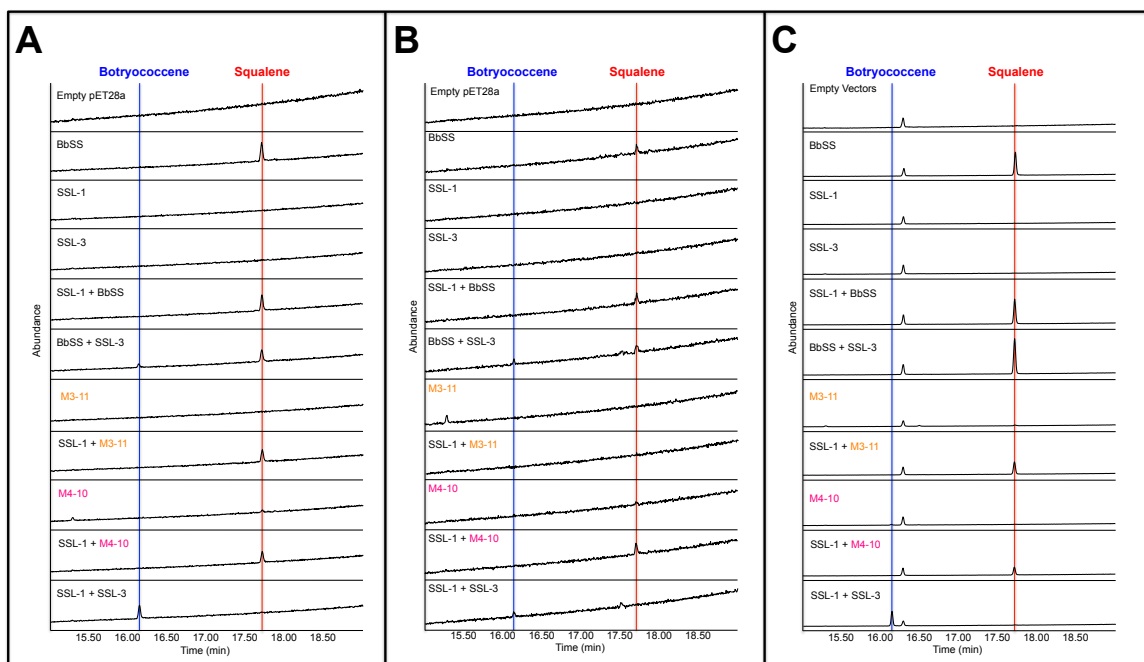
**Figure 2.16** M1, M3, and M4 series mutants follow-up *in vivo* assays. *In vivo* chemical profiling of BbSS and SSL-3 mutants expressed in *E. coli* BL21 (DE3) harboring pSB1:operon3. Depiction of general scheme used for *in vivo* analysis of a mutant triterpene synthase for the ability to catalyze the first (+ SSL-3), second (+ SSL-1), or both (mutant only) reaction steps of squalene or botryococcene biosynthesis. Mutant enzymes were expressed from the pET28a vector while co-expressed SSL-1 or SSL-3 was expressed from pACYCDuet-1 (A). M1, M3, and M4 series mutants used for this analysis are listed along with the backbone from which they are derived and their respective mutation(s) (B). Hexane extractable products were quantified after GC-MS product profiling of the wild-type and mutant enzymes. Results are representative of three independent experiments with two biological replicates for each (n=3) (C).





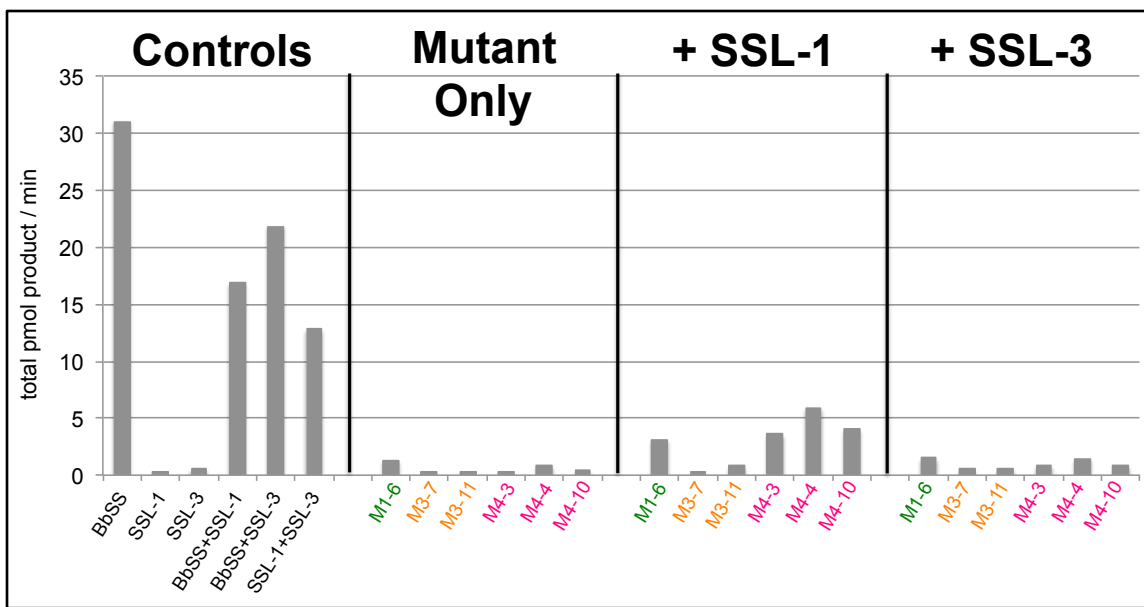
**Figure 2.17** M8 and M11 series mutants *in vivo* chemical profiling to assess contribution of highly conserved residues to first, second, and both steps of squalene and botryococcene biosynthesis. Chemical profiling of the M8 mutant series where the highly conserved Tyr residue in domain II was mutated to Phe or Ala in BbSS, SSL-1, SSL-2, and SSL-3 (A). Chemical profiling for the BbSS M11 series mutants where each conserved position of the S1 and S2 DxxED motifs were individually mutated to Ala (B). Chemical profiling for the SSL-1 and SSL-3 M11 series mutants where each conserved position of the S1 and S2 DxxED motifs were individually mutated to Ala (C). Data shown is representative of three independent replicates (n=3).

*GC-MS verification of in vitro and in vivo enzyme products.* GC-MS product profiling verified products from *in vitro* and *in vivo* analyses as squalene or botryococcene. Authentic standards of squalene and botryococcene were compared to the products from *in vitro* and *in vivo* assays in terms of retention time (**Figure 2.18**) and mass spectral fragmentation pattern (data not shown). Additionally, NADPH dependence of the mutant triterpene synthases was examined to verify this had not been altered due to the introduced mutation(s). These experiments demonstrated not only that products observed from *in vitro* and *in vivo* analyses were in fact squalene and botryococcene, but also the dependence of these enzymes on NADPH for the second-step reaction (**Figure 2.18**). Minor activities were observed in minus NADPH samples for *in vitro* reactions, perhaps due to NADPH present in the crude lysate used as the source of enzyme (**Figure 2.18B**).

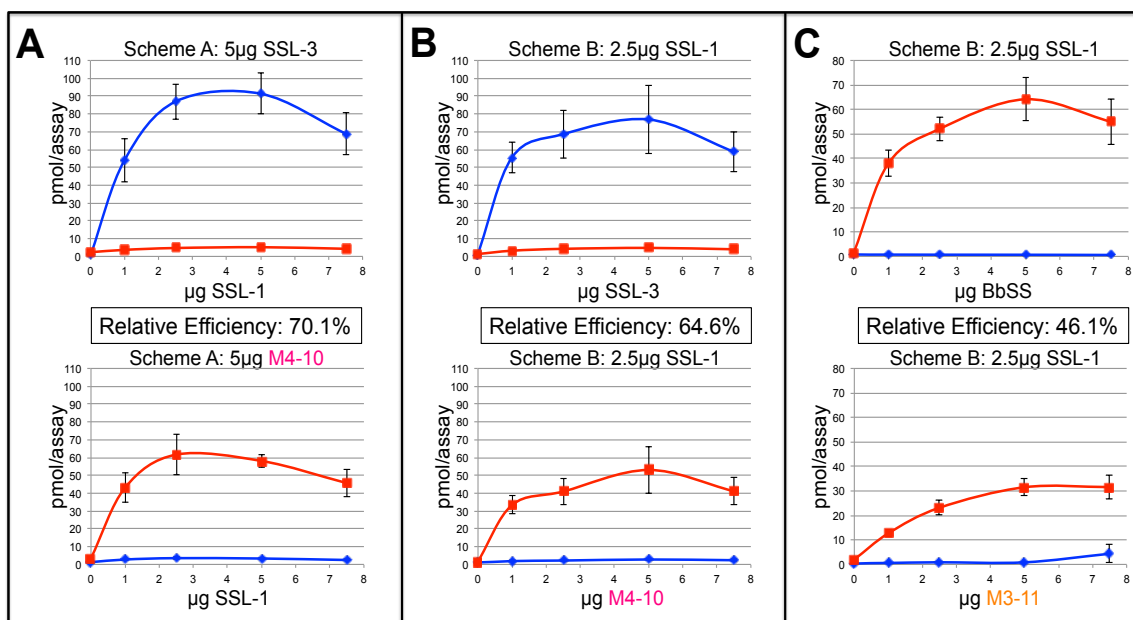


**Figure 2.18** GC-MS analysis of wild-type and mutant triterpene synthase product profiles. *In vitro* assays conducted using bacterial crude lysates with 50  $\mu$ M FPP in the presence (A) and absence (B) of 5 mM NADPH. *In vivo* analysis of triterpene synthase(s) co-expressed in bacteria as described in **Figure 2.16** (C). Chemical profiles represent hexane extractable products for the *in vitro* and *in vivo* analyses.

*Crude catalytic efficiency assessment of mutants relative to wild type.* Ideally, assessing enzyme catalytic efficiency involves purifying enzymes to homogeneity followed by a substrate-dependent analysis using Michaelis-Menten models. As stated earlier, PSPP is not a readily available substrate. Further complicating this effort is the loss of activity observed when enzymes are assayed after purification (**Figure 2.19**). This loss of activity was observed after multiple independent enzyme purifications. Due to these complications, a strategy from Greenhagen *et al.* (2003) was exploited whereby varying enzyme amounts in a coupled assay were used to glean information about the catalytic efficiency of one of the coupled enzymes in the mixture.<sup>66</sup> These authors faced similar obstacles in terms of substrate availability and rationalized this method to assess various P450 enzymatic activities with a given sesquiterpene scaffold. Using this approach, we estimated how efficiently M4-10 could catalyze squalene formation relative to botryococcene by SSL-3 *via* two different reaction schemes, A and B. Scheme A used increasing amounts of SSL-1 crude lysate (0, 1, 2.5, 5, and 7.5µg) with the intention of providing increasing amounts of PSPP to achieve a 'pseudo-kinetic' assessment of SSL-3 and M4-10; the second enzyme amount was held constant at 5 µg. Scheme A assessment of SSL-3 and M4-10 led to a relative efficiency of 70.1% product formation by the mutant M4-10 enzyme (**Figure 2.20A**). Relative efficiency was calculated as a ratio of product formed by mutant enzyme to product formed by wild-type enzyme for each corresponding data point then expressed as a percentage. Overall relative efficiency was obtained by averaging the relative efficiencies for each data point. Scheme B was conducted similarly to Scheme A with the difference that the SSL-1 amount was held constant in the first reaction while the amount of SSL-3, M4-10, BbSS, or M3-11 was varied (0, 1, 2.5, 5, and 7.5µg) in the second reaction. This revealed a relative efficiency of 64.6% for SSL-3 and M4-10 (**Figure 2.20B**). BbSS and M3-11 relative efficiency assessed using Scheme B was calculated as 46.1% (**Figure 2.20C**). Relative efficiencies for Scheme B assays were calculated identically to those for Scheme A.



**Figure 2.19** Purified wild-type and mutant triterpene synthase quality control assays. Simple assay to check activities of purified enzymes for botryococcene and squalene. Enzymes were incubated with 30  $\mu$ M FPP (S.A.  $\approx$  1000DPM/pmol) and 5 mM NADPH for 10 minutes followed by extraction with hexane from which an aliquot was used directly for scintillation spectroscopy. Results represent one replicate



**Figure 2.20** Relative efficiency assays for squalene (red line) and botryococcene (blue line) formation by M3-11 and M4-10 mutants. Scheme A – measurement of botryococcene and squalene activity with increasing amounts (0 – 7.5 µg) of SSL-1 crude lysate in first reaction and a constant amount (5 µg) of SSL-3 (upper) or M4-10 (lower) crude lysate in second reaction (A). Scheme B – botryococcene and squalene activity measurements with a constant amount (2.5 µg) of SSL-1 crude lysate in first reaction and increasing amounts (0 – 7.5 µg) of SSL-3 (upper) or M4-10 (lower) crude lysate in second reaction (B). Scheme B reactions for BbSS (upper) or M3-11 (lower) to measure relative efficiency of squalene and botryococcene catalysis (C). All assays contained 5 mM NADPH. Results are representative of six independent replicates (n=6) except for assays with 7.5 µg crude lysate in the second reaction (n=4).

## Discussion

Unraveling the intricate structure-function relationships among the head-to-head coupling class of triterpene synthases is a grand challenge. Structure-function maps for squalene synthase were first suggested by Robinson *et al.* (1993) using protein sequence alignments with other known squalene and phytoene synthases<sup>29</sup> that have been largely corroborated by crystallographic studies of HsSS and CrtM.<sup>26,30,57</sup> The structures are quite similar overall, with an all-alpha helical fold in which domains I-IV comprise the active site pocket. Notable features include the aspartate-rich (DxxED) motifs responsible for coordination of the prenyl diphosphate-Mg<sup>2+</sup>(Mn<sup>2+</sup>) functionality of the two identical substrates. Additionally, hydrophobic residues predominate in the rear of the pocket presumably for coordinating the prenyl chains through van der Waals interactions with a gradual shift to more polar/charged amino acids near the entry of the active site.

Structure-function studies conducted using site-directed mutants of *Rattus norvegicus* squalene synthase (RnSS) suggested a few residues that are important for the second (F288) or both (Y171) reaction steps based on negative phenotypes whereby the first, second or both reaction steps were decreased or eliminated.<sup>31</sup> We performed an analysis of the corresponding position in the *B. braunii* triterpene synthases where the native tyrosine residue was mutated to phenylalanine and alanine (**Figure 2.10** and **Figure 2.17A**). These studies suggested varying contributions of this residue to squalene or botryococcene catalysis in BbSS, SSL-1, SSL-2, and SSL-3. Importantly, the BbSS Y172F mutant retained significant activity with a slight decrease in catalytic activity for the Y172A mutant. These results suggest that the tyrosine hydroxyl does not play a direct role in catalysis. Instead, it may be contributing carbocation stabilizing effects from *pi* orbitals in its ring structure. Another puzzling observation regarding structure-function relationships in squalene synthase was made in *Arabidopsis thaliana* which contains two copies of squalene synthase, one that is active (AtSS1) and one that is not (AtSS2) despite 79% sequence identity between the two.<sup>67</sup> Interestingly, AtSS2 possesses a Ser in the position that corresponds to F288 in RnSS. This does not seem to explain the lack of squalene activity however, because an AtSS2 S287F mutant that was constructed and assayed showed no squalene activity. Moreover, when F287 was mutated to Ser in AtSS1, activity appears to be diminished but not completely lost indicating that other residues in AtSS2 are contributing to the loss of squalene activity.<sup>67</sup>

Our interconversion approach to study product specificity in BbSS and SSL-3 has yielded new insight into the amino acids controlling product specificity in the 1'-1 and 1'-3 linkage class of triterpene synthases in *B. braunii*. Residues N171 and G207 of SSL-3 contribute dramatically to botryococcene/squalene second-step reaction product specificity in this enzyme. When interconverted to match the corresponding positions in BbSS, the SSL-3 mutant makes squalene 70.1% as efficiently as wild-type SSL-3 makes botryococcene. The corresponding positions in BbSS, A177 and Q213, do not appear to be entirely

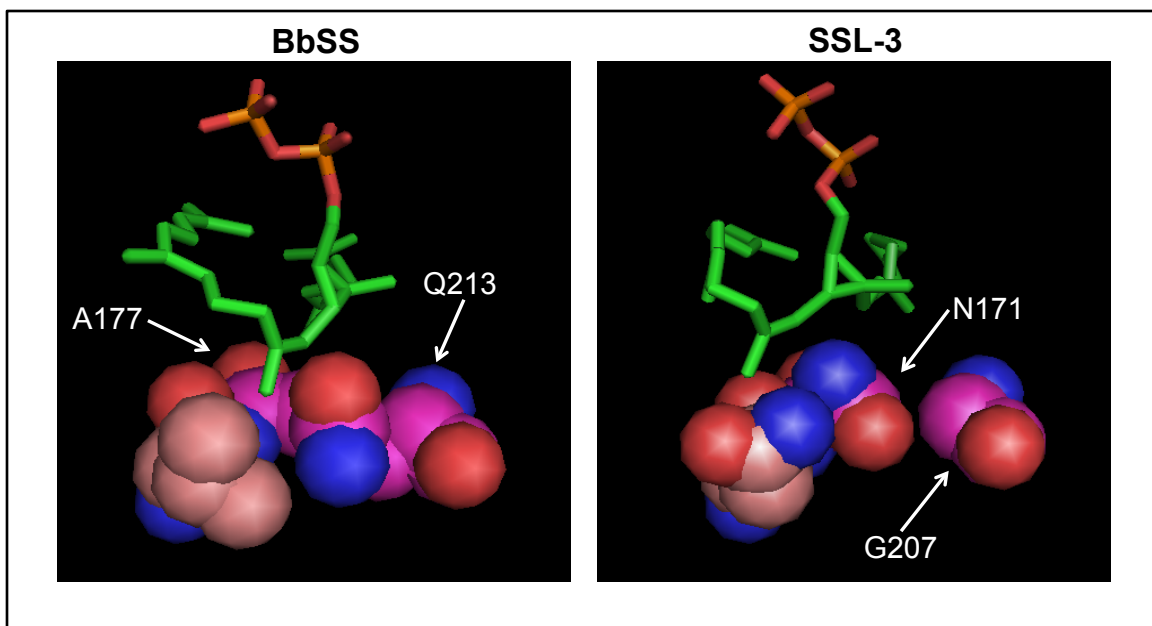
responsible for this enzyme's product specificity. Instead, A177 appears to contribute to first-step reaction capability (M1-6) while Q213 seems important for both steps of the reaction (M3-7). Interestingly, the M3-11 (A177N, Q213G) mutant retains more second-step reaction activity than the single-position mutant M3-7 (Q213G) pointing to a significant contribution from the carboxamide functional group of this residue in the active site of BbSS.

Although these results were neither the perfect reciprocal nor a complete structure-function map of these enzymes, the positions identified in this analysis are interesting and important to consider along with prior work in the field. Poulter (1990) surmised, largely through a chemical rationale, that squalene/botryococcene product specificity is likely maintained by strict regioselectivity of the hydride transfer from NADPH. This conclusion is in opposition to the idea that PSPP conformation in the enzyme-substrate complex plays a major role in the final 1'-1 or 1'-3 linkage.<sup>59</sup> Poulter (1990) also suggested that for the 1'-1 linkage to occur with the observed stereochemistry, a stereo-electronic barrier is required to preclude premature hydride transfer at the C1' carbocation position prior to the c1'-1-2 rearrangement (see **Figure 2.1** for reaction cascade) such that an inversion at C1' is caused by hydride transfer. This barrier was presumed to be unnecessary for the observed 1'-3 linkage.<sup>59</sup>

Results from the current studies suggest all of these possible mechanisms may contribute to the product specificity observed in BbSS and SSL-3 (**Figure 2.21**). For instance, the notion of a stereo-electronic barrier was supported in the context of the Q213 and G207 positions in BbSS and SSL-3, respectively. These positions reside on the solvent face of the cyclopropyl ring in the active site. The BbSS Q213 carboxamide could act as the stereo-electronic barrier needed to prevent premature hydride transfer from NADPH to C1' during squalene biosynthesis. In SSL-3, the corresponding G207 position lacks the carboxamide functionality, which fits well with the idea that this barrier is unnecessary for the 1'-3 linkage. Alternatively, these positions could aid in the positioning of the NADPH cofactor such that when Gln is present in this position of the active site, hydride transfer proceeds to the C1' position as observed in squalene. These circumstances alone, however, cannot fully be responsible for either the 1'-1 or 1'-3 linkage because in SSL-3, both mutations (N171A, G207Q) are necessary to convert product specificity from botryococcene to squalene. It seems likely that during the cascade of squalene biosynthesis, proper spatial requirements are necessary for the chemical gymnastics that occur during the c1'-2-3 → c1'-1-2 → 1'-1 reaction cascade. The proper amount of space seems important for this to occur and could be supplied by in part by the A177 position in BbSS, which is asparagine in the corresponding position of SSL-3. This would likely prevent the necessary spatial requirements during the rearrangement of PSPP to squalene in SSL-3.

In light of the spatial requirements discussed above, an interesting model was presented by Lin *et al.* (2010) regarding the dynamic movements of the FPP and PSPP diphosphate groups from one aspartate-rich motif to the other during the dehydrosqualene catalytic cascade based on structures solved with less reactive thio-FPP analogs and PSPP.<sup>26</sup> Their model suggests that the FPP-





**Figure 2.21** 3D models of BbSS and SSL-3 with PSPP docked into active site showing residues known to affect product specificity as spheres. Model depicts orientation of PSPP in relation to the residues identified in this study that control product specificity in SSL-3. Residues shown with carbon backbone sphere colored magenta could be acting to orient PSPP in the active site such that  $c1'-3$  bond breakage and reformation to  $c1'-1-2$  is favored in BbSS while  $c1'-2$  breakage to yield botryococcene is favored in SSL-3.

diphosphate in the S1 site (aspartate-rich motif proximal to CrtM N-terminus; present in domain I) is the prenyl-diphosphate that ionizes, followed by electrophilic attack of the C1' carbocation on the C2-C3 double bond of the FPP in the S2 site to yield PSPP. The PSPP-diphosphate is then proposed to migrate from the S2 site (aspartate-rich motif proximal to CrtM C-terminus; present in domain III) to the S1 site where again, the diphosphate is ionized to initiate cyclopropyl rearrangement to dehydrosqualene. Although second-step reactions of CrtM and SS yield slightly different products, the dynamics of the substrates described here are likely similar for both enzymes based on protein 3D similarities. Given these results, previous chemical rationale about the mechanism, and our mutagenesis results, it seems plausible that more flexibility within the active site is necessary for the  $c1'-2-3 \rightarrow c1'-1-2 \rightarrow 1'-1$  rearrangement of squalene biosynthesis. This flexibility could be provided by A177 in BbSS to allow for the necessary movements during catalysis of the 1'-1 linkage with Q213 playing an important role as a stereo-electronic barrier and/or in NADPH positioning. SSL-3 likely favors the 1'-3 linkage over the 1'-1 linkage by fixing the substrate in the active site *via* N171 and removal the stereo-electronic barrier and/or repositioning of NADPH by G207 possibly by constricting the size of the active site pocket.

Until the NADPH binding site of 1'-1 and 1'-3 triterpene synthases is empirically defined, using either crystallographic or mutagenic studies, the conundrum of 1'-1 and 1'-3 linkages observed in squalene and botryococcene will remain unresolved. Moreover, based on the large number of mutants rationally designed and tested for this study and the incompleteness of the structure-function map, a more high-throughput technique will be of considerable importance for understanding the complex structure-function relationships of these enzymes.

## Materials and Methods

**3D modeling.** Generation of three-dimensional models for the *B. braunii* triterpene synthases was accomplished using MODELLER 9v8.<sup>63</sup> Online tutorials for basic modeling were followed using empirically determined structures of HsSS<sup>30,57</sup> (PDB: 1EZf and 3VJ8) and CrtM<sup>26</sup> (PDB: 2ZCO) as templates. *B. braunii* GenBank accession numbers are as follows: BbSS – AF205791.1, SSL-1 - HQ585058.1, and SSL-3 - HQ585060.1. Models were generated by first performing an alignment of the template and *B. braunii* enzyme of interest using MODELLER 9v8. The *B. braunii* enzyme sequence was then trimmed at its N- and C-terminal ends to match the length of the template. A second alignment was performed with the shortened *B. braunii* sequence and original template; this alignment was then used for the modeling program. All parameters were left as default with five models generated for each run. Models with the lowest DOPE score were chosen for subsequent analyses.

**PSPP docking.** PSPP was docked into HsSS (PDB: 3VJ8) and the *B. braunii* 3D models of BbSS and SSL-3 (threaded onto 3VJ8) using Genetic Optimisation for Ligand Docking (GOLD; Cambridge Crystallographic Data Centre, CCDC; Cambridge UK).<sup>68</sup> Proteins were set up as follows: all hydrogens added, all waters and ligands removed, distance constraints (3.5 – 4.5Å) set to anchor the C11 prenyl chain to L184, C9 prenyl chain to Y280, and cyclopropyl methyl group to L212 (distance constraints based on measurements of CrtM-PSPP (PDB: 3ADZ); conserved residues listed are numbered according to BbSS). Binding site was defined as a sphere with a 10Å radius centered on V180 in domain II. Twenty-five runs were performed for each GOLD run with 1,000,000 operations per run. Hydrophobic constraints were put on all hydrophobic residues within the active site sphere and all hydrophobic ligand atoms during docking runs. The PSPP file used for docking was created using atom coordinates from PSPP co-crystallized with CrtM (PDB: 3ADZ). These coordinates were submitted to the PRODRG server<sup>69</sup> (University of Dundee, UK) with the 'chirality', 'EM', and 'charges' options selected as yes, reduced, and yes, respectively. The MDL Molfile output with all hydrogens was saved in mol2 file format and used for docking. All other parameters were left as default during docking runs.

**Wild-type *B. braunii* triterpene synthase cloning.** Primers used for most constructs can be found in **Table 2.1**; plasmid DNA (courtesy of Dr. T.D. Niehaus)<sup>36</sup> was used as template. All wild-type and mutant genes used in this study were directionally cloned into the pET28a vector (Novagen) NdeI and HindIII sites such that expressed enzymes contain N-terminal hexahistidine tags that could be used for purification and immunodetection purposes. SSL-3 had an internal HindIII site intentionally altered at the time it was cloned into pET28a; mutants subsequently constructed from it contain the HindIII site deletion. Because BbSS and SSL-2 were predicted to be membrane-bound enzymes,<sup>36,61</sup> they were truncated at their C-terminal ends 65 and 87 amino acids, respectively, to produce catalytically competent, soluble enzymes that could be expressed in *E. coli*.

**Table 2.1** Primers used for cloning and screening enzymology constructs. Restriction sites are underlined. Nucleotides in lower case represent mutations.

Primer	5' → 3' Sequence
5' NdeI BbSS	GGAATTCC <u>CATATG</u> GGGATGCTTCGCTGGGG
3' HindIII BbSStr64	CCCAAGCTTTT <u>AGCTCCTCAATTCG</u> TCAAAGG
5' NdeI SSL-1	GGAATTCC <u>CATATGACTATGCACCAAGACCACGG</u>
5' NcoI SSL-1	CATGCCATGGGCACTATGCACCAAGAC
3' HindIII SSL-1	CCCAAGCTTTC <u>ACTTGGTGGGAGTTGGGGCTGC</u>
5' NcoI His6 SSL-2	CATGCCATGGGCACTCACCATCACCATCACCATT CAGGAAGTATGGTGA <u>AACTCGT</u> CGAG
3' XhoI SSL-2tr87	CCGCTCGAGTTAGCTGAGAGCGGCTTTGCATG
5' SSL-2 (-NcoI)	GGGAGGCGAGTGGAGACCCaATGGTGGCCACA ACGCTGAAGC
3' SSL-2 (-NcoI)	GCTTCAGCGTTGTGGCCACCATtGGGTCTCCACTC GCCTCCCC
5' NdeI SSL-3	GGAATTCC <u>CATATGAAACTTCGGGAAGTCTTGCAG</u> C
5' NcoI SSL-3	CATGCCATGGGCAAACTTCGGGAAGTCTTGCAGC ACC
3' HindIII SSL-3	CCCAAGCTTTCAAGCACCCCTTAGCTGAAACC
5' SSL-3 (-HindIII)	GTTTGCAGACCCAAAaCTTCTGGACCGGGAG
3' SSL-3 (-HindIII)	CTCCCGGTCCAGAAGtTTTGGGTCTGCAAAC
5' M1-4 IF	gcttttacaataatGGGGTGGTGGGTCTCGGGCTTTCC C
3' M1-4 IR	caccccattatttgtaaaagcGTAAAGGTCGTAGTCCCCCA C
5' M1-5 IF	TACTGCCACTATAatGCTGGGGTGGTGGGTCTCGG GCTTTCCC
3' M1-5 IR	GGGAAAGCCCGAGACCCACCACCCAGCattATAG TGGCAGTA
5' M1-6 IF	TACTGCCACTATGTTaatGGGGTGGTGGGTCTCGG GCTTTCCC
3' M1-6 IR	GGGAAAGCCCGAGACCCACCACCCcattAACATAG TGGCAGTA
5' M1-7 IF	TACTGCCACTATAataatGGGGTGGTGGGTCTCGGG CTTTCCC
3' M1-7 IR	GGGAAAGCCCGAGACCCACCACCCcattattATAGT GGCAGTA
5' M3-7 IF	GGCCTCTTCCTTgggAAGACCAACATCATCC
3' M3-7 IR	GGATGATGTTGGTCTTcccAAGGAAGAGGCC
5' M3-8 IF	GGCCTCTTCCTTaatAAGACCAACATCATCC
3' M3-8 IR	GGATGATGTTGGTCTTattAAGGAAGAGGCC
5' M3-11 IF	USE M1-6 AND M3-7 PRIMERS
3' M3-11 IR	USE M1-6 AND M3-7 PRIMERS
5' M4-3 IF	GCCTTCACTAATgccGGGCCAGTTGCTATC

**Table 2.1** (continued)

3' M4-3 IR	GATAGCAACTGGCCCggcATTAGTGAAGGC
5' M4-4 IF	CATGGCCATGTTCTTGcagAAGATTAACGTCATCC
3' M4-4 IR	GGATGACGTTAATCTTctgCAAGAACATGGCCATG
5' M4-10 IF	USE M4-3 AND M4-4 PRIMERS
3' M4-10 IR	USE M4-3 AND M4-4 PRIMERS
5' XbaI op3DXS	GCTCTAGAATGAGCTTCGACATCGCCAAGTATCC
3' KpnI op3DXS	CGGGGTACCTCACGCCAGCCACGCCTTGATCTTC G
5' XbaI op3idi	GCTCTAGAATGCAGACCGAGCACGTCATCCTCC
3' KpnI op3idi	CGGGGTACCTCACTTGAGCTGGGTGAACGCCG
5' XbaI op3FPS	GCTCTAGAATGCAGCCGCACCACCACCATAAGG
3' KpnI op3FPS	CGGGGTACCTCACTTCTGGCGCTTGATAGATCTTC
5' SacI plac	AATTGCAGAGCTCATTATCAGGGTTATTG
3' XbaI plac	TTCTCTAGATTAATTTCTCCTCTTTAATTC

For *in vivo* analyses of mutants co-expressed with SSL-1 or SSL-3 in *E. coli*, SSL-1 and SSL-3 were cloned into MCS1 NcoI/HindIII sites of pACYCDuet-1 (Novagen). All constructs were verified by BigDye Terminator automated sequencing (ABI).

*PCR construction of mutant genes.* Sequence overlap extension polymerase chain reaction (SOE PCR)<sup>70</sup> afforded the ability to construct mutants described in this work as follows. Internal primers containing the altered nucleotide(s) of interest (IF, IR primers in **Table 2.1**) were used along with either the 5' or 3' end primer to first PCR amplify the gene of interest in two fragments in two separate reactions. The two gene fragments were gel purified and used as template in a second SOE PCR with the 5' and 3' end primers to reassemble the mutated gene. The reassembled gene was gel purified, digested with NcoI/HindIII and ligated into the same sites of pET28a. Reassembly was routinely performed with up to four fragments of a gene in an SOE PCR. Conditions for the first PCR and second SOE PCR were 98°C, 1 minute initial denaturation followed by 25 cycles of 98°C, 10 seconds → 55°C, 15 seconds → 72°C, 1min/kb with a final elongation step of 72°C, 10 minutes. All constructs were verified by BigDye Terminator automated sequencing (ABI).

*Construction of a methyl erythritol phosphate (MEP) pathway overexpression cassette.* pSB1:operon3 plasmid construction involved relocating the Tetracycline resistance (Tet<sup>R</sup>) marker from pBBR1MCS-3 to the pBBR1MCS-2 backbone.<sup>71</sup> To accomplish this, the kanamycin resistance (Kan<sup>R</sup>) marker was removed by first digesting pBBR1MCS-2 with BglII then treating with Klenow to obtain blunt ends. This product was then digested with KpnI to completely excise the Kan<sup>R</sup> gene from the vector followed by gel purification of the vector backbone. Tet<sup>R</sup> was then digested from pBBR1MCS-3 with KpnI/ZraI and ligated into the pBBR1MCS-2 backbone; clones containing the Tet<sup>R</sup> marker were isolated *via* tetracycline resistance selection on LB-Tet plates (12.5 mg/L). This vector was named pSB. The A1/03/04 Lac promoter was PCR amplified from pIND4<sup>72</sup> (courtesy of Dr. Judith Armitage) using primers 5' SacI plac / 3' XbaI plac and cloned into the SacI and XbaI sites of pSB resulting in the expression vector pSB1. To enhance *in vivo* titers of the substrate FPP, a synthetic operon consisting of a deoxyxylulose 5-phosphate synthase gene (DXS) and the isopentenyl diphosphate isomerase gene (*idi*) from *E. coli* and a farnesyl pyrophosphate synthase gene (FPS) from *Gallus gallus* was constructed. The genes were synthesized (GenScript Corporation) as a single polycistron with an optimized Shine Dalgarno sequence in front of each gene for expression in *E. coli* (AGGAGGA).<sup>73</sup> The operon consisting of DXS, *idi*, and FPS was digested from pUC57-operon3 as an XbaI/BamHI fragment and subcloned into the same sites of pUC19 to give construct pUC19:operon3. The entire operon3 was cloned into pSB1 as an XbaI/KpnI fragment to afford pSB1:operon3. This construct was functionally verified in *E. coli* BL21 (DE3) by co-expressing it with pET28a:BbSS and analyzing squalene levels. Additionally, each gene of operon3 was individually cloned into pSB1 using the XbaI/KpnI sites in pSB1 to yield the vectors pSB1:DXS, pSB1:*idi*, and pSB1:FPS. These constructs were verified by BigDye Terminator automated sequencing (ABI).

*Protein expression.* The following general scheme for protein expression was utilized throughout the course of this work for all wild-type and mutant enzymes. Fresh transformations of *E. coli* BL21 (DE3) (Life Technologies) cells with a given construct were always performed. Single colonies were selected to start 2 mL LB-Kan (50 mg/L) cultures that were incubated overnight at 37°C on a platform shaker. 500 µL of the overnight culture was used to inoculate 50 mL of Terrific Broth (TB) supplemented with 1% glycerol and 50 mg/L Kan in a 250 mL flask. Cultures were incubated at 37°C for 2 hours on a platform shaker then cooled on ice for 2-3 minutes and induced with IPTG (0.5 mM). Protein expression was allowed for 5 hours at 23°C on a platform shaker. Cultures were divided into 10 mL aliquots and centrifuged (2000 x g, 10 min) to pellet cells; growth media was decanted and pellets were stored at -80°C for use up to two weeks later.

*Hot <sup>3</sup>H-FPP in vitro enzyme assays.* Activity screening assays were conducted as described by Niehaus *et al.* (2011) and Okada *et al.* (2004).<sup>35,36</sup> General assays were performed as follows. Ten mL cell culture pellets were thawed on ice and reconstituted in 1 mL of cold reaction buffer (50 mM MOPS, pH 7.3, 2.5 mM β-mercaptoethanol, 20 mM MgCl<sub>2</sub>). Cells were lysed using a probe dismembrator (50% power, 3 bursts, 10 seconds each, 5 minutes ice between bursts) at 4°C followed by centrifugation clarification (4°C, 10,000 x g, 10 minutes). Five µL of crude lysate(s) were used in 50 µL assays with 30 µM FPP (S.A. ≈ 1250 DPM/pmol) and 5 mM NADPH. The reaction mixture was incubated at 37°C for 30 minutes followed by termination with 50 µL of stop reagent (0.4 M NaOH, 0.2 M EDTA). Hydrocarbon products were extracted with 200 µL of hexane. 150 µL of organic phase was removed to a fresh tube, dried to completion, and resuspended in 15 µL of hexane. All 15µL of resuspension was used for TLC (silica gel 60 plates, hexane solvent) to separate botryococcene and squalene. Authentic standards of squalene and botryococcene were ran alongside experimental samples to define zones from which the silica gel was scraped into a scintillation vial, dissolved in scintillation cocktail, and quantified using scintillation spectroscopy.

*Cold FPP in vitro enzyme assays.* Assays containing only one enzyme were incubated with 50 µM FPP in a volume of 400 µL for 30 minutes at which point 100 µL of 25 mM NADPH was added or 100 µL of reaction buffer for assays without NADPH. The reaction was then incubated at 37°C for 30 more minutes until 500 µL of stop reagent was added. Coupled assays performed with two enzymes were done identically as described except that 50 µL of BbSS, M3-11, SSL-3 or M4-10 crude lysate was added in addition to NADPH (50 µL of 50 mM) or buffer (50 µL) after the first 30 minute incubation at 37°C. Hydrocarbon product analysis consisted of two 1 mL extractions with hexane where 900 µL of the organic phase was removed after each extraction to a fresh vial and dried to completion. Samples were reconstituted in 50 µL of hexane and analyzed by GC-MS using 1 µL of the sample. GC-MS analyses were performed on an Agilent 7890A GC equipped with a HP-5MS column (30m x 0.25mm x 0.25µm) coupled to an Agilent 5975C MS insert XL MSD with Triple-Axis Detector (Agilent Technologies, Inc.). Initial oven temperature was held at 150°C for 1 minute

followed by two ramps: 10°C/min to 280°C → 5°C/min to 310°C and held for 1 minute. Mass spectral scan range was from 50-500 amu.

*In vivo chemical profiling.* Wild-type and mutant triterpene synthases were chemically profiled *in vivo* by co-expressing operon3, a triterpene synthase, and the pACYCDuet-1 expression vector containing SSL-1 or SSL-3 (for triterpene synthase only analyses, empty pACYCDuet-1 vector was present to account for metabolic burden of maintaining the additional plasmid in co-expression analyses). Initially, pSB1:operon3 was transformed into *E. coli* BL21 (DE3) and plated on LB-Tet (12.5 mg/L); from this transformation a single colony was selected to generate competent cells. Competent cells were then transformed with a pET28a:construct and pACYCDuet-1:construct or empty pACYCDuet-1. Transformations were plated on LB-Tet-Kan-Chlr (12.5 mg/L Tet, 25 mg/L Kan, 12.5 mg/L Chlr). Single colonies were chosen and cultured in 2 mL of LB-Tet-Kan-Chlr overnight at 37°C on a platform shaker. Fifty µL of overnight culture was used to inoculate 5 mL of TB-Tet-Kan-Chlr which was grown at 37°C for 2 hours on a platform shaker. Cultures were then cooled on ice for 1-2 minutes until IPTG was added (0.5 mM) followed by incubation at 23°C for 24 hours on a platform shaker. Samples were frozen at -80°C until needed. Extractions were done in 4 mL screw cap vials using 1 mL culture: 1 mL acetone: 1 mL hexane. Extraction mixtures were vortexed vigorously 5 times for 30 seconds before centrifugation at 2000 x g for 2 minutes. Three hundred fifty µL of organic phase was removed to a 2 mL GC vial insert where it was allowed to evaporate at room temperature. Samples were reconstituted in 50 µL of hexane and analyzed *via* GC-MS as described above.

*Radiolabeled crude relative efficiency assays.* Reaction schemes were based on those described by Greenhagen *et al.* (2003) where product accumulation was measured using incremental amounts of one enzyme and a constant amount of another enzyme in a coupled assay as an assessment of enzyme efficiency.<sup>66</sup> Crude lysates for these assays were prepared identically as described earlier. Total protein amounts in crude lysates were measured using a Bradford dye assay and quantified by comparison with a BSA standard curve. Scheme A assays were setup with increasing amounts (0, 1, 2.5, 5, and 7.5 µg) of SSL-1 crude lysate in a 40 µL reaction with 50 µM FPP (S.A. ≈ 250 DPM/pmol) in the same reaction buffer conditions described earlier. Reactions were incubated at 37°C for 30 minutes at which point 5 µL of 50 mM NADPH was added followed by 5 µg of SSL-3 or M4-10 crude lysate. Reaction final volumes were adjusted to 50 µL with reaction buffer and incubated at 37°C for 30 more minutes followed by termination, extraction, and analysis as described earlier. Scheme B assays were setup with a constant amount (2.5 µg) of SSL-1 crude lysate in a 40 µL reaction with 50 µM FPP (S.A. ≈ 250 DPM/pmol). The reaction was incubated at 37°C for 30 minutes until 5 µL of 50 mM NADPH was added followed by the addition of increasing amounts (0, 1, 2.5, 5, and 7.5 µg) of either SSL-3, M4-10, BbSS, or M3-11 crude lysate. Final volumes were adjusted to 50 µL with reaction buffer and incubated at 37°C for 30 more minutes. Calculation of relative efficiency was made as follows: pmol squalene / assay for M4-10 with 1 µg SSL-1 crude lysate divided by pmol botryococcene / assay for SSL-3 with 1



$\mu\text{g}$  SSL-1 crude lysate; this value was then multiplied by 100. Example:  $(43.09/53.98)*100 = 79.82\%$ . Percentage values for the other amounts of SSL-1 crude lysate tested were then calculated; all percentage values were then averaged to yield the final relative efficiency for the given Scheme (A or B) assessment.  $([\text{mutant activity} / \text{wild-type activity} * 100])$  for each varying amount then averaged to yield the final relative efficiency)

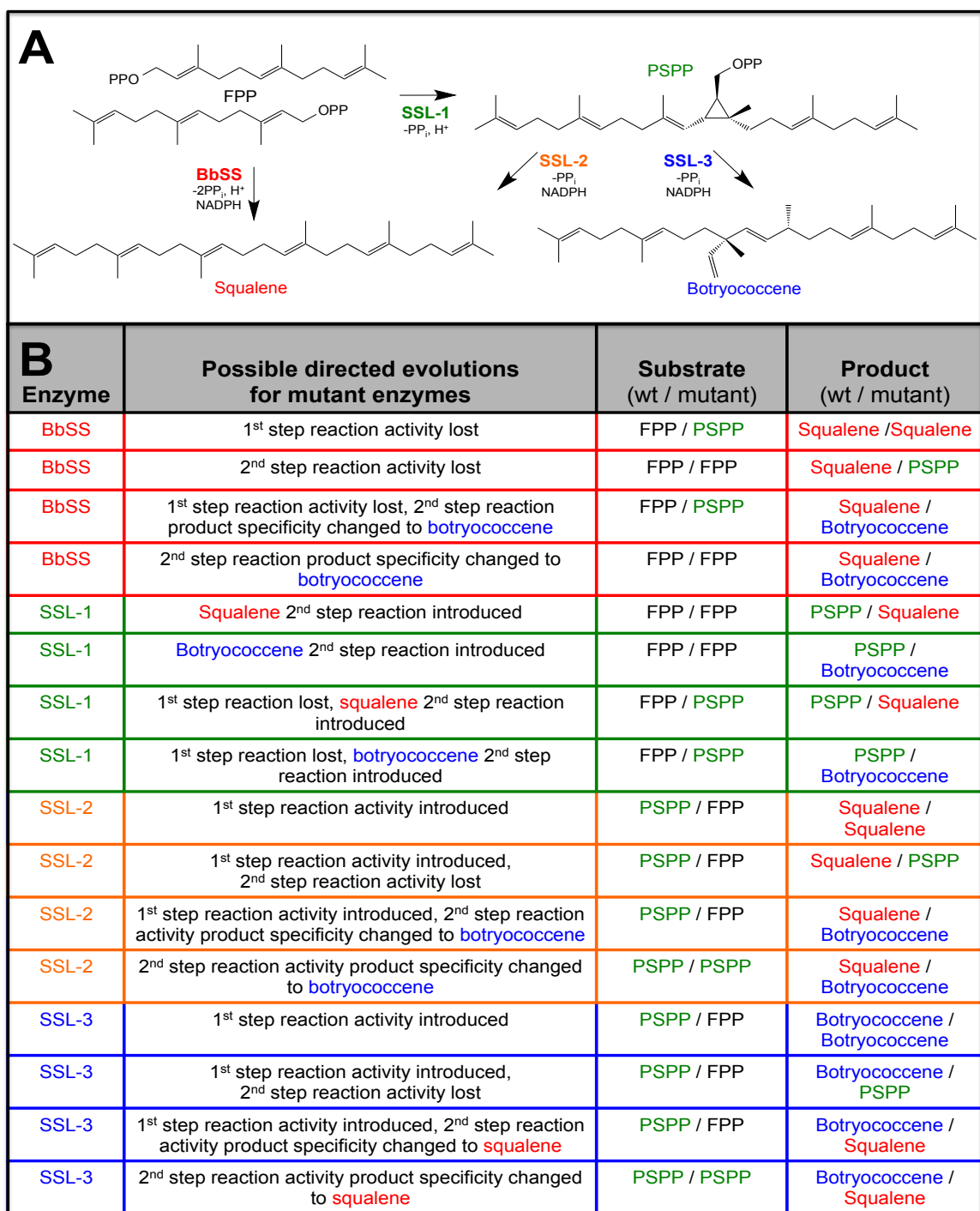
## Chapter Three: Development of a Selection Platform for the Directed Evolution of Triterpene Synthases

### Background and Introduction

Squalene synthase (EC 2.5.1.21) catalyzes the first dedicated step in eukaryotic sterol biosynthesis by coupling two molecules of farnesyl diphosphate (FPP) together to yield squalene.<sup>74</sup> The importance of squalene for sterol biosynthesis is well documented in plants, animals, and fungi. A squalene synthase gene knockout in organisms from these kingdoms is lethal.<sup>75,76,77</sup> Squalene synthase and other enzymes involved in sterol biosynthesis have drawn considerable attention in the past, particularly kingdom specific differences that can be exploited as targets for anti-fungal drugs.<sup>78</sup> Although differences are typically found in post-oxidosqualene steps of the pathway, interest remains in squalene synthase at the molecular level.<sup>25</sup> Understanding how to regulate squalene synthase activity could have important medical implications for controlling cholesterol levels in humans.<sup>7</sup> Alternatively, deregulating this enzyme's activity has industrial application for the production of squalene and other triterpenes used in vaccine adjuvant formulations and cosmetic products.<sup>79,80</sup> Squalene and squalene analogs could even be used as a feedstock for the production of liquid transportation fuels.<sup>81</sup>

More knowledge of the structure-function map for squalene synthase could translate to better rational design of mutant enzymes with more desirable characteristics (*i.e.*, improved catalytic efficiency, thermostability, or even tailored substrate/product specificity). The design of small molecule inhibitors targeting squalene synthase could also benefit from an improved understanding of this enzyme. Squalene biosynthesis has been of interest for decades prompting detailed kinetic and structural assessments under various conditions.<sup>19,23-25,30,37,56,65</sup> Collectively, these studies support a two-step mechanism for squalene synthase where both steps take place in the same active site cavity without release of the intermediate. First, two molecules of FPP are joined head-to-head to form the cyclopropyl intermediate presqualene diphosphate (PSPP).<sup>22,55</sup> PSPP is then reductively rearranged to yield a 1'-1 linked squalene with a requirement for hydride transfer from NAD(P)H in the second step.<sup>82</sup> **(Figure 3.1)** Although many groups have reported mutations that affect squalene synthase activity,<sup>6,31-57</sup> many of these have been loss-of-function mutations that only allow for relatively constrained inferences. In contrast, gain-of-function mutants could provide positive evidence for ascribing a contribution of a residue or domain to a catalytic role, which is the intent of the work reported in this chapter.

Squalene synthase-like (SSL) enzymes from the green alga *Botryococcus braunii* race B synthesize squalene and a 1'-3 linked molecular relative, botryococcene.<sup>36</sup> Interestingly, these enzymes adopt a similar mechanism to squalene synthase but instead accomplish it with two separate enzymes. SSL-1 catalyzes the first step by joining two molecules of FPP to produce PSPP. This PSPP can then be acted on by either SSL-2 to produce squalene or SSL-3 to form botryococcene with both reactions requiring NADPH **(Figure 3.1)**.



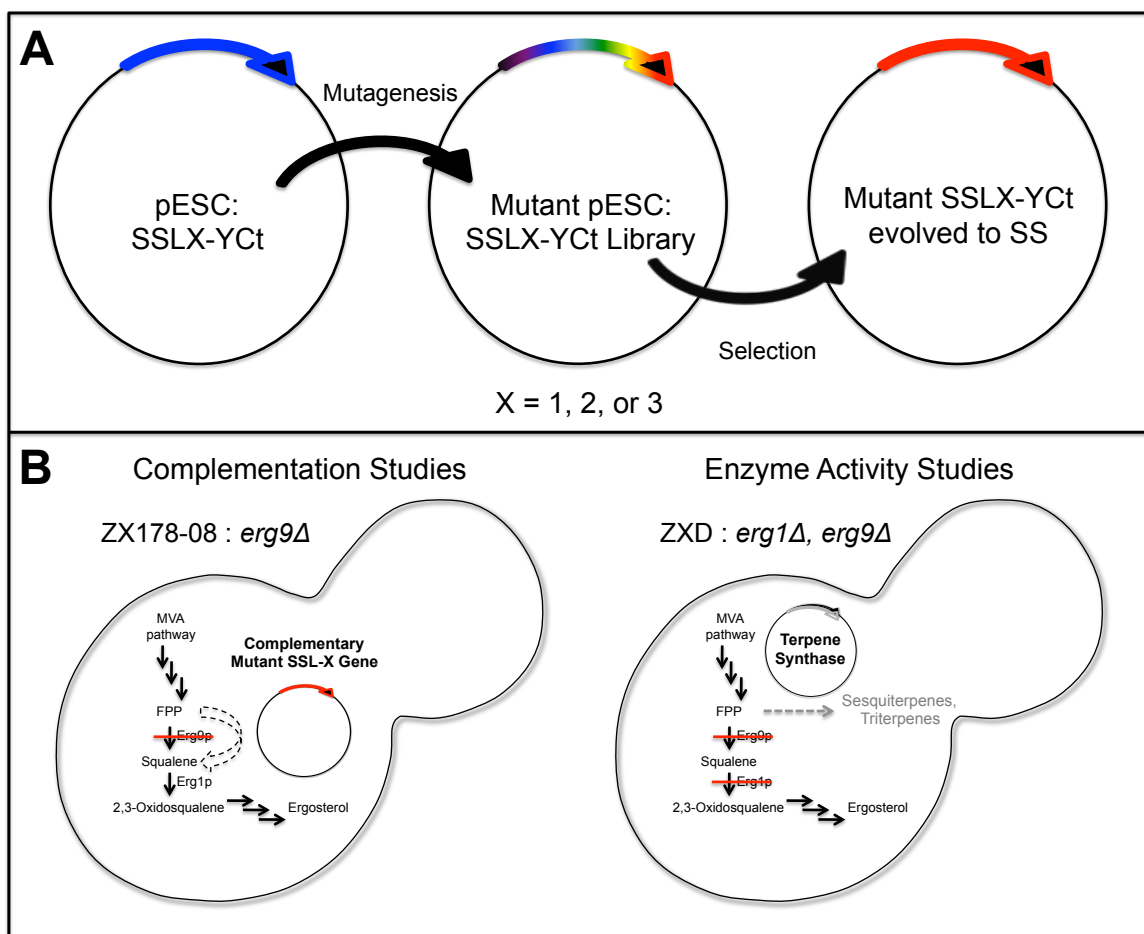
**Figure 3.1** Two-step reaction mechanism for the biosynthesis of squalene and botryococcene and possible directed evolutions for SSL enzymes. SSL-1 forms PSPP from two molecules of FPP in the first step followed by reductive rearrangement of PSPP to squalene by SSL-2 or botryococcene by SSL-3 in the second step. BbSS performs both steps of squalene biosynthesis within one enzyme active site (A). Possible effects of mutations introduced in BbSS, SSL-1, SSL-2, and SSL-3 on substrate and product specificity (B). List is not comprehensive.

SSL-2 also catalyzes the formation of an ether linkage between two molecules of FPP to produce bis-farnesyl ether (bFE) by a novel  $SN_2$  Williamson ether synthesis-type reaction.<sup>36</sup> *B. braunii* also harbors a bona fide squalene synthase (BbSS) that likely supports necessary requirements for sterol biosynthesis.<sup>61</sup> For reasons not entirely clear, *B. braunii* evolved to become very proficient at the production of these triterpene hydrocarbons,<sup>62</sup> likely through the action of the SSL enzymes though this has not been unequivocally demonstrated in the alga itself. The oils associate with an extracellular matrix housing the algal cells, comprise 30% (w:w) of *B. braunii*'s biomass, and give it the ability to float in undisturbed water.<sup>9,62</sup>

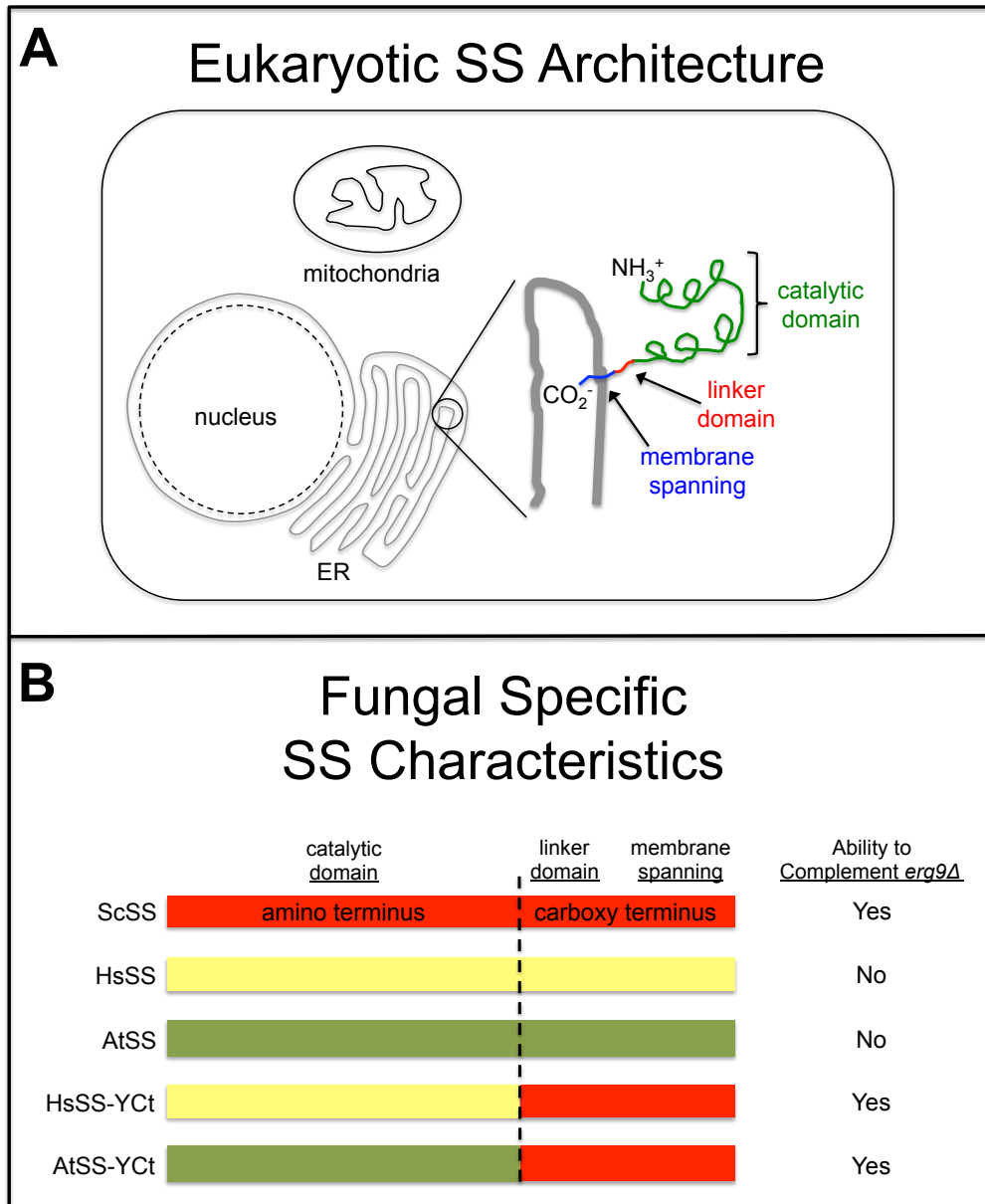
Regardless of the evolutionary advantage these enzymes might offer to *B. braunii*, SSL-1, SSL-2 and SSL-3 present a unique opportunity to study the two reaction steps for squalene and botryococcene biosynthesis separately. Rational mutant design had been employed with BbSS, SSL-1, SSL-2 and SSL-3 with intentions of interconverting substrate and/or product specificities among the BbSS and SSL enzymes by introducing reciprocal mutations in putatively critical positions within the active site cavity that differed between the respective enzymes. While not a saturating assessment, differences between two active site residues of BbSS and SSL-3 were shown to influence catalytic outcomes. Notably, a double mutant (N171A, G207Q) in SSL-3 was identified that converted product specificity from botryococcene to squalene; reciprocal mutations in the corresponding positions of BbSS (A177N, Q213G) did not, however, convert its product specificity from squalene to botryococcene. These and aforementioned studies provide valuable insight into the structure-function map for triterpene synthases. While the SSL enzymes present a great avenue through which structure-function studies can be conducted, the lack of sequence identity among the protein sequences (~60%) coupled with the number of possible outcomes for given SS and SSL mutant enzymes (**Figure 3.1B**) has made rational design approaches difficult at this stage.

A robust screening platform allowing for the identification and selection of 1'-1 and 1'-3 triterpene biocatalysis from large pools of mutants would be desirable. Such a selection platform would need to meet a variety of criteria (**Figure 3.2**). Studies with *Saccharomyces cerevisiae* have demonstrated that while a squalene synthase knockout (*erg9Δ*) is typically lethal, exogenous ergosterol can complement the *erg9* knockout as long as the yeast also harbors a second mutation allowing for the uptake of exogenous sterol.<sup>83</sup> Interestingly, previous work with *erg9Δ* yeast lines has shown that expression of a gene encoding a non-fungal squalene synthase in an *erg9Δ* yeast line does not genetically complement the loss of the native *ERG9* gene. However, catalytic domains of heterologous squalene synthases fused with the carboxy terminus from a fungal squalene synthase does complement (**Figure 3.3**).<sup>84-86</sup> Given this information, we asked if wild-type SSL enzymes that already make squalene could complement an *erg9* knockout in *S. cerevisiae* and if so, could a mutant library of SSL-3 be created and screened with the *erg9Δ* yeast line to isolate the previously described SSL-3 N171A, G207Q double mutant as well as others. The present work was undertaken to establish a proof-of-principle for a yeast

selection platform that would allow mutant SSL enzyme libraries to be screened by *erg9* $\Delta$  complementation.



**Figure 3.2** Using *S. cerevisiae* as a selection platform for directed evolution of SSL enzymes. Mutant enzyme libraries are constructed in the pESC-URA vector with SSL-1, SSL-2, or SSL-3 appended to the ScSS carboxy-terminal domain (YCt) followed by selection for mutants that complement *erg9Δ* (A). Description of ZX178-08 and ZXD yeast lines used for complementation testing and chemical profiling of constructs used in this study. *ERG1* – squalene epoxidase, *ERG9* – squalene synthase. ZXD yeast line is used for chemical profiling to prevent triterpene products from further modification (B).



**Figure 3.3** Architecture of squalene synthase in the eukaryotic cell. Squalene synthase is a membrane-bound enzyme localized to the endoplasmic reticulum (ER) by its carboxy terminus.<sup>29,67,87,88</sup> This domain consists of a membrane spanning (blue) and linker region (red) that attaches to the amino terminus, or catalytic domain (green), which lies on the cytoplasmic face of the ER (A).<sup>84</sup> Complementation of *erg9Δ* in *S. cerevisiae* requires the carboxy terminus from ScSS or another fungal squalene synthase (B).<sup>29,67,84,86</sup>

## Results

*Establishing the yeast selection platform.* The yeast line ZX178-08 was developed by Dr. Xun Zhuang in the Chappell laboratory by first selecting for enhanced sterol uptake from the culture media when cells were grown under aerobic conditions.<sup>89</sup> Next, a deletion of the resident *ERG9* gene was introduced to disable squalene biosynthesis in the yeast. This line is derived from the laboratory yeast strain BY4741. Yeast line ZXD is derived from ZX178-08 by the deletion of the resident squalene epoxidase gene (*ERG1*), thus eliminating further metabolism of any squalene produced *via* the ergosterol biosynthetic pathway.

The first choice of promoter for gene expression was a strong, constitutive promoter to circumvent the additional steps needed to initiate gene expression with an inducible promoter. Initially, a divergent GPD/TEF promoter setup made by interchanging the Gal1/Gal10 divergent promoter in the pESC-URA vector series was utilized.<sup>89,90</sup> The initial constructs were made to assess the complementation ability of SSL-1 and SSL-2 in the *erg9Δ* yeast genetic background (ZX178-08).<sup>89</sup> However, upon transformation into ZX178-08 it was observed that any vector containing a SSL-1 construct never yielded transformants while colonies were always obtained with vectors lacking this gene (data not shown). From this observation, we surmised that the product of the SSL-1 enzyme, PSPP, might be toxic when made in excess and therefore chose to use the galactose-inducible gene expression system provided by the pESC-URA vector. With this setup, transformants were readily obtained under non-inducing conditions, allowing us to obtain the necessary independent transformants for subsequent analyses.

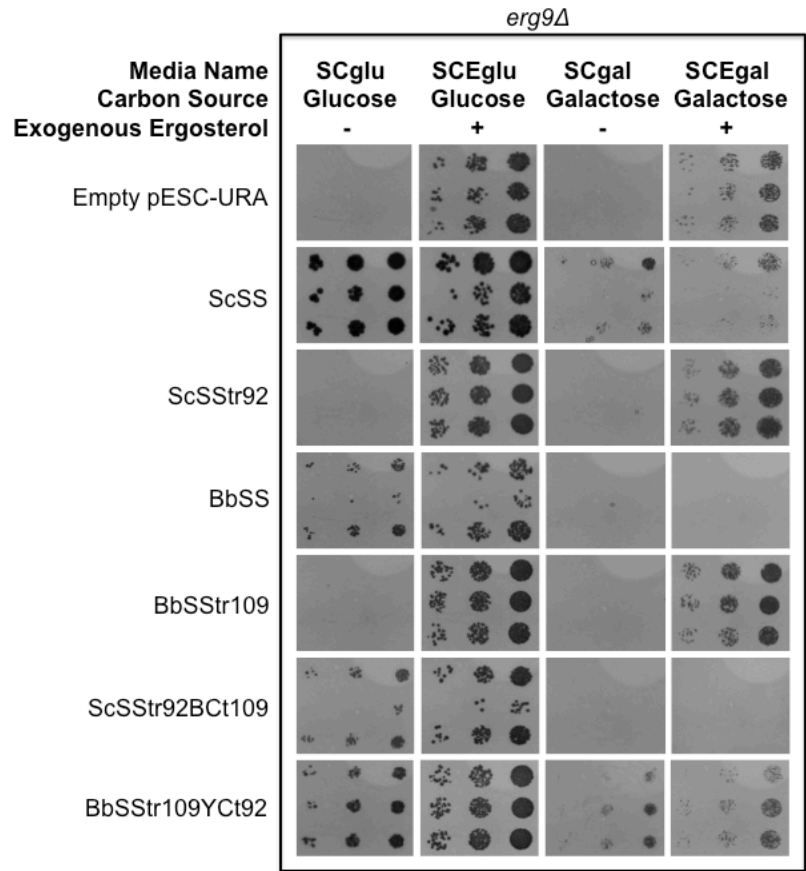
*Validation of inducible gene expression system and complementation testing in ZX178-08.* To validate the inducible gene expression system and provide confidence the complementation test was working, positive and negative control transformants were assessed. As previously mentioned, the non-catalytic C-terminal region of squalene synthase anchors the enzyme to the ER membrane and possibly aids in metabolon formation with other sterol biosynthetic machinery.<sup>84,86</sup> This region may be kingdom specific as plant and animal squalene synthases can only complement a fungal squalene synthase knockout when a fungal squalene synthase C-terminal (YcT) is substituted for the corresponding region. In an attempt to mirror work on squalene synthase by Niehaus (2011),<sup>84</sup> six constructs were made under control of the Gal1 promoter: two full-length enzymes (BbSS and ScSS), two C-terminally truncated enzymes (BbSStr109 and ScSStr92 – numbers correspond to amino acids truncated from the carboxy-termini), and two chimeric enzymes where the C-terminal regions were interchanged (BbSStr109YcT92 and ScSStr92BcT109). The ability of ZX178-08 harboring these plasmids to grow on four different media types were assessed: SCglu; SCEglu; SCgal; and SCEgal. SC represents minimal media, E represents the presence of exogenous ergosterol in the media, and glu/gal denote which carbon source is present for non-inducing (glu) and inducing (gal) gene expression conditions. Three independent biological transformants selected



for each construct were subsequently grown in SCEglu liquid media for three days, then 1 mL of cells were pelleted, washed and resuspended in water, then diluted 50, 250, and 1250 times (to an  $\sim$ OD<sub>600</sub> of 0.1, 0.02, 0.004, respectively). Five  $\mu$ L of each dilution for each replicate was spot plated on all four media types and grown at 28°C for five days, then transferred to room temperature for another ten days; pictures were taken at five days. The results validating the expression and complementation system as designed are shown in **Figure 3.4**.

As expected, ZX178-08 that contained empty vector (pESC-URA) or the truncated squalene synthases (BbSStr109 and ScSStr92) grew only in the presence of exogenous ergosterol independent of carbon source. Expression of ScSS (full length yeast squalene synthase) showed vigorous growth on SCglu and SCEglu but marginal growth on SCgal and SCEgal media types. These results were both surprising and unexpected. First, complementation when the yeast cells were grown on glucose containing media without exogenous ergosterol was unexpected because the *GAL1* promoter driving expression of the ScSS gene is not expected to be very significant. Leaky gene expression with the *GAL1* promoter has been previously reported in non-induced (glucose) conditions.<sup>91,92</sup> This result suggested that only a modest level of squalene synthase gene expression was necessary to complement the knockout of the native *ERG9* (squalene synthase) gene. When these same lines were grown in SCgal and SCEgal, media that should support strong expression of the heterologous yeast squalene synthase gene, growth was significantly compromised. Assuming that the heterologous yeast squalene synthase gene was vigorously expressed and producing catalytically competent enzyme, this result suggests that too much squalene production by the membrane-targeted squalene synthase imposed some level of toxicity that limited growth. The notion of site-specific production of squalene having an impact on growth is consistent with the complementation effects of carboxy-truncated heterologous yeast squalene synthase, ScSStr92. This construct yields a functionally soluble form of squalene synthase but is not able to complement *erg9* $\Delta$  either with weak expression (SCglu), or when it is strongly overexpressed (SCgal). Importantly, when these transformants were grown in media containing exogenous ergosterol vigorous growth was exhibited.

Heterologous expression of the algal full-length and truncated squalene synthase (BbSS and BbSStr109) corroborates these inferences. The growth pattern of ZX178-08 yeast lines expressing the truncated form of the algal squalene synthase resembles the pattern seen for the heterologously expressed yeast squalene synthase that is truncated. Leaky expression of the algal full-length squalene synthase in yeast grown on SCglu media was significantly less than when a wild-type yeast squalene synthase gene was used, but growth was evident. When these yeast lines were grown under strong inducing conditions, SCgal or SCEgal, no viable yeast were observed. These results again support an interpretation that complementation of the *ERG9* deletion is sensitive to both the expression level of the heterologous squalene synthase and its targeting to the ER membrane.



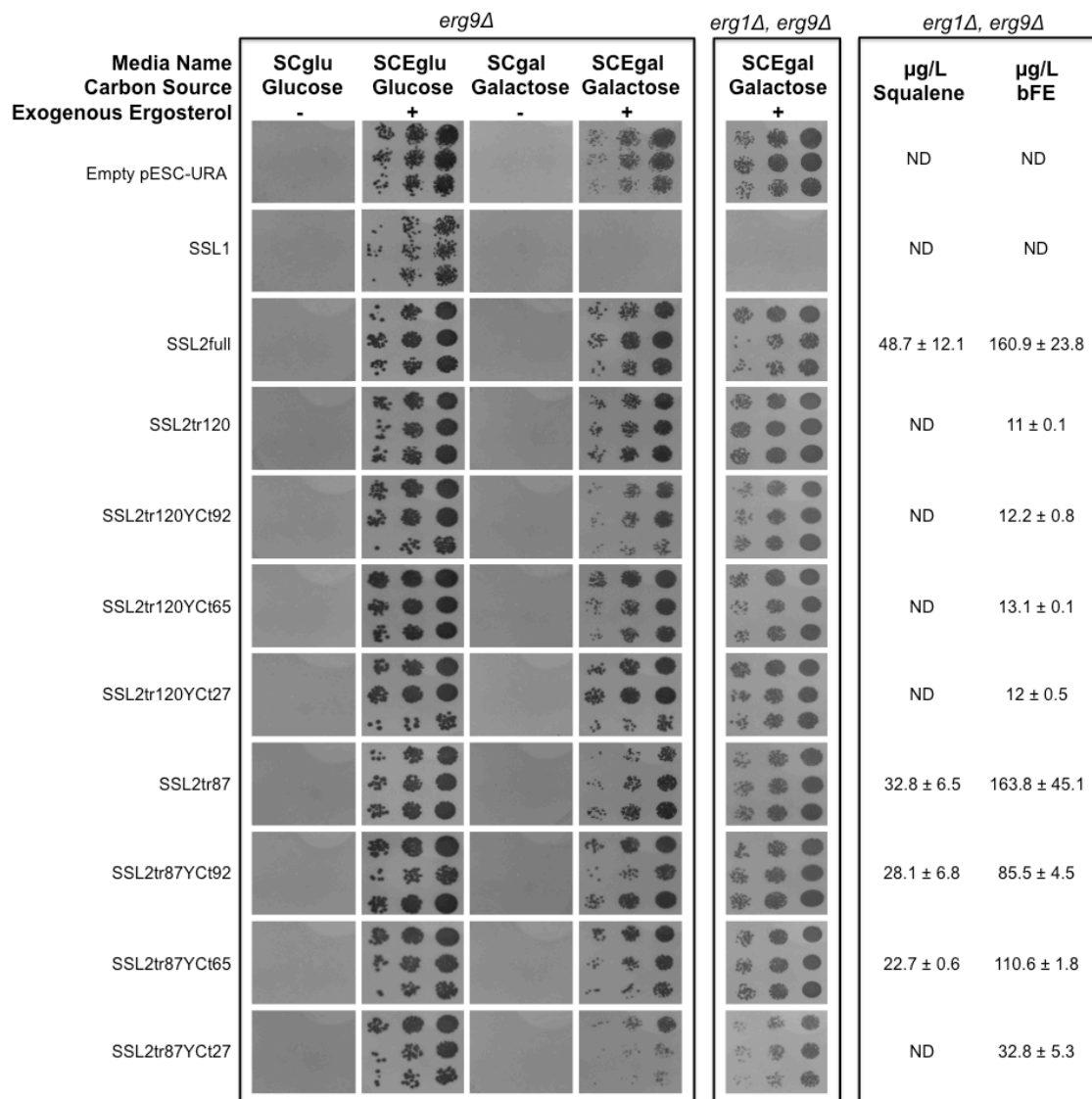
**Figure 3.4** Complementation and growth testing of ScSS and BbSS full-length, truncated, and chimeric constructs in ZX178-08 on four different media types. ZX178-08 harbors a deletion of the *ERG9* gene as denoted by the *erg9Δ* designation at the top of the figure frame. Three biological replicates (top to bottom within each panel) for each construct were grown for three days in liquid SCEglu media; 1 mL of cell cultures were pelleted, washed with water, then diluted 50X, 250X, and 1250X. Five  $\mu$ L of each dilution (lowest to highest dilutions from right to left within each panel) for each biological replicate was spot plated on the four different media types as indicated at the top of each column. Pictures were taken after five days of growth at 28°C.

Complementation with chimeric enzymes showed identical growth patterns as their full-length counterpart. BbSSStr109Yct92, which has its carboxy-terminal tail domain substituted with the yeast domain grew on all four media types as the full length ScSS did. In contrast, when the yeast squalene synthase had its carboxy-terminal domain replaced with that from the alga, the transformants harboring the ScSSStr92Bct109 construct grew only on SCglu and SCEglu as did the full-length BbSS construct. To confirm constructs were enzymatically active, chemical profiling was done by overexpressing the various constructs in the *erg1Δ*, *erg9Δ* yeast line (ZXD)<sup>89</sup> under SCEgal conditions then analyzing hexane-extractable products by GC-FID. All showed squalene product accumulation, albeit varying amounts were observed (data not shown). Additionally, when growth of all these ZXD lines in SCEgal media were monitored, they all grew comparable to the empty vector control lines, further suggesting that a product downstream from squalene could be toxic (data not shown).

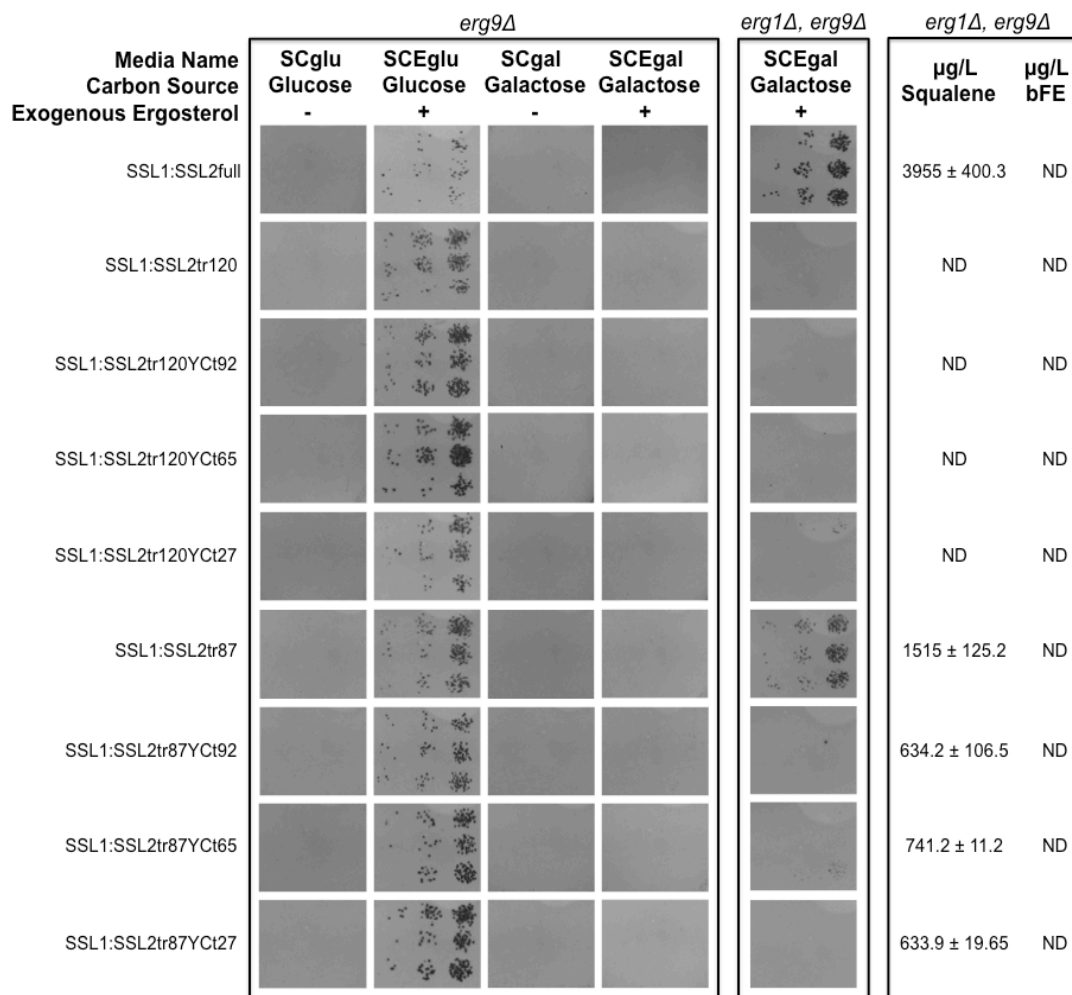
*Complementation testing of SSL-1 and SSL-2 single-gene and co-expressed constructs in ZX178-08.* Co-expression of the algal SSL-1 and SSL-2 genes in an industrial yeast line capable of high level FPP accumulation was previously shown to produce squalene.<sup>36</sup> Given the unexpected observation above that the site of squalene biosynthesis plays an important role in whether the squalene produced can be utilized by the resident ergosterol biosynthetic pathway, we wanted to determine if co-expression of SSL-1 for PSPF biosynthesis and SSL-2 for the conversion of PSPF to squalene could complement the *ERG9* deletion. The SSL-1 wild-type gene was expressed under control of the Gal1 promoter, while all the SSL-2 constructs were under control of the Gal10 promoter in the dual promoter pESC-URA vector. Because SSL-2 has a predicted membrane-anchoring region, we tested the full-length enzyme and two different C-terminal truncations with and without three permutations of the yeast carboxy terminal (Yct) region: the membrane-spanning region (Yct65); the linker domain (Yct27); and the complete carboxy terminus (Yct92). For reasons previously demonstrated with full-length BbSS, the resulting transgenic yeast lines were tested for complementation on SCglu/SCgal and growth on SCEglu/SCEgal. As controls, transformants harboring empty pESC-URA vector, SSL-1, and SSL-2 were assessed individually in ZX178-08. None complemented the squalene synthase knockout mutation when grown on SCglu or SCgal. All SSL-2 constructs and empty pESC-URA vector lines grew on SCEglu and SCEgal, while vectors with SSL-1 only grew on SCEglu (**Figure 3.5A**).

These results were not entirely unexpected based on earlier observations wherein vectors harboring the SSL-1 gene under transcriptional control of the strong, constitutive GPD promoter, comparable to the inducible Gal promoters, never yielded yeast transformants. These two results were consistent with a toxicity effect of the SSL-1 gene, perhaps caused by the accumulation of PSPF. Co-expression of SSL-1 and the SSL-2 constructs in ZX178-08 revealed the inability of these enzymes to complement the *erg9* knockout when expressed at low (SCglu) or high levels (SCgal). Growth was observed on SCEglu for all construct combinations but not on SCEgal, again suggesting a possible toxic

effect of these enzymes when overexpressed in the presence of high FPP levels (**Figure 3.5B**). Interpretation of these results is somewhat confounded by the earlier observation that high level expression of BbSS (a complete squalene synthase) inhibited growth of ZX178-08, as did the high level expression of SSL-1. Because the reaction products of the two respective enzymes, squalene for BbSS and PSPP for SSL-1, are different, the cause of toxicity might actually be different for the two genes.



**Figure 3.5A** Complementation, growth, and chemical profile analysis of SSL-1 and SSL-2 single-gene constructs in ZX178-08 and ZXD yeast lines. ZX178-08 harbors a deletion of the *ERG9* gene while ZXD contains deletions of both *ERG9* and *ERG1* genes as denoted by the *erg9Δ* and *erg1Δ, erg9Δ* designation across the top of the figure frame, respectively. Three biological replicates (top to bottom within each panel) for each construct in ZX178-08 or ZXD were grown for three days in liquid SCEglu media. One mL of cell culture was pelleted, washed, and then diluted 50X, 250X, and 1250X. Five μL of each dilution (lowest to highest dilutions from right to left within each panel) for each biological replicate was spot plated on the four different media types as indicated at the top of each column. Pictures were taken after five days of growth at 28°C. Chemical profiles were generated from the same three ZXD biological replicates used for the SCEgal growth test. Cultures were grown in liquid SCEgal media for seven days at room temperature; hexane-extractable products were quantified *via* GC-FID in comparison to authentic standards.

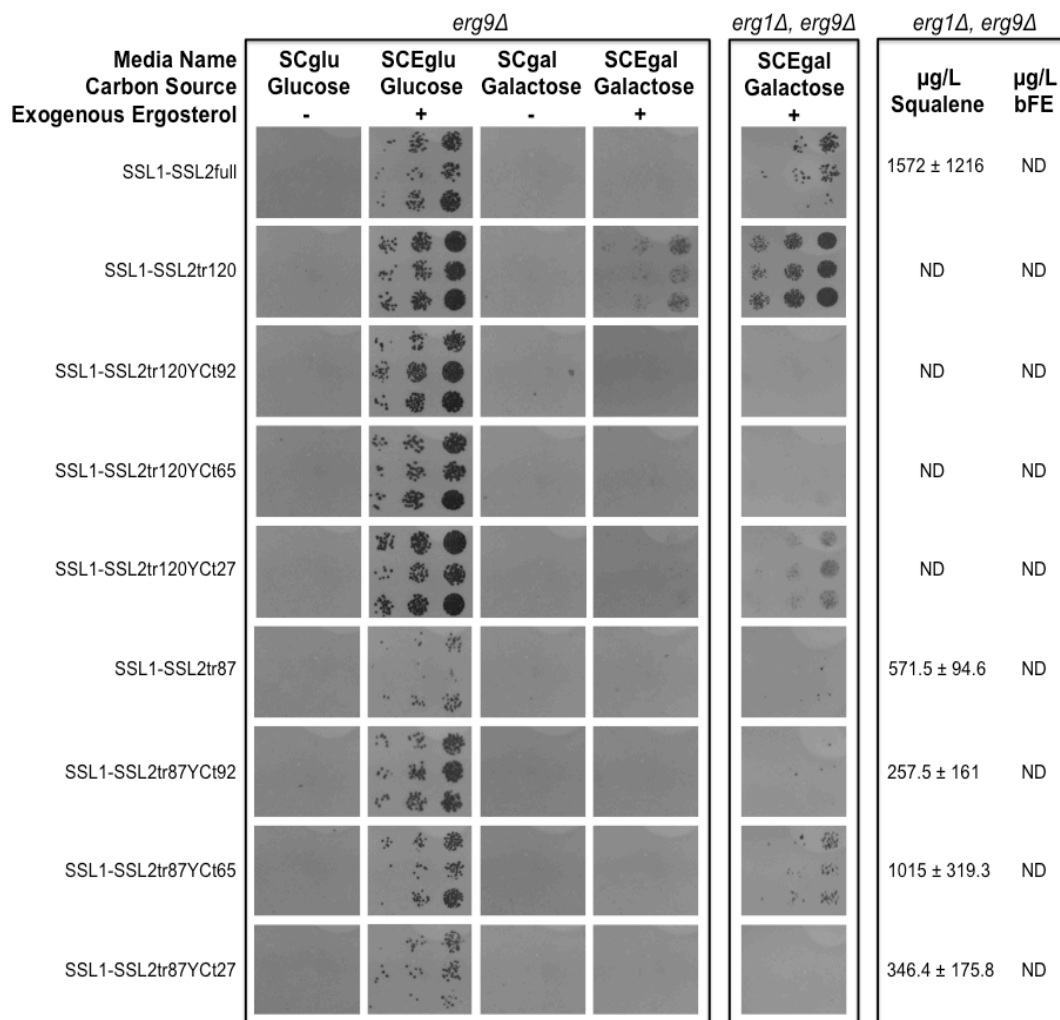


**Figure 3.5B** Complementation, growth, and chemical profile analysis of SSL-1 and SSL-2 co-expressed constructs in ZX178-08 and ZXD yeast lines. ZX178-08 harbors a deletion of the *ERG9* gene while ZXD contains deletions of both *ERG9* and *ERG1* genes as denoted by the *erg9Δ* and *erg1Δ, erg9Δ* designation across the top of the figure frame, respectively. Three biological replicates (top to bottom within each panel) for each construct in ZX178-08 or ZXD were grown for three days in liquid SCEglu media. One mL of cell culture was pelleted, washed, and then diluted 50X, 250X, and 1250X. Five μL of each dilution (lowest to highest dilutions from right to left within each panel) for each biological replicate was spot plated on the four different media types as indicated at the top of each column. Pictures were taken after five days of growth at 28°C. Chemical profiles were generated from the same three ZXD biological replicates used for the SCEgal growth test. Cultures were grown in liquid SCEgal media for seven days at room temperature; hexane-extractable products were quantified *via* GC-FID in comparison to authentic standards.

*Chemical profiling and growth assessment of SSL-1 and SSL-2 single-gene and co-expressed constructs in ZXD.* To verify functional enzymes were expressed in the yeast, SSL-1 and SSL-2 constructs were expressed alone and together in ZXD, followed by analysis of hexane-extractable products via GC-FID. Detecting the presence of presqualene alcohol (PSOH), the dephosphorylated form of PSPP, would indicate the expression of functional SSL-1. SSL-2, in addition to catalyzing the conversion of PSPP to squalene, also possesses the ability to catalyze the conversion of FPP to bis-farnesyl ether (bFE). Detection of bFE would indicate the expression of functional SSL-2 enzyme. Formation of squalene in yeast co-expressing SSL-1 and SSL-2 would demonstrate the functionality of both enzymes working together to catalyze the conversion of FPP to squalene. Finally, yeast line ZXD expressing the SSL-1 and SSL-2 genes singularly and together were assessed for growth on SCEgal. This experiment was intended to determine if the toxic effects observed in ZX178-08 were the result of SSL-1 producing excessive PSPP or SSL-1 and SSL-2 providing too much squalene to the ergosterol biosynthetic pathway.

Growth assessments for single-gene constructs showed that ZXD overexpressing SSL-1 in the presence of ergosterol (SCEgal) was unable to grow while all SSL-2 single-gene constructs were able to grow (**Figure 3.5A**). Only ZXD lines overexpressing SSL1:SSL2full and SSL1:SSL2tr87 were able to grow on SCEgal; the SSL1:SSL2tr120 construct series and SSL1:SSL2tr87Yct construct series showed no growth on SCEgal plates (**Figure 3.5B**). Because ZXD overexpressing SSL-1 would not grow on plates or in liquid culture, obtaining a chemical profile was not possible. Squalene and bFE product accumulation levels are listed in **Figure 3.5A** and **Figure 3.5B** for the single-gene and co-expressed gene constructs, respectively.

*Complementation testing of SSL1-SSL2 fusion constructs in ZX178-08.* Because of the inability of SSL-1 and SSL-2 to complement the *erg9* knockout in ZX178-08 when co-expressed as separate enzymes and the observation that SSL-1 overexpression caused yeast cell death, we supposed these enzymes either did not couple together well or the substrate affinity of SSL-2 for PSPP was so low that PSPP could not be converted to squalene efficiently enough before it caused cell death. To circumvent the former, we constructed fusion enzymes of SSL-1 and SSL-2 under control of the Gal1 promoter. SSL-1 wild-type sequence was used with a GGSGGGSGGGSG linker in frame with SSL2full (full-length enzyme), SSL2tr120 (SSL-2 with the carboxy-terminal 120 amino acids deleted), and SSL2tr87 (SSL-2 with the carboxy-terminal 87 amino acids deleted) with and without the Yct regions previously described. Using this approach, complementation was still not observed in ZX178-08 even when the low-level and high-level expression capability of the Gal1 promoter was exploited (**Figure 3.5C**). All fusion constructs grew on SCEglu, although the SSL1-SSL2tr87 fusion series appeared to be somewhat growth compromised. Much like the co-expressed SSL-1 and SSL-2 enzymes in ZX178-08, no growth was observed on SCEgal for any of the fusion constructs with the exception of SSL1-SSL2tr120.



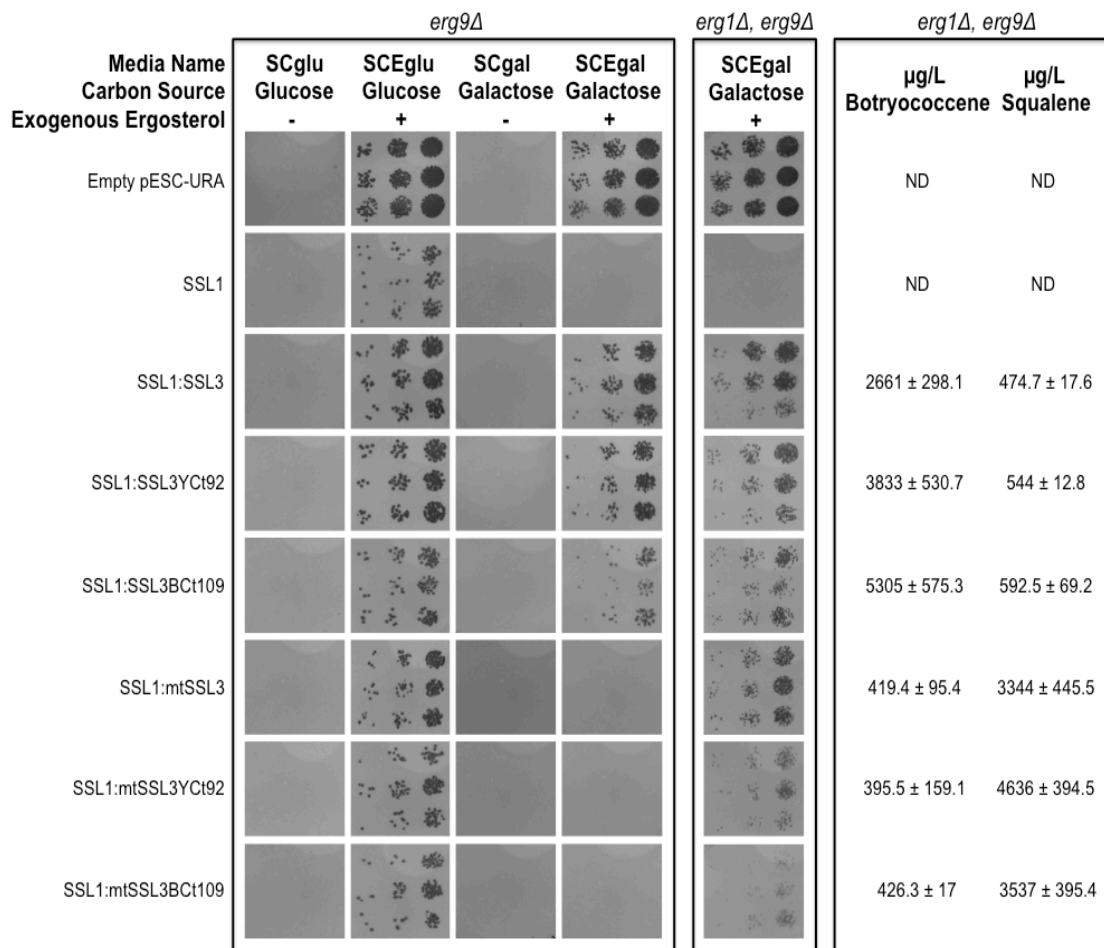
**Figure 3.5C** Complementation, growth, and chemical profile analysis of SSL-1 and SSL-2 gene-fusion constructs in ZX178-08 and ZXD yeast lines. ZX178-08 harbors a deletion of the *ERG9* gene while ZXD contains deletions of both *ERG9* and *ERG1* genes as denoted by the *erg9Δ* and *erg1Δ, erg9Δ* designation across the top of the figure frame, respectively. Three biological replicates (top to bottom within each panel) for each construct in ZX178-08 or ZXD were grown for three days in liquid SCEglu media. One mL of cell culture was pelleted, washed, and then diluted 50X, 250X, and 1250X. Five μL of each dilution (lowest to highest dilutions from right to left within each panel) for each biological replicate was spot plated on the four different media types as indicated at the top of each column. Pictures were taken after five days of growth at 28°C. Chemical profiles were generated from the same three ZXD biological replicates used for the SCEgal growth test. Cultures were grown in liquid SCEgal media for seven days at room temperature; hexane-extractable products were quantified *via* GC-FID in comparison to authentic standards.



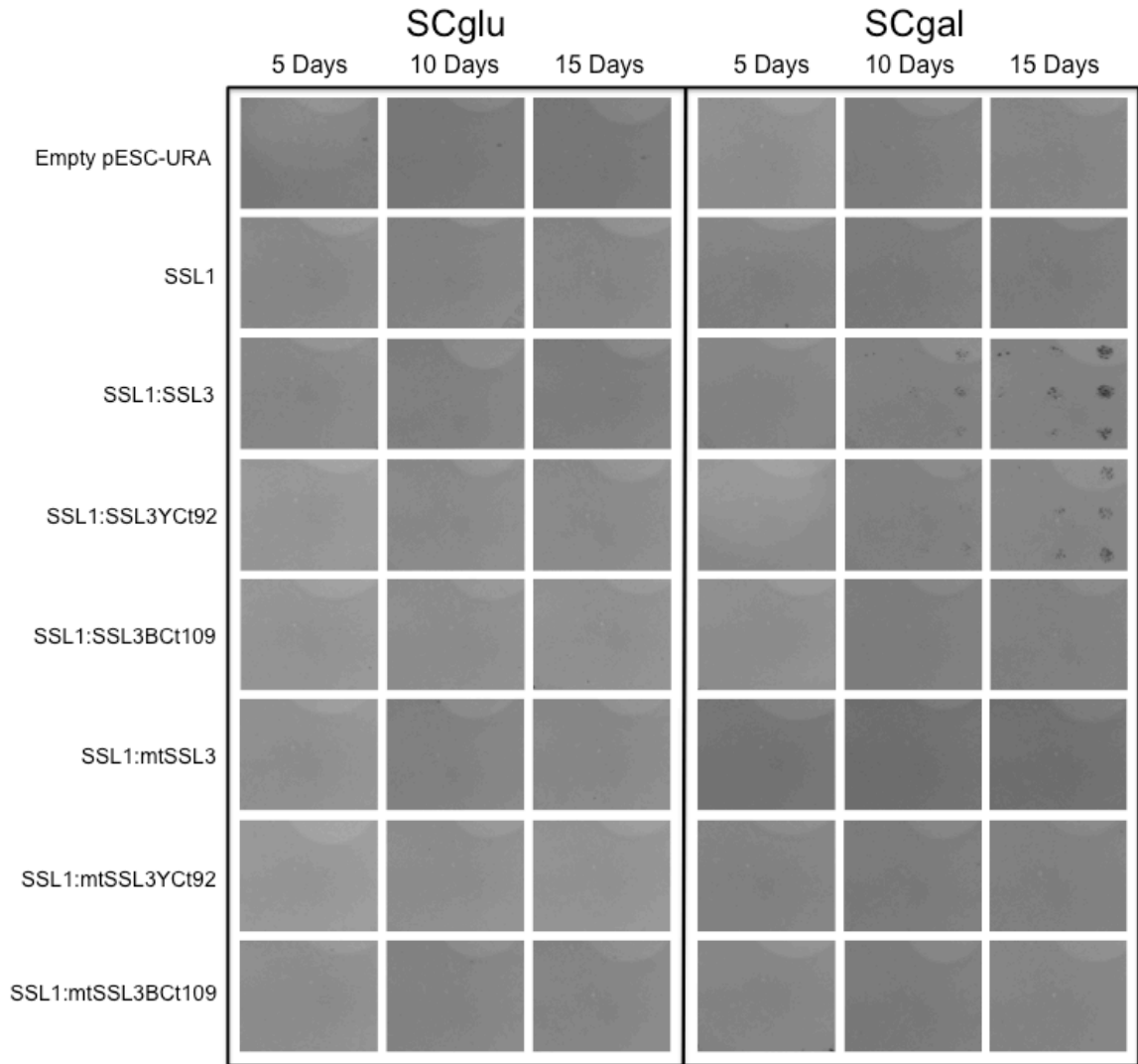
*Chemical profiling and growth assessment of SSL1-SSL2 fusion constructs in ZXD.* Similar to the results for co-expressed enzymes in ZXD on SCEgal media, SSL1-SSL2full exhibited growth. However, unlike their co-expressed counterparts in ZXD, SSL1-SSL2tr120, SSL1-SSL2tr120Yct27, and SSL1-SSL2tr87Yct65 fusion constructs were able to grow on SCEgal while SSL1-SSL2tr87 was not (**Figure 3.5C**). All other fusion constructs in ZXD were unable to grow on SCEgal like their co-expressed counterparts. Chemical profiles for the fusion constructs in ZXD showed a similar pattern as those for the co-expressed constructs, which can be seen in **Figure 3.5C**. No bFE was detected in any of the cultures expressing the SSL1-SSL2 fusion constructs.

*Complementation testing of SSL-1, SSL-3, and mtSSL-3 co-expressed constructs in ZX178-08.* SSL-3 predominantly produced botryococcene (~95%) and a small amount of squalene (~5%). mtSSL3 (N171A, G207Q) described in Chapter 2 was about ~70% efficient as SSL-3, but with ~90% of the reaction product as squalene. Given this and the lack of success demonstrating the ability of SSL-1 and SSL-2 to complement the *erg9* knockout, SSL-1 with either SSL-3 or mtSSL-3 was tested for their ability to supply squalene to the ergosterol pathway and thus complement the *erg9* deletion. SSL-3 and mtSSL-3 behave as functionally soluble enzymes because they do not possess a carboxy-terminal membrane-spanning region. Hence, no C-terminal truncations were tested for the SSL-3 or mtSSL-3 enzymes. Only Yct92 and BCt109 were appended to the C-terminus to test for their contribution to complementation of the *erg9* deletion. The promoters used to drive gene expression were the Gal1 promoter for SSL-1 (as in earlier studies), and the Gal10 promoter for SSL-3 and mtSSL-3 constructs.

As shown in **Figure 3.6A**, the pESC-URA vector without a transgene (empty vector control) expressed in ZX178-08 was unable to grow on SCglu and SCgal, but showed normal growth on media that contained ergosterol. ZX178-08 harboring the SSL-1 gene only grew on SCEglu as previously observed. SSL-1 and mtSSL3 constructs co-expressed in ZX178-08 only grew on SCEglu, again suggesting a toxic effect of either PSPP or a downstream product of squalene in the ergosterol biosynthetic pathway. SSL-1 and SSL-3 constructs co-expressed in ZX178-08 showed no growth on SCglu or SCgal but demonstrated normal growth on SCEglu and SCEgal at the five-day mark (**Figure 3.6A**). Interestingly, ZX178-08 that co-expressed SSL1:SSL3 or SSL1:SSL3Yct92 constructs showed growth after ten days on SCgal, which was more apparent at fifteen days (**Figure 3.6B**). While only a partial restoration of growth in comparison to yeast expressing ScSSfull on SCgal media (growth observable within three days), these results demonstrated the ability of SSL-1 and SSL-3 to complement the *erg9* knockout in *S. cerevisiae* with or without a carboxy-terminal, membrane-spanning domain (**Figure 3.6B**).



**Figure 3.6A** Complementation, growth, and chemical profile analysis of SSL-1, SSL-3, and mtSSL-3 co-expressed constructs in ZX178-08 and ZXD yeast lines. ZX178-08 harbors a deletion of the *ERG9* gene while ZXD contains deletions of both *ERG9* and *ERG1* genes as denoted by the *erg9Δ* and *erg1Δ, erg9Δ* designation across the top of the figure frame, respectively. Three biological replicates (top to bottom within each panel) for each construct in ZX178-08 or ZXD were grown for three days in liquid SCEglu media. One mL of cell culture was pelleted, washed, then diluted 50X, 250X, and 1250X. Five μL of each dilution (lowest to highest dilutions from right to left within each panel) for each biological replicate was spot plated on the four different media types as indicated at the top of each column. Pictures were taken after five days of growth at 28°C. Chemical profiles were generated from the same three ZXD biological replicates used for the SCEgal growth test. Cultures were grown in liquid SCEgal media for seven days at room temperature; hexane-extractable products were quantified via GC-FID in comparison to authentic standards.

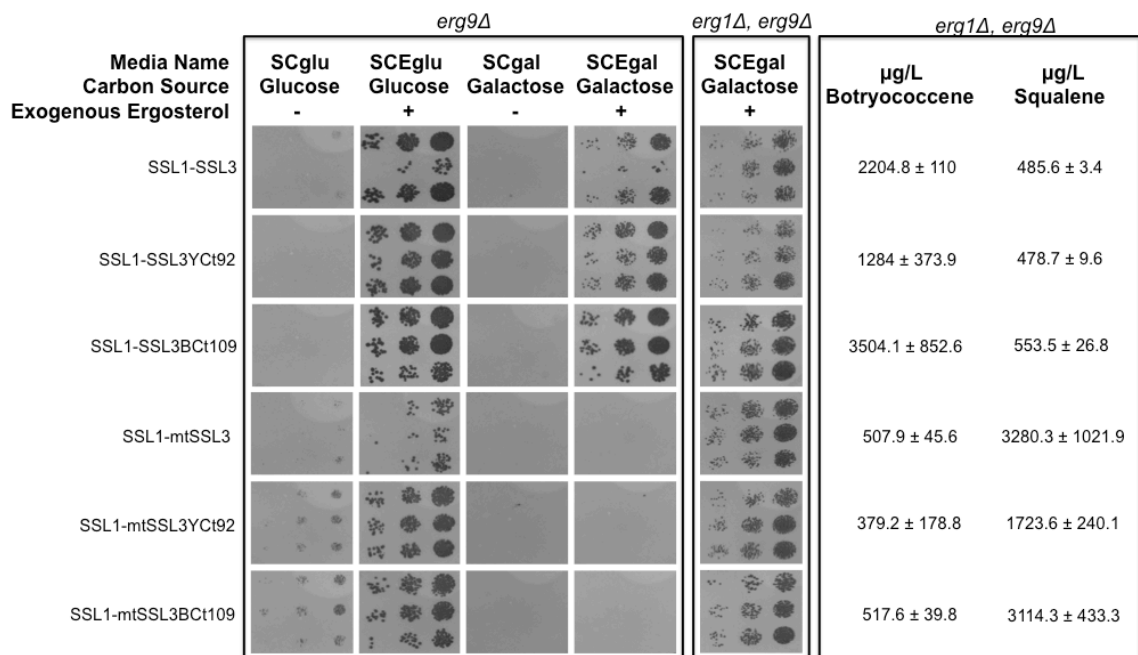


**Figure 3.6B** SSL-1, SSL-3 and mtSSL-3 co-expressed constructs after fifteen days of growth. Constructs were transformed into ZX178-08 (*erg9Δ*) and grown on SCglu and SCgal media for five days at 28°, then transferred to room temperature for ten more days.

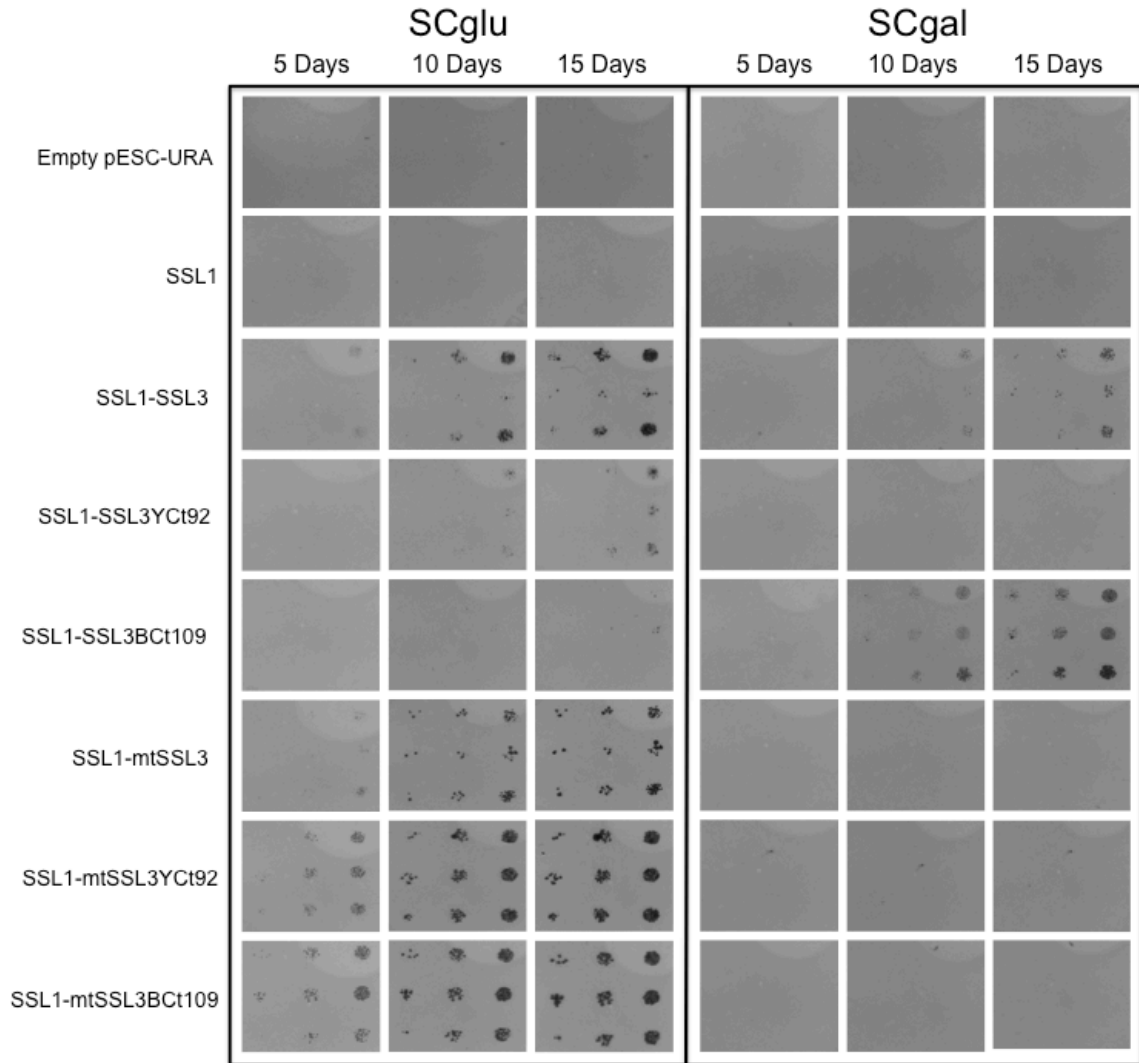
*Chemical profiling and growth assessment of SSL-1, SSL-3, and mtSSL-3 co-expressed constructs in ZXD.* SSL-1 overexpressed in ZXD, as already described, was unable to grow on SCEgal. However, when SSL-1 and SSL-3 were co-expressed, the yeast lines were able to grow on SCEgal. Similarly, all co-expressed SSL-1 and mtSSL-3 ZXD yeast lines demonstrated growth. Product profiling for ZXD lines that overexpressed empty vector and single-gene constructs, except for SSL-1, which does not grow in SCEgal, revealed no squalene or botryococcene accumulation (**Figure 3.6A**). ZXD that co-expressed SSL-1 and SSL-3 constructs showed all constructs to be functional based on botryococcene and squalene accumulation as shown in **Figure 3.6A**. SSL-1 and mtSSL3 overexpressed in ZXD also displayed enzymatic activity based on botryococcene and squalene accumulations as shown in **Figure 3.6A**.

*Complementation testing of SSL1-SSL3 and SSL1-mtSSL3 fusion constructs in ZX178-08.* Previous experiments with SSL-1 and SSL-2 demonstrated the inability of co-expressed and fusion enzymes with varying C-terminal tail arrangements expressed at low and high levels to complement the *erg9* knockout in ZX178-08. Moreover, SSL-1 and mtSSL-3 constructs that predominantly produce squalene were also unable to complement when co-expressed as separate enzymes with or without two different C-terminal tails. SSL-1 and SSL-3 however, when co-expressed with or without the carboxy-terminal tail from ScSS (YCt92), did display the ability to complement *erg9Δ* after ten days on SCgal media (high-level expression). In an attempt to improve complementation ability of SSL-1 and (mt)SSL3 constructs, fusion enzymes were again employed with varying C-terminal regions. As before, all fusion constructs were under control of the Gal1 promoter in pESC-URA, assessed on all four media types, and observed for growth after five days at 28°C, followed by ten days of growth at room temperature.

Unlike their co-expressed counterparts, SSL1-mtSSL3 fusion constructs were able to complement on SCglu regardless of C-terminal tail. SSL1-mtSSL3BCt109 complemented with the most vigorous growth, followed by SSL1-mtSSL3YCt92, and SSL1-mtSSL3 (**Figure 3.6C**). Complementation for these constructs was not observed when expressed at a high level (SCgal media). Growth on SCEglu was comparable to empty vector while no growth was observed on SCEgal for the SSL1-mtSSL3 constructs. After five days of growth, only the SSL1-SSL3 construct displayed signs of growth on SCglu media, which became more apparent at ten and fifteen days (**Figure 3.6D**). For ZX178-08 expressing SSL1-SSL3YCt92 at a low level (SCglu), growth was observed after ten days. No growth was seen when this construct was expressed at a high level (SCgal). Only the SSL1-SSL3BCt109 construct expressed at a high level in ZX178-08 complemented the *erg9* deletion with growth observed after ten days (**Figure 3.6D**). All ZX178-08 lines that expressed SSL1-SSL3 constructs grew normally on SCEglu and SCEgal.



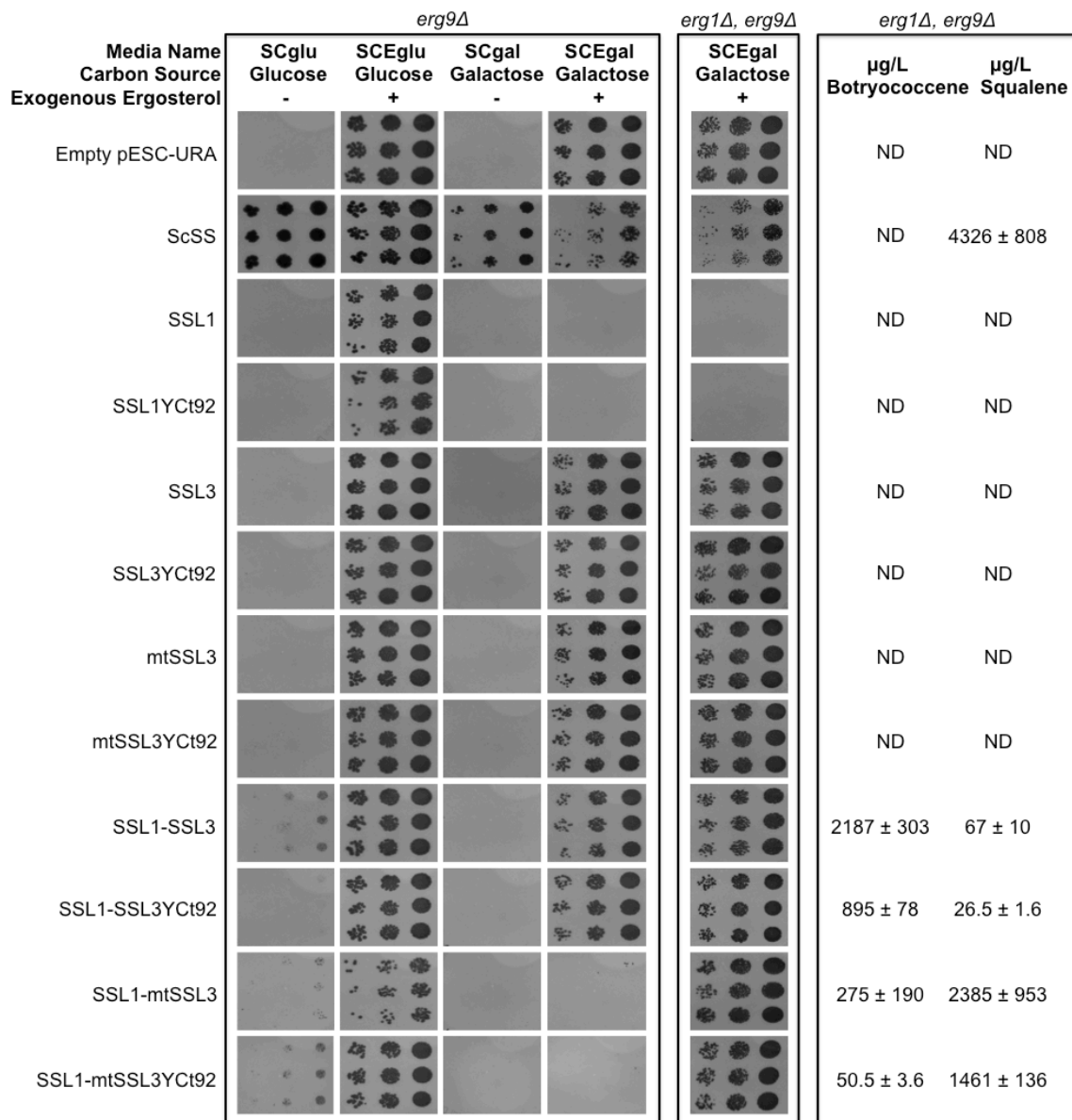
**Figure 3.6C** Complementation, growth, and chemical profile analysis of SSL1-SSL3 and SSL1-mtSSL3 fusion constructs in ZX178-08 and ZXD yeast lines. ZX178-08 harbors a deletion of the *ERG9* gene while ZXD contains deletions of both *ERG9* and *ERG1* genes as denoted by the *erg9Δ* and *erg1Δ, erg9Δ* designation across the top of the figure frame, respectively. Three biological replicates (top to bottom within each panel) for each construct in ZX178-08 or ZXD were grown for three days in liquid SCEglu media. One mL of cell culture was pelleted, washed, then diluted 50X, 250X, and 1250X. Five μL of each dilution (lowest to highest dilutions from right to left within each panel) for each biological replicate was spot plated on the four different media types as indicated at the top of each column. Pictures were taken after five days of growth at 28°C. Chemical profiles were generated from the same three ZXD biological replicates used for the SCEgal growth test. Cultures were grown in liquid SCEgal media for seven days at room temperature; hexane-extractable products were quantified *via* GC-FID in comparison to authentic standards.



**Figure 3.6D** SSL1-SSL3 and SSL1-mtSSL3 fusion constructs in ZX178-08 after fifteen days of growth. Constructs were transformed into ZX178-08 (*erg9Δ*) and grown on SCglu and SCgal media for five days at 28°, then transferred to room temperature for ten more days.

*Chemical profiling and growth assessment of SSL1-SSL3 and SSL1-mtSSL3 fusion constructs in ZXD.* All ZXD lines overexpressing SSL1-SSL3 and SSL1-mtSSL3 constructs were viable, and the introduced enzyme fusions appeared functional based on the chemical profiling of the respective cultures. These results are found in **Figure 3.6C**.

*Complementation and growth analysis of SSL-1, SSL-3, and mtSSL-3 single-gene and fusion constructs with normalized cell concentrations.* Because previous figures with yeast lines tested for complementation and growth were not normalized before spot plating, select lines with specific constructs were chosen for subsequent analyses with normalized cell concentrations to obtain a comparison for a given construct's effect on complementation and/or growth. Empty pESC-URA and ScSS full-length were chosen as negative and positive controls, respectively. SSL-1, SSL-3, and mtSSL-3 single-gene constructs with and without YCt92 were assessed along with SSL1-SSL3 and SSL1-mtSSL3 fusion constructs with and without YCt92. All the ZX178-08 lines were tested for complementation and growth on all four media types; ZXD lines were grown only on SCEgal. Chemical profiles were generated as previously described for ZXD lines. As illustrated in **Figure 3.7**, these results clearly demonstrated the ability of the fusion enzymes to complement the *erg9* knockout in ZX178, albeit with only a partial restoration of growth when compared to complementation of the same genetic background transformed with the heterologous full-length ScSS construct grown on SCglu media. Growth tests for ZXD lines on SCEgal media and chemical profiles were generated from the corresponding ZXD yeast lines for botryococcene and squalene accumulation, which can be seen in **Figure 3.7**. SSL-1, SSL-3, and mtSSL-3 single-gene constructs did not produce any detectable botryococcene or squalene while ScSS and the fusion constructs did demonstrating the expression of functional enzymes.



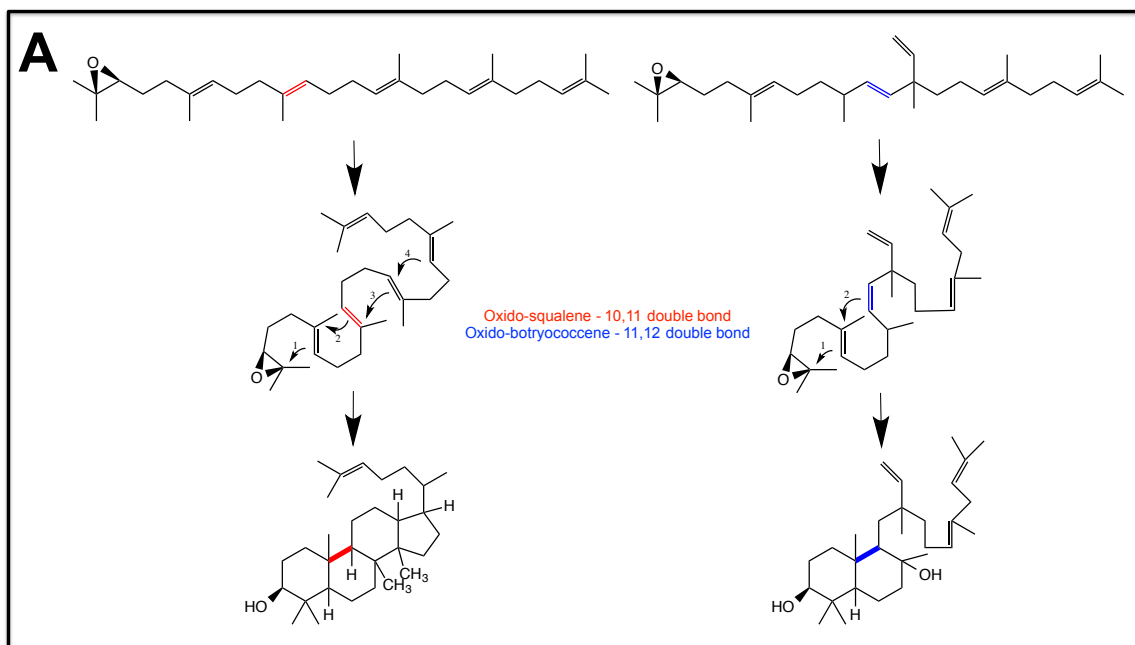
**Figure 3.7** Quantitative comparison of yeast lines ZX178-08 and ZXD transformed with the indicated constructs for their ability to complement the *erg9* deletion. ZX178-08 harbors a deletion of the *ERG9* gene while ZXD contains deletions of both *ERG9* and *ERG1* genes as denoted by the *erg9Δ* and *erg1Δ, erg9Δ* designation across the top of the figure frame, respectively. Biological replicates (top to bottom within each panel) were grown for three days in SCEglu liquid media, one mL of cells were pelleted, washed, and resuspended in sterile water, then used for OD<sub>600</sub> normalization to 0.1, 0.02, 0.004. Five μL of each dilution for each biological replicate was plated on all four media types for the ZX178-08 lines and SCEgal for the ZXD lines. Chemical profiles were generated from the same three ZXD biological replicates that were used for the SCEgal growth tests. Cultures used for chemical profiling were grown in liquid SCEgal media for seven days at room temperature; hexane-extractable products were quantified *via* GC-FID in comparison to authentic standards.



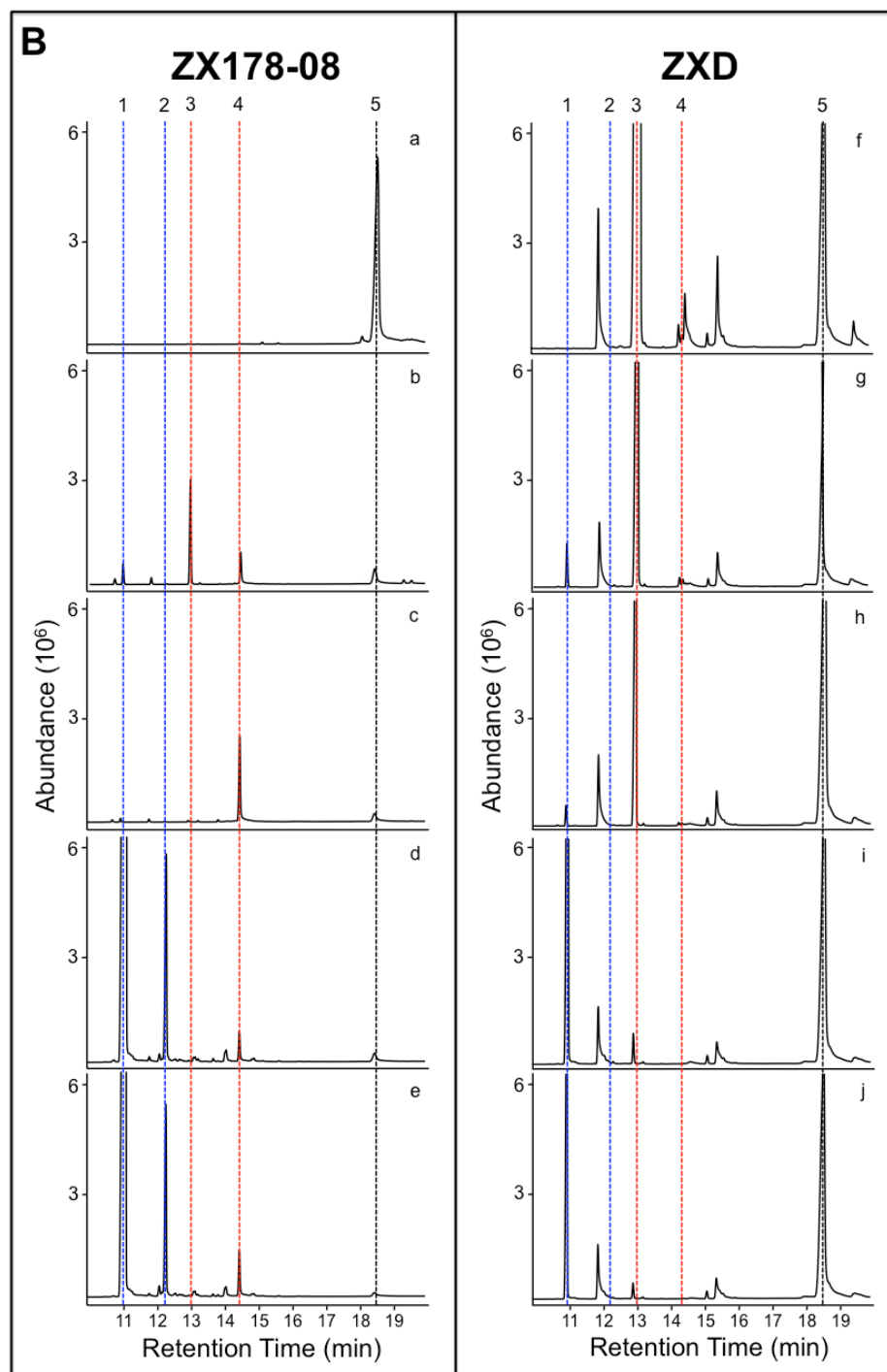
*Sterol profiling of the squalene synthase knockout lines complemented with mutant squalene synthase-like genes.* As noted earlier, the ability of SSL-3 to complement the *erg9* knockout despite only catalyzing 5% squalene formation and 95% botryococcene left open the possibility that botryococcene might serve as a surrogate for the biosynthesis of novel sterols (**Figure 3.8A**). To determine if squalene or botryococcene was leading to the production of bona fide ergosterol or alternative sterols, chemical profiles were examined for ZX178-08 expressing (SCglu media) ScSS, SSL1-SSL3, SSL1-SSL3YCt92, SSL1-mtSSL3, and SSL1-mtSSL3YCt92. Cells from starter cultures grown in SCEglu media were collected, washed with sterile water, and used to inoculate SCglu media which was grown for seven days at 28°C. Cells were then collected, washed, and lyophilized. Equal weights of freeze dried cells were saponified (methanolic KOH, 85°C), extracted with petroleum ether, concentrated, and analyzed by GC-MS in comparison to squalene, botryococcene and ergosterol authentic standards. No samples were derivatized although considerations were made to do so but ultimately abandoned given the excellent ability to detect ergosterol without derivatization (data not shown). As shown in **Figure 3.8B**, yeast line ZX178-08 harboring either SSL1-SSL3 or SSL1-mtSSL3 fusion enzymes with and without the YCt domain were able to channel squalene to ergosterol biosynthesis and accumulation. A comparison of the growth rates of ZX178-08 expressing these five constructs grown in SCglu media was also determined (**Figure 3.9C**), which showed an approximate 5-fold difference in growth at the seven day time point for the lines that complemented with the SSL1-(mt)SSL3 (except SSL1-mtSSL3, 52-fold difference) constructs and the native squalene synthase gene (ScSS). When compared to the quantified product profiles however (**Figure 3.9D**), these results were not consistent with the accumulation of ergosterol levels for the complementing ScSS (4950 ng ergosterol / mg dry cell weight) and fusion constructs: SSL1-SSL3(YCt92) (~93 ng ergosterol / mg dry cell weight; 53-fold difference) and SSL1-mtSSL3(YCt92) (~300 ng ergosterol / mg dry cell weight; 16.5-fold difference).

Close examination of the GC-MS chromatograms also revealed another very interesting correlation. That is, the site of squalene production, either targeted to the ER membrane *via* the YCt domain or defaulted to the cytoplasm for constructs without the YCt, influenced the conversion of squalene to its epoxide form, a necessary transformation for squalene to feed into the ergosterol biosynthetic pathway. This observation can be seen in the two additional peaks present in the chromatograph analysis. Inspection of the mass spectral fragmentation pattern revealed squalene (*m/z* 410) and botryococcene (*m/z* 410) like properties, but with a parent ion of *m/z* 426, suggesting the addition of an oxygen atom to the molecule. Comparison of these peaks to an oxidosqualene standard revealed the identity of one of the peaks, and by inference we tentatively identified the second unknown peak as oxidobotryococcene. The disappearance of these two peaks when the respective constructs were expressed in the ZXD yeast line that contained a knockout of the squalene epoxidase gene as well as squalene synthase corroborated the chemical

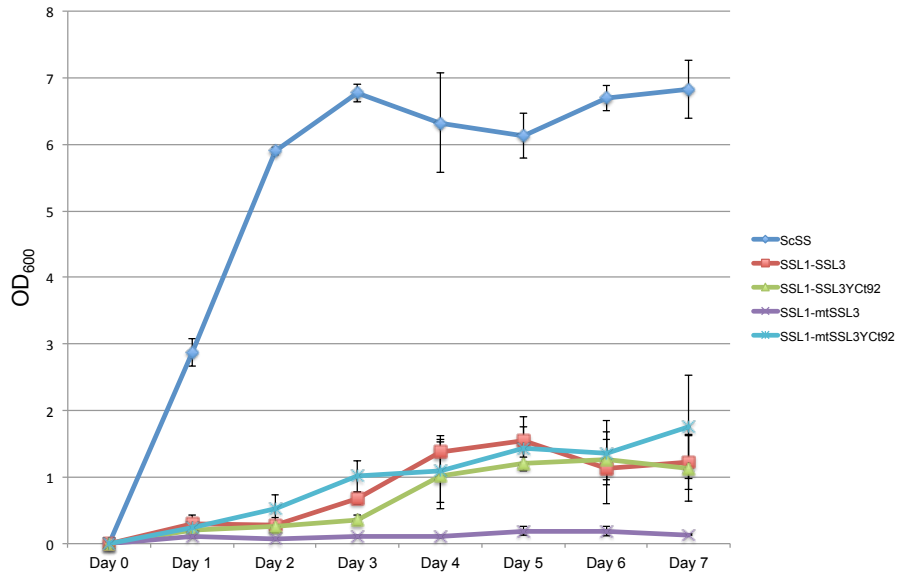
identification of these peaks as oxidosqualene and oxidobotryococcene (**Figure 3.8B**).



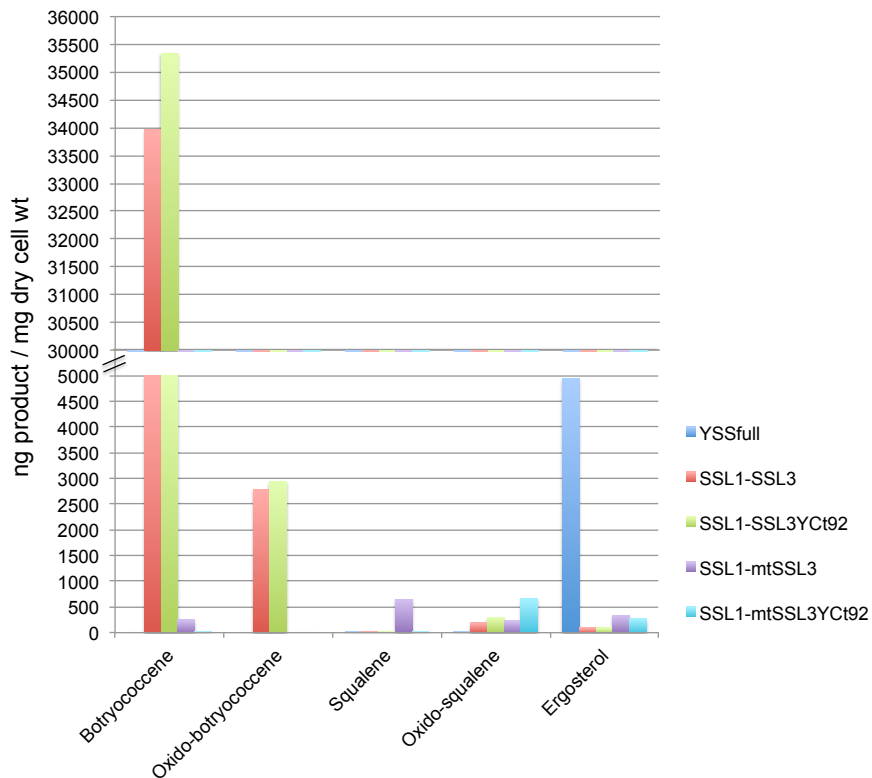
**Figure 3.8A** Proposed cyclization of botryococcene after epoxidation by Erg1p. Only two rings are likely to be formed from oxidobotryococcene; the additional OH in the cyclized botryococcene might originate from water quenching of the tertiary carbocation in that position to terminate the reaction.



**Figure 3.8B** GC-MS sterol profiling for ZX178-08 and ZXD with complementing constructs. ZX178-08 (a-e) and ZXD (f-j) yeast lines expressing ScSS (a,f), SSL1-mtSSL3 (b,g), SSL1-mtSSL3Yct92 (c,h), SSL1-SSL3 (d,i), and SSL1-SSL3Yct92 (e,j). ZX178-08 cultures were grown in SCglu media (low-level expression) while ZXD cultures were grown in SCEgal (high-level expression). Peak numbers correspond to botryococcene (1), putative oxidobotryococcene (2), squalene (3), oxidosqualene (4), and ergosterol (5).



**Figure 3.8C** Comparison of growth rates for the ZX178-08 yeast line expressing ScSS, SSL1-SSL3, SSL1-SSL3Yct92, SSL1-mtSSL3, and SSL1-mtSSL3Yct92 in SCglu media. Data points are representative of three biological replicates from one independent experiment.



**Figure 3.8D** Quantified GC-MS results for sterol profiling of the ZX178-08 (*erg9Δ*) yeast line complemented with ScSS, SSL1-SSL3, SSL1-SSL3Yct92, SSL1-mtSSL3, and SSL1-mtSSL3Yct92. Note the y-axis break to emphasize lower levels of certain products.

The importance of oxidosqualene formation for the complementation phenotype was apparent when the levels of this intermediate were compared between lines where squalene synthase activity was targeted and not targeted to the ER membrane (**Figure 3.8B**, panels a-e). First, when a full-length ScSS targeted to the ER membrane was used for complementation, very little if any squalene or squalene epoxide was evident, yet ergosterol levels were very high (panel a). The growth rate of this line was also very robust (**Figure 3.8C**). In contrast, significant levels of squalene and its epoxide form were apparent in yeast complemented with the SSL1-mtSSL3 gene not targeted to the membrane (panel b). When this same enzyme was targeted to the ER membrane *via* the YCt targeting domain (SSL1-mtSSL3YCt92), no squalene accumulated, yet greater levels of the epoxide were evident (compare panels b and c). The growth rate of the yeast line harboring the non-targeted SSL1-mtSSL3 enzyme also exhibited the slowest growth rate compared to the other lines, approximately one-tenth of SSL1-mtSSL3YCt, which was targeted to the ER membrane. Equally striking was when the levels of these intermediates were considered in relation to the growth rates of the respective yeast lines with botryococcene synthase (SSL1-SSL3). When the SSL1-SSL3 gene was used to complement the *erg9* knockout, no squalene accumulated regardless of whether or not the encoded enzyme was targeted to the ER membrane (compare panels d and e). However, when the SSL1-SSL3 enzyme was directed to the ER membrane, slightly higher levels of the epoxide were observed which was even more evident when the SSL1-mtSSL3YCt complementing line was compared to the SSL1-mtSSL3 in ZX178-08. Both of the SSL1-SSL3 lines grew similarly to yeast complemented with the SSL1-mtSSL3YCt and certainly exhibited more vigorous growth than the yeast line harboring the SSL1-mtSSL3 construct not targeted to the ER membrane, which accumulated significantly more squalene than any of the other lines. These results suggested that while a soluble form of squalene synthase enzyme activity can complement the *erg9* knockout, cellular growth rates are negatively impacted by either the burden of excess squalene biosynthesis. Alternatively, the effects observed might be an effect of the inability to provide the ergosterol biosynthetic pathway with enough squalene, which further indicates the importance of controlling carbon flux into this pathway for optimal cellular performance.

*Library construction and validation.* With the demonstration that SSL1-mtSSL3Yct92 could complement the *erg9* knockout in *S. cerevisiae*, a mutant enzyme library was constructed to validate the idea that ZX178-08 could be used as a complementation screen to identify mutants with altered catalytic activities. As discussed in Chapter 2, two positions in SSL-3 were identified that contribute to product specificity for either botryococcene or squalene. The wild-type enzyme possesses an asparagine at position 171 (N171) and a glycine at position 207 (G207) and catalyzes the biosynthesis of botryococcene with ~95% specificity. A SSL-3 mutant harboring an asparagine to alanine mutation at position 171 and a glycine to glutamine mutation at position 207 exhibited ~90% specificity for the biosynthesis of squalene. Thus, if the complementation screen is possible, the two mutations that convert product specificity in SSL-3 (N171A, G207Q) should be isolated from the mutant library. To assess all possible amino acid combinations in these two positions, both sites were mutated to saturation (N171X, G207X) for a total of  $20^2 = 400$  possible combinations. Mutations were introduced by employing sequence overlap extension (SOE) PCR with oligonucleotides that contained the NNS codon in positions 171 and 207 of SSL-3. N171X was introduced first followed by G207X and the final reassembled PCR products were inserted into an intermediate cloning vector (**Figure 3.9A**).

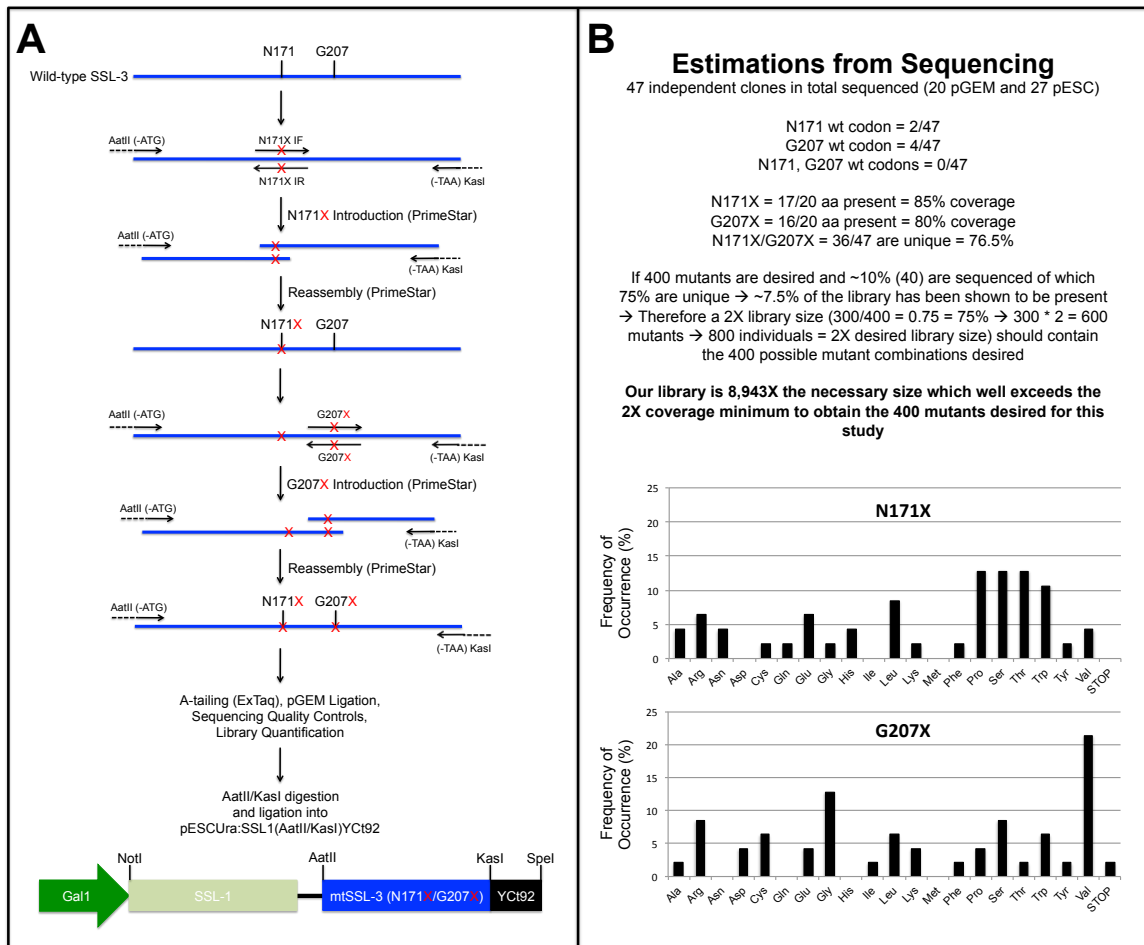
To assess the randomness and biases of the SOE PCR technique, 20 independent clones were chosen randomly for sequencing. Eleven different codons were found in both positions 171 and 207. The mtSSL-3 sequences were then released in bulk from the intermediate cloning vector by restriction digestion and ligated into the corresponding restriction sites in-frame and in between the SSL-1 gene and the yeast carboxy-terminal sequence, Yct92, in the yeast expression vector pESC-URA. The recombinant vector reconstituted a fusion gene configured as SSL1-mtSSL3-Yct92 encoding for a multifunctional presqualene diphosphate (PSPP) synthase followed by a triterpene synthase capable of using PSPP for the biosynthesis of variable triterpene products (botryococcene, squalene, and other unknowns) and harboring a carboxy-terminal domain that targeted the enzyme complex to the endoplasmic reticulum. Expression of the chimeric gene was placed under the control of the Gal1 promoter. Twenty-seven independent clones of the assembled constructs were sequenced; all were found to be properly assembled, in-frame gene constructions. Codons for 16 different amino acids were found at position 171 and codons for 14 different amino acids were found at position 207. Of the total 47 independent clones sequenced from both libraries, 17 of 20 amino acids were encoded for at position 171 and 16 of 20 amino acids were coded for in position 207 (**Figure 3.9B**). To ensure 100% library coverage (all possible combinations =  $20^2$ ), it was estimated based on sequencing results that a 2X library size (800 members) would ensure complete coverage (400 total mutants). From library size estimations based on CFU/cm<sup>2</sup>, the constructed library was calculated to be 8,943X the necessary size which well exceeded the desired 2X coverage.

*Library screening in ZX178-08 for *erg9*Δ complementation.* ZX178-08 yeast cells were transformed with the pESC-URA:SSL1-mtSSL3-Yct92 (N171X, G207X) mutant enzyme library and directly selected on SCglu and SCgal media

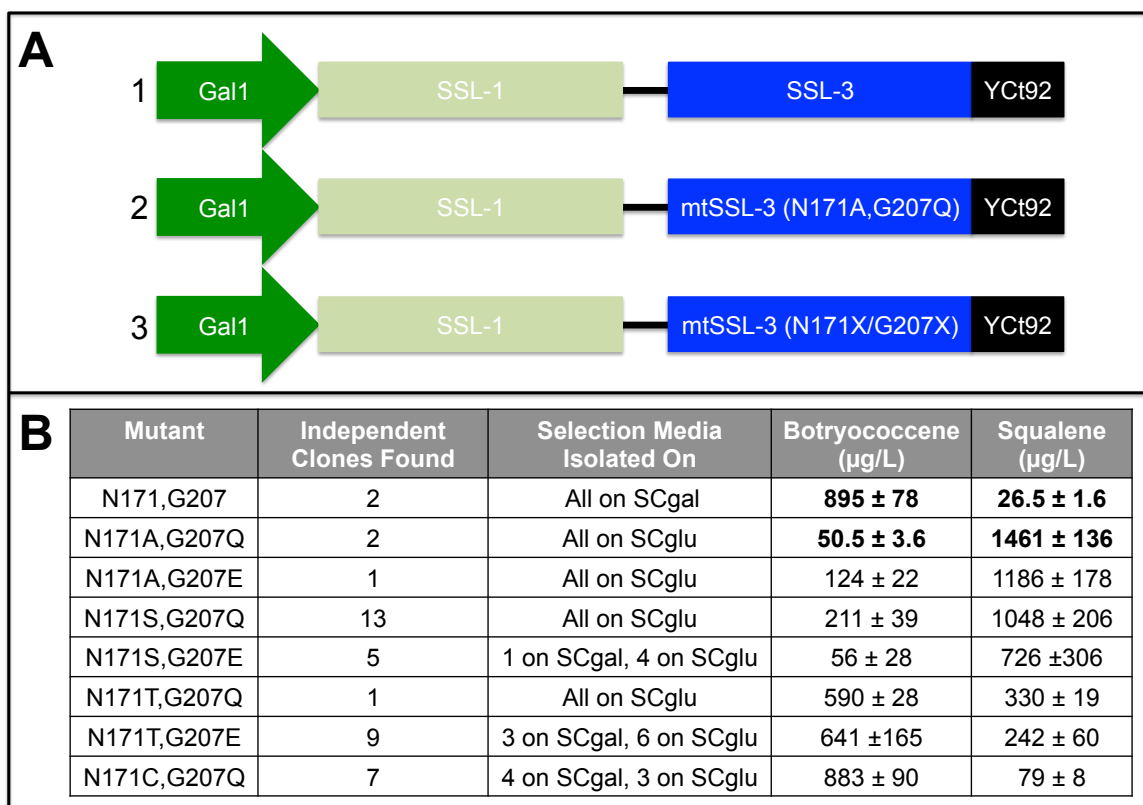
for mutants able to complement *erg9Δ*. By selecting on low- and high-level expression media, mutants with low squalene synthase activity would be isolated on SCgal media and mutants more efficient at making squalene would be found on SCglu. If the selection worked as envisioned, recovery of the SSL1-SSL3 wild-type and original SSL1-mtSSL3 mutant (N171A, G207Q) sequences were expected (**Figure 3.10A**). Many colonies arose on these plates at various times after plating. Some colonies were observed five days after plating, while others were only visible fifteen days after plating. Additionally, although not quantified, some colonies seemed to grow more robustly only after the plates were transferred to room temperature from 28°C, which suggested temperature sensitivity in either the mutant enzymes or the ZX178-08 yeast line itself. At ten and twenty days, colonies were restreaked from the original plates onto fresh selection plates. An independent colony was picked from these plates for growth in liquid selection media followed by plasmid DNA extraction and transformation into *E. coli*. Single bacterial colonies were chosen, grown in selection media, mini-prepped, and sequenced. Independent colonies failing to grow at any one of these steps were discarded from subsequent analyses.

Ultimately, forty independent colonies that arose on the original selection plates were sequenced. Two colonies selected on SCgal that were isolated were found to contain the wild-type protein sequence for SSL-3 and two colonies selected on SCglu contained the N171A, G207Q mutations. Other mutations were also identified at both positions 171 and 207 which include: A171, E207; S171, Q207; S171, E207; T171, Q207; T171, E207; and C171, Q207. **Figure 3.10B** provides a breakdown of the number of independent colonies for each mutant isolated, selection media mutants were isolated on, and product profile results. Confirmation that each clone sequenced was not a false positive and that both SSL-1 and the mtSSL-3 sequence were required for complementation, complementation and growth testing as well as chemical profiling were performed for each mtSSL-3 constructs as well as the entire fusion construct in ZX178-08 and ZXD as described earlier. Shown in **Figure 3.10C**, these results further verified the authenticity of the mutants isolated from the selection as well as provided information regarding their chemical profiles.





**Figure 3.9** Construction of the mutant enzyme library to test the ZX178-08 selection methodology. Mutants were constructed by SOE PCR to incorporate mutagenic oligonucleotides at positions 171 and 207 of SSL-3 (A). During construction, independent mutants were randomly selected from the pGEM and pESC-URA libraries for sequencing to assess the fidelity of regions outside targeted mutagenic positions and the introduction of mutations in the mtSSL-3 sequences. The library size was estimated from bacterial CFU (B) before preparation of the plasmid DNA.



**Figure 3.10** Mutant library selection in ZX178-08. Sequences already shown to complement  $\Delta\text{erg9}$  (1 and 2) that were expected to be isolated if selection works. Additional mutants (3) were hypothesized to be isolated if the mutant library and selection platform were robust enough (A). Results from selection of the mutant library included six new mutants previously uncharacterized, number of independent clones isolated for each mutant, selection media mutants were isolated on, and their chemical productivities for botryococcene and squalene (B).

Media Name Carbon Source Exogenous Ergosterol	<i>erg9Δ</i>				<i>erg1Δ, erg9Δ</i>	<i>erg1Δ, erg9Δ</i>	
	SCglu Glucose	SCEglu Glucose	SCgal Galactose	SCEgal Galactose	SCEgal Galactose	μg/L Botryococcene	μg/L Squalene
	-	+	-	+	+		
Empty pESC-URA						ND	ND
SSL1-mtSSL3 N171S, G207Q						211 ± 39	1048 ± 206
SSL1-mtSSL3 N171S, G207E						56 ± 28	726 ± 306
SSL1-mtSSL3 N171T, G207Q						590 ± 28	330 ± 19
SSL1-mtSSL3 N171T, G207E						641 ± 165	242 ± 60
SSL1-mtSSL3 N171C, G207Q						883 ± 90	79 ± 8
SSL1-mtSSL3 N171A, G207E						124 ± 22	1186 ± 178
mtSSL3 N171S, G207Q						ND	ND
mtSSL3 N171S, G207E						ND	ND
mtSSL3 N171T, G207Q						ND	ND
mtSSL3 N171T, G207E						ND	ND
mtSSL3 N171C, G207Q						ND	ND
mtSSL3 N171A, G207E						ND	ND

**Figure 3.10C** Complementation, growth, and chemical profile analyses for mutants isolated from ZX178-08 selection of mutant library. OD<sub>600</sub> normalized cells were used for complementation and growth testing in ZX178-08 and ZXD. Three biological replicates (top to bottom within one panel) were grown for three days in SCEglu liquid media, one mL of cells were pelleted, washed, and resuspended in sterile water then used for OD<sub>600</sub> normalization to 0.1, 0.02, 0.004. Five μL of each dilution for each biological replicate was plated on all four media types for ZX178-08 lines and SCEgal for ZXD lines. Chemical profiles were generated from the same three ZXD biological replicates that were used for the SCEgal growth tests. Cultures were grown in liquid SCEgal media for seven days at room temperature; hexane-extractable products were quantified *via* GC-FID in comparison to authentic standards.

## Discussion

Squalene synthases are critically important in sterol metabolism for all eukaryotic organisms but have been difficult enzymes to study. Initially, technical challenges were faced because these enzymes are membrane-bound making enzyme kinetic studies intractable due to difficulties in protein purification. Recombinant DNA technologies have helped clear this hurdle by allowing removal of the C-terminal membrane-spanning region from the open reading frame. The truncated form was then overexpressed in bacteria, a technique commonly employed nowadays. Since then, very detailed kinetic assessments have been achieved for the *S. cerevisiae* squalene synthase while crystallographic studies have revealed the overall structure of the human squalene synthase in addition to substrate and inhibitor binding properties. Structure-function studies of these enzymes are still complicated because of the two-step reaction mechanism for squalene that occurs all within one active site of one enzyme. The squalene synthase-like triterpene synthases from *B. braunii* offer an advantage in studying this mechanism because they catalyze one reaction step or the other for squalene and botryococcene catalysis. A rationally designed mutant study of these enzymes (Chapter 2) revealed some structure-function relationships; however, more complete and unbiased assessments are desirable. Thus, a more robust approach for studying the SSL enzymes of *B. braunii* was developed by employing mutant enzyme libraries and a selection for the evolution of squalene biosynthesis. The initial proof-of-principle established the ability of SSL fusion enzymes capable of squalene biosynthesis to complement the *erg9* knockout in ZX178-08. Once this was accomplished, a second proof-of-principle was demonstrated where a mutation already known to convert SSL-3 product specificity from botryococcene to squalene was isolated from a mutant library screen.

Several unexpected observations were made throughout the course of this work that concerned not only the original selection platform devised for SSL enzyme directed evolution, but also the ability of non-fungal squalene synthases to complement the *erg9* knockout in *S. cerevisiae*. For instance, during the establishment of positive and negative controls for this work, a non-fungal squalene synthase (the algal BbSS) was shown to complement the *erg9* knockout without the carboxy-terminal region from a fungal squalene synthase. This was unexpected because a large body of work has been established in this area suggesting otherwise. Previous work showed that when yeast lines with an *erg9* knockout were used to overexpress a non-fungal squalene synthase, the cells were unable to grow in the absence of ergosterol. The cells overexpressing these non-fungal squalene synthases were neither assessed for growth in the presence of ergosterol nor were their catalytic activities determined or if they had the ability to produce squalene *in vivo*. Other key differences between these previous studies and those conducted here include the promoters used for gene expression. Earlier studies all used strong, constitutive promoters (PMA1, ADH1), while we used a strong, inducible promoter (GAL1). With this promoter, we were able to exploit leaky expression in non-inducing conditions that allowed

for “low-level” gene expression as well as “high-level” gene expression when the yeast were grown in the presence of galactose.

Using this promoter system, both algal (BbSS) and yeast (ScSS) full-length squalene synthases were able to complement the *erg9* knockout in ZX178-08 under low-level expression conditions. However, when assessed under high-level expression conditions (galactose media), only the ScSS gene was able to complement. Moreover, when BbSS was expressed at a high level in the presence of exogenous ergosterol, no growth was observed. Taken together, these results suggested the production of a toxic metabolite downstream from oxidosqualene. The dependence of BbSS and ScSS on the presence of a C-terminal region for complementation was demonstrated by testing truncated enzymes, which actively produced squalene but were unable to complement. Chimeric BbSS and ScSS enzymes where their C-terminal regions were interchanged resulted in a reversal of the phenotypes observed for BbSS and ScSS. Chimeric BbSS behaved as ScSS did which grew on all four media types, while chimeric ScSS mirrored BbSS behavior only growing on glucose media. One possible explanation for these observations is that either the amount of squalene synthase protein or squalene synthase activity tightly regulates squalene levels in *S. cerevisiae*, which has important implications for the production of a toxic intermediate or shunt product. Such controls are likely functioning through a sequence specific region present only in the fungal squalene synthase, perhaps the linker region as defined by Niehaus (2011).<sup>84</sup> Thus, if this interpretation is correct, other non-fungal squalene synthases in their wild-type forms should be able to complement the *erg9* knockout if expressed at a low level in *S. cerevisiae*.

Another intriguing observation made at the outset of this work was the inability of ZX178-08 to grow in the presence of exogenous ergosterol when SSL-1 was overexpressed. Because ZX178-08 produces high levels of the substrate FPP, it was presumed that when SSL-1 was overexpressed much of this FPP is converted to PSPP. Only when SSL-1 was co-expressed with SSL-3 in ZX178-08 were the cells able to survive, implicating PSPP in the yeast cell death. Niehaus *et al.* (2011) showed the presence of presqualene alcohol (PSOH, dephosphorylated PSPP) when SSL-1 was overexpressed in a similar yeast line. We were unable to obtain an analogous chemical profile in the current studies because of the complete lack of growth for these yeast lines, although we could detect PSOH in samples where SSL-1 and SSL-3, mtSSL3, or SSL2 constructs were co-expressed. This suggests that PSOH by itself might not be cytotoxic, at least when produced within a certain threshold.

In this work, we were able to demonstrate specific combinations of the SSL enzymes that could complement the *erg9* knockout in ZX178-08. Although SSL-1 (FPP to PSPP synthase) and SSL-2 (PSPP to squalene synthase) seemed to be incapable of complementing the *erg9* knockout, we were able to show several arrangements of SSL-1 and SSL-3 (PSPP to >90% botryococcene, ~10% squalene) or mtSSL-3 (PSPP to <10% botryococcene, ~90% squalene) could restore ergosterol prototrophy, albeit less efficient than ScSS. The lack of growth for the ZXD lines expressing SSL-1 and SSL-2 constructs suggested the

production of a toxic intermediate made either by SSL-1 or the unique combination of SSL-1 and SSL-2. Alternatively, this effect could be the result of too much squalene being made by SSL-1 and SSL-2 that is fed into the ergosterol pathway to produce a different toxic metabolite or finally, it could be a combination of both of these possibilities. SSL-1 and SSL-3 setups that showed successful complementation were somewhat unexpected, but not entirely surprising as these enzymes are known to produce small amounts of squalene. More surprising was ability of the SSL-1 and (mt)SSL-3 enzyme arrangements to complement regardless of C-terminal tail which suggested that SSL-3 may have some ability to associate itself with the ergosterol biosynthetic machinery, perhaps a vestige of its supposed evolution from BbSS. SSL-3 C-terminal truncations tested using the complementation system described in this work could help resolve this issue.

This work provided new mechanistic insight into the controlled regulation of squalene and its enzyme, squalene synthase in fungal systems, which might extend to other eukaryotic organisms. Most importantly however, the intention of this work was to develop a directed evolution selection platform for the SSL enzymes. In general, this was accomplished, but with an important caveat. The range of squalene levels that can be channeled into sterols and viably tolerated is important – too little squalene and cells either die or grow so slow that growth is not apparent; too much squalene and a toxic intermediate/shunt product occurs that leads to cell death. Although this could limit the robustness of this approach to successfully evolve the SSL enzymes, using the Gal promoter system can help circumvent this downfall as evidenced by the new mutant enzymes that were isolated from the two-position mutant library using the ZX178-08 screen. In position 171 of SSL-3, only small polar amino acids cysteine, serine, and threonine substitutions were isolated, while only glutamine or glutamate were found in position 207 of SSL-3. Important to note here is that significance should not be attributed to the number of independent clones identified for any one mutant because this may be a reflection of codon bias for those particular mutants in the library. Nonetheless, the diversity of amino acids found in position 171 possibly suggest the importance of this position in shaping of the active site surface to favor squalene biosynthesis. Identification of only glutamine or glutamate substitutions in position 207 suggests a possible role of the carbonyl functionality in these two amino acid R-groups, as well as the proper length of the R-group. In Chapter 2, the Q213N mutation was observed to have substantially lower activity. The recent report by Liu *et al.* (2014) corroborates this observation by a Q212N mutation in the human squalene synthase. Additionally, Liu *et al.* (2014) documented HsSS Q212E to retain 85% activity. However, these interpretations are not sufficient to explain all the current observations. Based on the above rationale, the N171C, G207Q mutant would be expected to favor squalene production over botryococcene, as does the SSL-3 N171A, G207Q mutant. However, this mutation still favors product specificity for botryococcene.

The next step in this work will be utilization of the selection platform developed here to directly evolve single SSL enzymes towards fully functional

squalene synthases. Any number of methods can be used for construction of the mutant libraries; these include but are not limited to structure guided, phylogenetic based, insertion-deletion (indel), or random mutagenesis (see Goldsmith and Tawfik (2012) for review).<sup>93</sup> This stage will likely be the most critical for the successful use of the selection platform, however with careful thought, consideration, and enough quality control checks, successful directed evolution of the SSL enzymes will be achieved. If future studies of mutant SSL enzyme libraries are successful, other squalene synthase-like enzymes such as dehydrosqualene synthase and phytoene synthase could also be studied for their ability to be evolved towards fully functional squalene synthases.

## Materials and Methods

*Plasmids and yeast strains.* All constructs used for this work were cloned into this pESC-URA vector (Agilent). BbSS, ScSS, SSL-1, and all fusion constructs were cloned into the MCS1 NotI and SpeI sites. SSL-2 constructs were cloned into the MCS2 BamHI and XhoI sites while SSL-3 constructs were cloned into the Sall and KpnI sites. Chimeric and fusion constructs were assembled by sequence overlap extension (SOE) PCR. As an example, BbSSStr109Yct92 was constructed as follows: plasmid DNA templates containing the full-length BbSS and ScSS coding sequences provided generously by Niehaus *et al.* (2011)<sup>84</sup> were used to amplify the BbSSStr109 section with 5' NotI BbSS and 3' BbSSStr109-Yct92 IR primers with BbSSfull as template in one reaction while the Yct92 section was amplified with 5' BbSSStr109-Yct92 IF and 3' SpeI ScSSfull primers with ScSSfull as template in a separate reaction. Appropriately sized PCR products were gel purified then used as template in a second reaction with 5' NotI BbSS and 3' SpeI ScSSfull primers to assemble the entire chimeric BbSSStr109Yct92 construct with 5' NotI and 3' SpeI sites. This product was gel purified, digested with NotI/SpeI (NEB), and ligated into NotI/SpeI digested pESC-URA vector. All oligonucleotides used for the construction of genes in this work can be found in **Table 3.1**. All constructs were sequence verified before use. ZX178-08 and ZXD yeast lines were provided courtesy of Dr. Xun Zhuang; description of their derivations and genetic backgrounds can be found in Zhuang (2013).<sup>89</sup>



**Table 3.1** Primers used for cloning and screening yeast constructs.

Primer	5' → 3' Sequence
5' NotI BbSS	AAGGAAAAAGCGGCCGCAAACAATGGGGATGC TTCGCTGGGGAGTGG
3' SpeI BbSSfull	GGACTAGTTTAGGCGCTGAGTGTGGGTCTAGGTG AGG
3' SpeI BbSSStr109	GGACTAGTTCAAAGCTTCTCTGCGAAGTTGAGG
5' BbSSStr-YCt92 IF	CGTCATTTCTCAACTTCGCAGAGAAGCTTAAATC TAAATTGGCTGTGCAAGATCC
3' BbSSStr-YCt92 IR	GGATCTTGACAGCCAATTTAGATTTAAGCTTCTC TGCGAAGTTGAGGAAATGACG
3' XhoI BbSSfull	CCGCTCGAGTTAGGCGCTGAGTGTGGGTCTAGG
3' KpnI BbSSfull	GGGGTACCTTAGGCGCTGAGTGTGGGTCTAGGTG AGG
5' NotI ScSS	AAGGAAAAAGCGGCCGCAAACAATGGGAAAGC TATTACAATTGGCATTGC
3' SpeI ScSSfull	GGACTAGTTCACGCTCTGTGTAAAGTGTATATATA ATAAAACC
3' SpeI ScSSStr92	GGACTAGTTCAGATATCACGTAAGTAATAGTCAAA AATC
5' ScSSStr92-BCt109 IF	GACTACTTACGTGATATCGAAGTCAGATGCAA CACCGAGAC
3' ScSSStr92-BCt109 IR	GTCTCGGTGTTGCATCTGACTTCGATATCACGTAA GTAATAGTC
3' XhoI ScSSfull	CCGCTCGAGTCACGCTCTGTGTAAAGTGTATATAT AATAAAACC
3' KpnI ScSSfull	GGGGTACCTCACGCTCTGTGTAAAGTGTATATATA ATAAAACC
5' NotI SSL-1	AAGGAAAAAGCGGCCGCAAACAATGACTATGC ACCAAGACCACGGAG
3' SpeI SSL-1	GGACTAGTTCACTTGGTGGGAGTTGGGGCTGCG
5' SSL1-YCt92 IF	CTTTCTGCGCAGCCCCAACTCCCACCAAGAAATCT AAATTGGCTGTGCAAGATCC
3' SSL1-YCt92 IR	GGATCTTGACAGCCAATTTAGATTTCTTGGTGGG AGTTGGGGCTGCGCAGAAAG
3' SSL1-Linker	ACCAGAACCACCACCAGAACCACCACCAGAACCA CCCTTGGTGGGAGTTGGGGCTGCGC
5' SSL1-AatII-KasI- YCt92 IF	GGTGGTTCTGGTGACGTCGGCGCCAAATCTAAAT TGGCTGTGCAAGATCC
3' SSL1-AatII-KasI- YCt92 IR	CCAATTTAGATTTGGCGCCGACGTCACCAGAACC ACCACCAGAACCACCACC
5' BamHI SSL-2	CGGGATCCAAAACAATGGTGAAACTCGTCGAGGT TTTGC

**Table 3.1** (continued)

3' XhoI SSL2full	CCGCTCGAGCTACTGCTTGAAGAAGCAGAGGTG AGC
3' XhoI SSL2tr120	CCGCTCGAGCTACAACCTCGTTCGCGAGTCGGAGA TATGTTTGG
3' XhoI SSL2tr87	CCGCTCGAGTTAGCTGAGAGCGGCTTTGCATG
3' SpeI SSL2tr87	GGACTAGTTTAGCTGAGAGCGGCTTTGCATGACT TCTGG
3' SpeI SSL2full	GGACTAGTCTACTGCTTGAAGAAGCAGAGGTGA GC
3' SpeI SSL2tr120	GGACTAGTCTACAACCTCGTTCGCGAGTCGGAGAT ATG
5' SSL2tr120-YCt92 IF	CAAACATATCTCCGACTCGCGAACGAGTTGAAATC TAAATTGGCTGTGCAAGATCC
3' SSL2tr120-YCt92 IR	GGATCTTGACACAGCCAATTTAGATTTCAACTCGTT CGCGAGTCGGAGATATGTTTG
5' SSL2tr120-BCt109 IF	CTCCGACTCGCGAACGAGTTGGAAGTCAGATGCA ACACCGAGACCAGCG
3' SSL2tr120-BCt109 IR	CGCTGGTCTCGGTGTTGCATCTGACTTCCAACCTC GTTTCGCGAGTCGGAG
5' SSL2tr87-YCt92 IF	CCAGAAGTCATGCAAAGCCGCTCTCAGCAAATCT AAATTGGCTGTGCAAGATCC
3' SSL2tr87-YCt92 IR	GGATCTTGACACAGCCAATTTAGATTTGCTGAGAGC GGCTTTGCATGACTTCTGG
3' 1st SSL2tr120- YCt27	AATTTTAAGAAATTTGGATCTTGACACAGCCAATTTA GATTTCAACTCGTTCGCGAGTCGG
3' 2nd SSL2tr120- YCt27	TCCATAAACTGTTTCGATCTTGGAGATTTGAATGTT CAATTTTAAGAAATTTGGATCTTGC
3' 3rd SpeI YCt27	GGACTAGTTCATTCTTCCATAAACTGTTTCGATCTT GG
3' 3rd XhoI YCt27	CCGCTCGAGTCATTCTTCCATAAACTGTTTCGATCT TGGAGATTTGAATGTTCAATTTAA
5' SSL2tr120-YCt65 IF	CATATCTCCGACTCGCGAACGAGTTGATGTACCA GGATAAATTACCTCCTAACGTG
3' SSL2tr120-YCt65 IR	CACGTTAGGAGGTAATTTATCCTGGTACATCAACT CGTTCGCGAGTCGGAGATATG
5' SSL2tr87-YCt65 IF	CCAGAAGTCATGCAAAGCCGCTCTCAGCATGTAC CAGGATAAATTACCTCCTAACG
3' SSL2tr87-YCt65 IR	CGTTAGGAGGTAATTTATCCTGGTACATGCTGAGA GCGGCTTTGCATGACTTCTGG
3' 1st SSL2tr87-YCt27 IR	AAATTTGGATCTTGACACAGCCAATTTAGATTTGCT GAGAGCGGCTTTGCATGACTTCTGG
5' Linker-SSL2	GGTGGTTCTGGTGGTGGTTCTGGTGGTGGTTCTG GTATGGTGAAACTCGTCGAGGTTTTG

**Table 3.1** (continued)

5' Sall SSL-3	ACGCGTCGACAAAACAATGAAACTTCGGGAAGTC TTGCAGCACC
5' SSL3-YCt92 IF	GGAAAGGTTTCAGCTAAGGGTGCTAAATCTAAAT TGGCTGTGCAAGATCC
3' SSL3-YCt92 IR	GGATCTTGCACAGCCAATTTAGATTTAGCACCCCTT AGCTGAAACCTTTCC
5' SSL3-BCt109 IF	GGAAAGGTTTCAGCTAAGGGTGCTGAAGTCAGAT GCAACACCGAGACCAGC
3' SSL3-BCt109 IR	GCTGGTCTCGGTGTTGCATCTGACTTCAGCACCC TTAGCTGAAACCTTTCC
5' SSL-3 N171X	GCCTTCACTAATNNSGGGCCAGTTGCTATC
3' SSL-3 N171X	GATAGCAACTGGCCCSNNATTAGTGAAGGC
5' SSL-3 G207X	CATGGCCATGTTCTTGNNSAAGATTAACGTCATC C
3' SSL-3 G207X	GGATGACGTTAATCTTSNNCAAGAACATGGCCAT G
3' SpeI SSL-3	GGACTAGTTCAAGCACCCCTTAGCTGAAACCTTTC C
3' KpnI SSL-3	GGGGTACCTCAAGCACCCCTTAGCTGAAACCTTTC C
5' AatII SSL-3 (no start)	GGTGGTTCTGGTGACGTCAAACCTTCGGGAAGTCT TGCAGCACC
3' KasI SSL-3 (no stop)	CCAATTTAGATTTGGCGCCAGCACCCCTTAGCTGA AACCTTTCC
5' Linker-SSL3	GTTCTGGTGGTGGTTCTGGTGGTGGTTCTGGTAT GAAACTTCGGGAAGTCTTGCAGCACC

*Maintenance of ZX178-08 and ZXD yeast lines, transformation, and media.* Through the course of this work, ZX178-08 and ZXD yeast lines were maintained on YPDE plates at room temperature; plates were restreaked every two weeks. For transformation of ZX178-08 or ZXD, a fresh plate was restreaked and incubated at 28°C for 3 days. Three colonies were then selected with pipette tips to inoculate 3 mL of YPDE media in a 14 mL culture tube and grown for 52-54 hours at 28°C with 220 rpm orbital shaking. 500 µL of the 3 mL cultures were then used to inoculate 50 mL of YPDE media in a 500 mL flask and grown for 12-13 hours at 28°C with 220 rpm orbital shaking. The culture with highest visible cell density was then used to prepare competent cells with Solutions 1 and 2 from the Zymo Frozen-EZ Yeast Transformation kit as described by the manufacturer. 50 µL of cells were mixed with 5 µL of plasmid DNA and Solution 3 from the Zymo Frozen-EZ Yeast Transformation kit in a 14 mL culture tube that was incubated for 3 hours at 28°C with 220 rpm orbital shaking. 1 mL of room temperature YPDE media was added to the transformation mixture and then allowed to recover for 6 hours at 28°C with 220 rpm orbital shaking. Transformation mixtures were transferred to sterile 2 mL tubes, cells were pelleted, washed with water, then plated on SCEglu (-uracil) media. Plates were incubated at 28°C for 5 days; colonies were typically visible by 3 days. YPD(E) and SC(E) media recipe were used as described by Zhuang (2013).<sup>89</sup>

*Complementation and growth testing.* All complementation (- ergosterol media) and growth (+ ergosterol media) experiments were performed as follows. Three biological replicates from fresh transformations (five days after plating) for each construct were selected to inoculate 3 mL of SCEglu media in a 14 mL culture tube; cultures were grown for 3 days at 28°C with 220 rpm orbital shaking. 1 mL of cultures was transferred to 1.5 mL tube, cells were pelleted, media was discarded, cells were washed in sterile water once, pelleted again, then resuspended in 1 mL of sterile water. For complementation and growth testing shown in **Figures 3.4** thru **3.6**, cells were simply diluted 50X in sterile water followed by two 5-fold dilutions. Cells for complementation and growth testing shown in **Figures 3.7** and **3.10C** were first normalized to OD<sub>600</sub> 0.1 then diluted 5-fold two times to OD<sub>600</sub> 0.02 and 0.004. 5 µL of each dilution was then used for spot plating on the four different SC media types. Plates were incubated at 28°C for 5 days until pictures were taken then transferred to room temperature to observe for 15 more days.

*Chemical profiling of hydrocarbon products.* All chemical profiling was performed for constructs in the ZXD yeast line. In general, three biological replicates from fresh transformations for a given construct were selected with a pipette tip to inoculate 3 mL of SCEglu media then grown for 3 days at 28°C with 220 rpm orbital shaking. 30 µL of this culture was then used to inoculate 3 mL of SCEgal media in a 14 mL culture tube then grown at room temperature for 1 week with 220 rpm orbital shaking. To extract hydrocarbon products, 1 mL of culture was vortexed periodically for 30 minutes with 1 mL of acetone in a 4 mL glass screw cap vial. 1 mL of hexane was then added and vortexed periodically for 30 more minutes. After separation of layers, 350 µL of the hexane phase was added to a GC vial insert in a 2 mL GC vial and allowed to evaporate overnight.

50  $\mu$ L of iso-octane was then used to reconstitute the sample; 1  $\mu$ L was analyzed via GC-FID containing an HP-5 column with the following oven program: initial oven temperature held at 150°C for 1 minute followed by a 10°C/min ramp to 280°C then a 5°C/min ramp to 310°C which was held for 1 minute. Hydrocarbon products were identified by comparison of retention times to authentic standards of botryococcene and bis-farnesyl ether provided by Niehaus *et al.* (2011) or squalene (Sigma). Hydrocarbon products were quantified by comparison to a squalene standard curve.

*Sterol profiling for complementing constructs.* ScSS, SSL1-SSL3, SSL1-SSL3Yct92, SSL1-mtSSL3, SSL1-mtSSL3Yct92 constructs in pESC-URA were transformed into ZX178-08. 5 days after transformation a single colony for each construct was selected by pipette tip to inoculate 3 mL of SCEglu media in a 14 mL culture tube. Starter cultures were grown for 3 days at 28°C with 220 rpm orbital shaking. At 3 days, 1 mL of each starter culture was transferred to a sterile 1.5 mL tube, cells were pelleted, washed with 1 mL of water, then reconstituted in 1 mL of sterile water which was used to inoculate 100 mL of SCglu media in a 500 mL flask. 100 mL SCglu cultures were grown at 28°C with 220 rpm orbital shaking for 7 days until cells were pelleted, washed with 50 mL of water twice, then lyophilized. 30 mg of freeze dried cells in a 4 mL glass screw cap vial were saponified as described by Guan *et al.* (2010).<sup>94</sup> Briefly, cells were dissolved in 1 mL of methanol with 30% KOH and 0.25% pyrogallol. This was incubated at 85°C for two hours with gentle vortexing every 30 minutes. Mixture was cooled to room temperature then extracted with 2 mL of petroleum ether; 1 mL of the petroleum ether extract was dried to completion then resuspended in 50  $\mu$ L of iso-octane. 1  $\mu$ L was analyzed on an Agilent 7890A GC equipped with a HP-5MS column (30m x 0.25mm x 0.25 $\mu$ m) coupled to an Agilent 5975C MS insert XL MSD with Triple-Axis Detector (Agilent Technologies, Inc.). Initial oven temperature was held at 200°C for 0.5 minutes then ramped 10°C/min to 270°C followed by a second 3°C/min ramp to 320°C and held for 10 minutes. Mass spectral scan range was from 50-500 amu.

*Growth analysis for complementing constructs.* ScSS, SSL1-SSL3, SSL1-SSL3Yct92, SSL1-mtSSL3, SSL1-mtSSL3Yct92 constructs in pESC-URA were transformed into ZX178-08. 5 days after transformation, 3 biological replicates for each construct were selected by pipette tip to inoculate 3 mL of SCEglu media in a 14 mL culture tube. Starter cultures were grown for 3 days at 28°C with 220 rpm orbital shaking. At 3 days, 1 mL of each starter culture was transferred to a sterile 1.5 mL tube, cells were pelleted, washed with 1 mL of water, then reconstituted in 1 mL of sterile water. 500  $\mu$ L of washed cells were used to inoculate 50 mL of SCglu media in a 250 mL flask. 50 mL SCglu cultures were grown at 28°C with 220 rpm orbital shaking for 7 days. OD<sub>600</sub> measurements were taken every 24 hours for each biological replicate to generate the data in **Figure 3.8C**.

*Mutant library construction.* A general depiction of the process for constructing the two-position saturation library can be seen in **Figure 3.9A**. To begin, a special pESC-URA yeast vector was constructed with SSL1Yct92 cloned into the NotI/SpeI sites of MCS1. Between the two coding sequences for

SSL-1 and YCt92, AatII and KasI restriction sites were introduced *via* SOE PCR using the 5' SSL1-AatII-KasI-YCt92 IF and 3' SSL1-AatII-KasI-YCt92 IR primers. This construct was sequence verified and titled as the pESC-URA:SSL1 (AatII/KasI)YCt92 vector. Next, pET28a:SSL-3 template DNA (Chapter 2) was used in the first two reactions to introduce N171X mutations with the 5' SSL-3 N171X and 3' SSL-3 N171X primers. After reassembly and gel purification, the SSL-3 N171X PCR product was used as template for two more reactions to introduce the G207X mutations with the 5' SSL-3 G207X and 3' SSL-3 G207X primers. The final PCR product containing both N171X and G207X mutations was A-tailed by amplifying with ExTaq polymerase (Takara); primers 5' AatII SSL-3 (no start) and 3' KasI SSL-3 (no stop) were used such that AatII and KasI sites were introduced at the 5' and 3' ends, respectively. A-tailed PCR product was quantified by spectroscopy then used for ligation into pGEM-T Easy (Promega) at a ratio of 3:1 (insert:vector); 10 identical reactions were setup in parallel and allowed to incubate at 16°C for 24 hours. Ligations were pooled and transformed into freshly made DH5 $\alpha$  competent cells. Recovered cells were divided evenly and plated between 20 LB Amp/IPTG/X-gal plates which were grown at 37° for ~15 hours until colonies were distinctly blue or white such that they could easily be scored. At this point, forty independent white colonies were randomly selected for a colony PCR analysis; twenty positive clones were then selected for a sequencing analysis. With satisfactory results from the sequencing analysis, all colonies from the pGEM-T Easy plates were scraped and used to inoculate four 100 mL LB Amp cultures that were grown at 37°C for ~15 hours. These cells were used to prepare a pGEM:mtSSL-3 (N171X,G207X) plasmid DNA library. 50  $\mu$ g of this library was used for a digestion with AatII and KasI restriction enzymes; released insert of the appropriate size was gel purified, quantified, and used to set up 10 identical ligations with the specially made pESC-URA:SSL1(AatII/KasI)YCt92 vector. A 3:1 (insert:vector) ratio was used for these ligations that were then incubated at 16°C for 24 hours. Ligations were pooled and transformed into freshly made DH5 $\alpha$  competent cells. Recovered cells were divided evenly and plated between ten 150 cm LB Amp plates which were grown at 37°C for ~15 hours. 40 randomly selected colonies from these plates were first analyzed *via* colony PCR to 1) quantify the percentage of colonies that had mtSSL-3 inserts and 2) identify colonies for a subsequent sequencing analysis. With satisfactory results from the sequencing analysis, an estimate of the library size was calculated by first counting colonies in a randomly chosen 10 cm<sup>2</sup> region for each 150 cm plate. Colonies per 10 cm<sup>2</sup> areas were then averaged and used to calculate colonies/plate based on the area of a 150cm plate; colonies/plate was then multiplied by the number of plates (10) to obtain the total number of colonies which was multiplied by 0.75 to account for only 75% of the colonies having a mtSSL-3 insert. All colonies from these plates were scraped and used to start four 100 mL LB Amp cultures which were grown for ~15 hours at 37°C. These cells were then used to prepare a pESC-URA:SSL1-mtSSL3 (N171X,G207X) plasmid DNA library.

*Mutant library selection in ZX178-08.* Selection of the mutant library was conducted by transforming 2.5  $\mu$ g aliquots of the pESC-URA:SSL1-mtSSL3

(N171X,G207X) plasmid DNA library into 50  $\mu$ L of ZX178-08 competent cells then plating on SCglu and SCgal media. Transformations of the library into ZX178-08 were carried out as described above except recovery of the cells was allowed for 18 hours and cells were plated on SCglu and SCgal 150 cm plates.

## Chapter Four: Metabolic Engineering for Triterpene Accumulation in the Green Alga *Chlamydomonas reinhardtii*

### Background and Introduction

*Botryococcus braunii* race B is a green, colony forming alga with the ability to accumulate high levels squalene and botryococcene, linear triterpenes that can be further modified with additional methyl groups.<sup>9</sup> The triterpenoids produced by *B. braunii* are an excellent feedstock for the hydrocracking and distillation processes currently used to produce liquid transportation fuels.<sup>14,15</sup> Cultivation of *B. braunii* on a large scale for renewable production of these molecules has long been considered unfeasible because of the alga's slow growth rate, although recent studies may suggest otherwise.<sup>42,95</sup> Nonetheless, engineering high-level accumulation of squalene or botryococcene in a faster growing algal host such as *Chlamydomonas reinhardtii* could help provide basic knowledge about engineering triterpene metabolism in green algae. Such efforts will be important for the future prospects of using green algae as a production platform for sustainable, renewable energy. This chapter will examine progress in nuclear transgene expression methodologies in *C. reinhardtii* then discuss these in relation to how they might be used for successful triterpene metabolic manipulations in this alga.

Metabolic engineering in *C. reinhardtii* has been approached by introducing transgenes into either the nuclear or chloroplast genomes using a variety of methodologies (for review, see Wang et al. 2012, Potvin and Zhang 2010).<sup>96,97</sup> Recombinant DNA is thought to integrate randomly into the nuclear genome through non-homologous end joining<sup>98-100</sup>, whereas integration into the chloroplast genome has been shown to proceed through homologous recombination.<sup>101</sup> While the latter offers better genetic precision and higher levels of protein accumulation, no significant manipulations of chloroplast metabolism have been reported other than those associated with loss-of-function mutants.<sup>102</sup> Causes for this may include protein misfolding although molecular chaperones are found in *C. reinhardtii* chloroplasts, or lack of the proper post-translational modifications that are needed for enzyme activity.<sup>103</sup> Although less precise, nuclear expressed transgenes offer advantages not only for protein folding and post-translational modifications, but also the ability to target enzymes to various subcellular locales. Unfortunately, nuclear genetic engineering in *C. reinhardtii* has been plagued by low-level transgene expression.<sup>104,105</sup>

Transgene promoter analyses have suggested that the transgene expression problems in *C. reinhardtii* might be due to nuclear gene silencing, which occurs through a chromatin based, epigenetic mechanism (histone methylation/acetylation).<sup>106</sup> With the Hsp70A promoter upstream of the RbcS2 promoter and Ble gene, Strenkert et al. (2013) showed that less repressive chromatin states were present near the chimeric promoter and also proposed an important role for heat shock factor 1 (HSF1) in promoting Ble transgene expression.<sup>106</sup> In addition, a study by Kumar et al. (2013) demonstrated optimal luciferase transgene expression when the *C. reinhardtii* Psad gene promoter was



used in comparison to other regularly cited promoters; importantly, these authors observed that the terminator played a more critical role than previously thought.<sup>104</sup> The *PsaD* gene is a highly expressed, intron-less nuclear gene that directs its protein product to the chloroplast *via* an amino-terminal signal sequence.<sup>107</sup> Furthermore, Kumar *et al.* (2013) also determined that transformation with linear DNA excised from the original vector backbone was more efficient for recovery of transgenic lines and transgene expression.<sup>104</sup>

Rasala *et al.* (2012) developed an alternative approach to promoter/terminator choice for increasing transgene expression in *C. reinhardtii* by exploiting the foot-and-mouth-disease-virus (FMDV) 2A self-cleavage linker sequence.<sup>18</sup> When the FMDV 2A sequence is present in an mRNA transcript, a peptide bond in the FMDV 2A linker sequence is skipped by the ribosome during translation.<sup>109</sup> Rasala *et al.* (2012) placed the FMDV 2A sequence in between the *Ble* and *GFP* genes that were under control of one promoter such that a polycistronic mRNA was produced.<sup>108</sup> When translated, two separate polypeptides were produced due to the FMDV 2A sequence. In effect, the cell was forced to express both proteins for survival because of the imposed antibiotic selection.<sup>18</sup> These authors were able to obtain zeocin resistant colonies that exhibited diffuse *GFP* fluorescence, which suggested that *Ble* and *GFP* were expressed and functional. They also utilized this methodology to express an active xylanase enzyme that accumulated in the media due to a secretion signal sequence appended to the xylanase gene.<sup>108</sup> Presumably, this technology could also be used to target enzymes to various subcellular locales such as the endoplasmic reticulum or chloroplast.<sup>105,108</sup>

In addition to construct design and transformation methodology, *C. reinhardtii* strains UVM4 and UVM11 were developed by Neupert *et al.* (2009) for higher-efficiency expression of non-selectable transgenes.<sup>110</sup> First, an emetine resistance gene was introduced that was capable of supporting only low levels of antibiotic resistance (5 mg/L).<sup>110</sup> The resulting transgenic line (Elow47) was then mutagenized with ultra-violet (UV) irradiation and selected for mutants capable of withstanding much higher emetine levels (120 mg/L). Two mutant lines, UVM4 and UVM11, were isolated from this effort. Nuclear encoded fluorescent protein genes were subsequently introduced in the Elow47 and UVM lines. They were shown to accumulate more in the UVM4 and UVM11 lines (~0.2% total soluble protein) *versus* the Elow47 parent line (not detected).<sup>110</sup> More recently, Kong *et al.* (2013) integrated the *C. reinhardtii* squalene synthase (*CrSS*) into the nuclear genome of the UVM strains under control of an *Hsp70A/RbcS2* chimeric promoter.<sup>111</sup> Protein levels of the epitope-tagged *CrSS* assessed in these studies were found to be higher when compared to the control strain; no squalene synthase enzyme activity measurements or squalene product accumulation determinations were reported.<sup>111</sup> Other important aspects identified for successful expression of nuclear encoded transgenes in *C. reinhardtii* include the presence of introns, codon optimization, and exogenous vs. endogenous CDS (reviewed by Gimpel *et al.* (2013), Georgianna *et al.* (2012), Wang *et al.* (2012) and Hyka *et al.* (2012)).<sup>41,44,97,112</sup> However, despite all the advances in algal biotechnology, no precedent setting examples have been published for algae

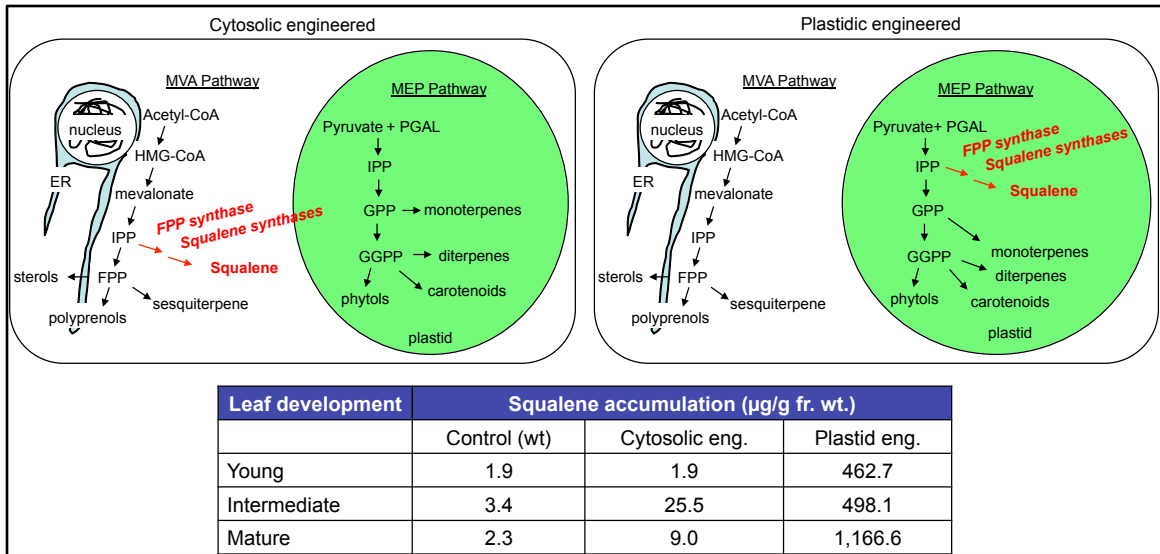
metabolically engineered to produce high levels of terpenoids from nuclear encoded transgenes.<sup>113–118</sup>

Examples from higher plants have shown that exploitation of subcellular targeting can greatly affect terpene product accumulation. Wu *et al.* (2006) demonstrated that targeting farnesyl diphosphate synthase (FPS) and patchouli synthase (PTS, sesquiterpene synthase) to the chloroplast of *Nicotiana tabacum* (tobacco) *via* nuclear encoded transgenes with chloroplast localization sequences afforded the highest levels of terpenes recorded yet.<sup>119</sup> A similar effort by the same group also demonstrated high levels of squalene (triterpene) accumulation in tobacco using the same subcellular localization strategy (**Figure 4.1**).<sup>120</sup> These results suggested more stringent regulation of prenyl diphosphate substrate channeling *via* the cytoplasmic MVA pathway in higher plants, whereas flux through the plastidic MEP pathway into non-native sesquiterpene and triterpene products experienced much less regulation. With such results for higher plants, we asked whether a similar approach might be utilized for engineering triterpene metabolism in *C. reinhardtii*.

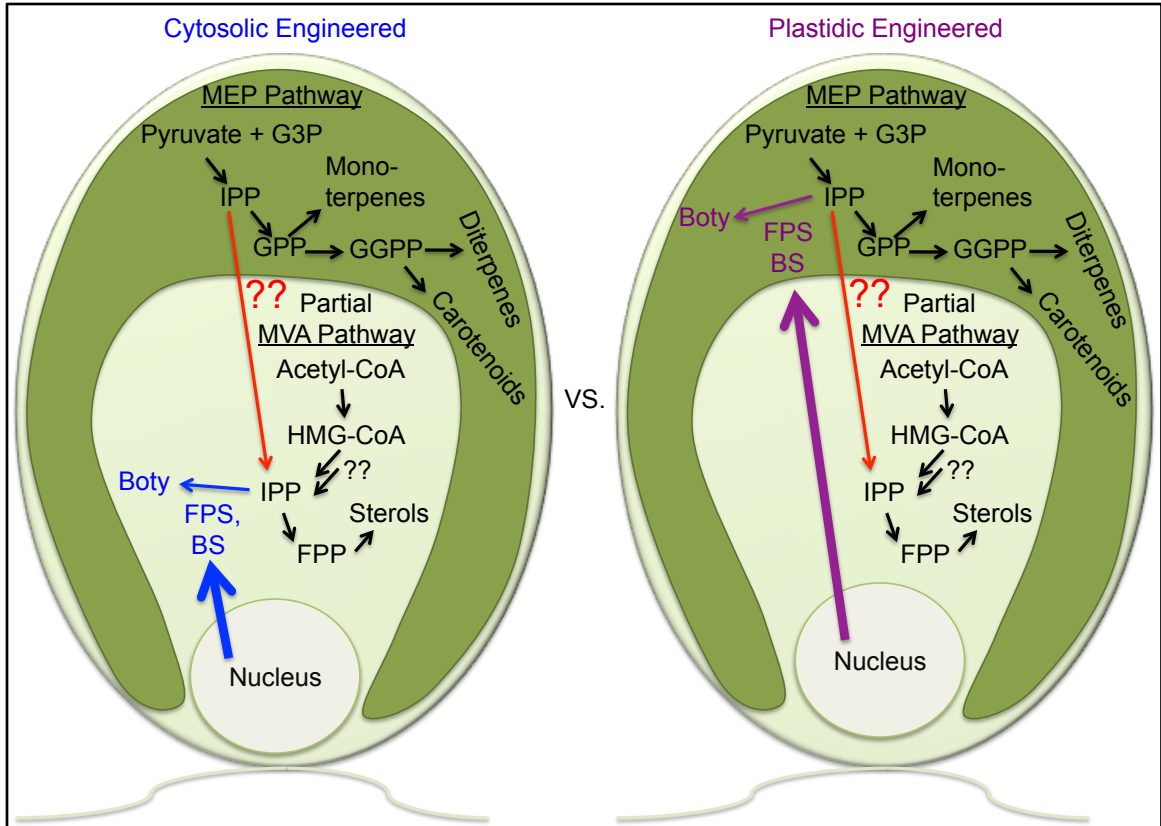
Unlike higher plants, however, the MVA and MEP biosynthetic pathways responsible for supplying isopentenyl diphosphate (IPP) are dispersed differently in algae, a variable that must be taken into consideration before an engineering strategy is chosen. Primary endosymbiotic algae such as glaucophyta and rhodophyta (red algae) typically possess both the MVA and MEP pathways, whereas chlorophyta (green algae) only harbor the plastidic MEP pathway. Secondary endosymbiotic algae phyla euglenophyta, chlorarachniophyta, and heterokontophyta generally utilize both pathways (for review, see Lohr *et al.* 2012).<sup>121</sup> A second variable to consider is the metabolite of interest chosen for engineering. Enzymes are known that efficiently catalyze squalene and botryococcene,<sup>36</sup> thus either metabolite could presumably be engineered for accumulation in an algal host. From an endpoint perspective, engineering botryococcene accumulation might be easier because this molecule would represent a metabolic dead end in *C. reinhardtii*. Squalene, however, could be channeled into and potentially corrupt the native sterol biosynthetic pathway, which might lead to undesirable consequences.

While the long-range objective has been ultimately to engineer novel triterpene biosynthesis into *C. reinhardtii*, our initial strategy was influenced by previous success in developing a yeast terpene production platform. That effort utilized a relatively simple screen for farnesol (FOH) accumulation, the dephosphorylated form of FPP.<sup>89</sup> FOH readily forms in yeast when FPP accumulates because of the naturally occurring, high phosphatase activity. Hence, we assumed that a first step in learning how to engineer terpene metabolism in *C. reinhardtii* might be to overexpress a robust FPP synthase (FPS) followed by FOH accumulation measurements. If we were successful in engineering FPS activity and observing FOH accumulation, then we could use that information to couple overexpression of a particular triterpene synthase with FPS to achieve elevated triterpene biosynthetic capabilities (**Figure 4.2**). Hence, our initial approach was to demonstrate robust nuclear transformation efficiencies followed by FOH measurements for transgenic lines expressing a heterologous,

robust FPS with and without a chloroplast targeting sequence. Unfortunately, we were never able to reliably observe elevated FOH levels. Subsequent efforts were undertaken to introduce both FPS and botryococcene synthase activities into *C. reinhardtii* but were also unsuccessful.



**Figure 4.1** Assessment of the strategy for squalene metabolic engineering in tobacco where enzymes were directed to the cytosol or chloroplast compartments to affect carbon flux through the MVA (cytosol) and MEP (chloroplast) pathways.<sup>119,120</sup> Genes encoding FPP synthase and squalene synthase were introduced into the nuclear genome of tobacco cells under control of strong, constitutive promoters and targeted to either the cytoplasm (left panel) or chloroplast (right panel). The transgenic lines regenerated were propagated in the greenhouse and leaves at different stages of development (*i.e.* small, upper leaves = young; larger, lower, fully-expanded leaves = mature) were sampled for squalene content by GC-MS.

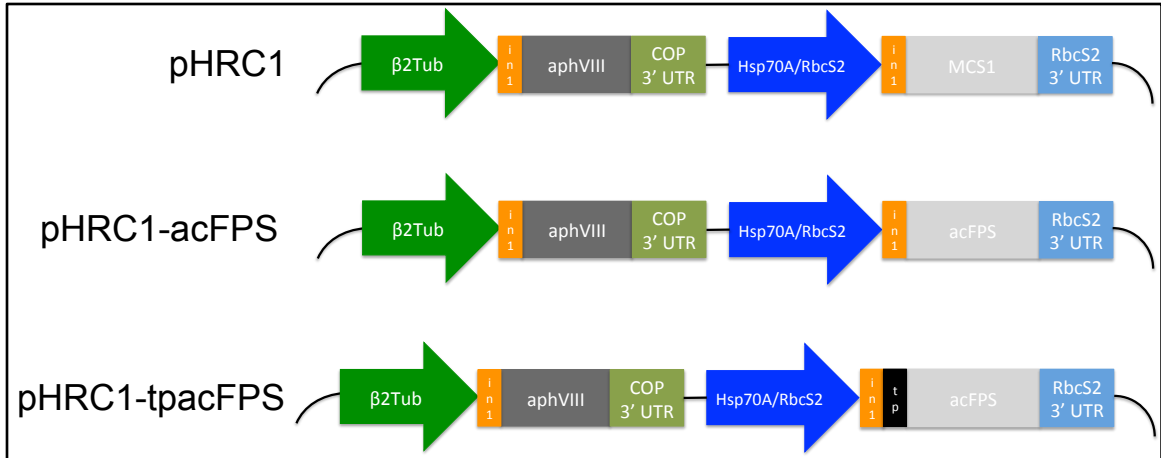


**Figure 4.2** Proposed metabolic engineering scheme to engineer production of botryococcene in the green alga *Chlamydomonas reinhardtii*. Transgenes would be introduced into the nuclear genome in order to target enzymes to the cytoplasm (blue arrows) or chloroplast (purple arrows), allowing for an assessment of carbon flux through the chloroplast localized MEP pathway or the partial MVA pathway resident in the cytoplasm.<sup>122</sup>

## Results

*First-generation nuclear transformation vectors.* To determine if robust nuclear transformation efficiencies could be achieved and if elevated farnesol accumulation could be detected, a series of vectors were constructed from the pHsp70A/RbcS2-Chlamy<sup>123</sup> (pHRC) vector obtained from chlamycollection.org. This vector contained a luciferase gene under the control of an Hsp70A/RbcS2 chimeric promoter equipped with intron1 from the RbcS2 gene and the RbcS2 3' UTR. To construct the pHRC vector series, the luciferase gene was excised and replaced with a synthetic FPS construct (acFPS) that encoded the *Gallus gallus* FPP synthase. acFPS was optimized for expression in *C. reinhardtii* by altering the coding sequence to match *C. reinhardtii* codon bias and high GC-content. The *G. gallus* FPS protein sequence was used to avoid possible endogenous regulatory mechanisms that might be imposed on a native or similar green algal FPS. Targeting to the chloroplast compartment was directed by the chloroplast signal sequence (tp) from the *C. reinhardtii* RbcS2 gene placed upstream of acFPS. Because it was previously shown that transformation efficiency was highest when both gene-of-interest and antibiotic resistance gene were introduced into the nuclear genome from the same plasmid *versus* co-transformation with two separate plasmids,<sup>123</sup> we relied on the double-gene expression constructs. The paromomycin resistance gene from pKS-aphVIII-lox<sup>123</sup> (also obtained from chlamycollection.org) was therefore PCR amplified with 5' EcoRV and 3' XbaI sites and introduced into the corresponding sites in the pHRC vector backbone. The newly created vectors that contained the paromomycin resistance gene (aphVIII)<sup>124</sup> under control of the beta2-tubulin promoter ( $\beta$ 2Tub) and chlamyopsin1 (COP1) 3' UTR were named pHRC1-acFPS (cytosolic targeted) and pHRC1-tpacFPS (chloroplast targeted) (**Figure 4.3**). An empty vector (pHRC1) was also created for use as a control by introducing a multiple cloning site (MCS1) in place of the acFPS gene in pHRC1-acFPS.

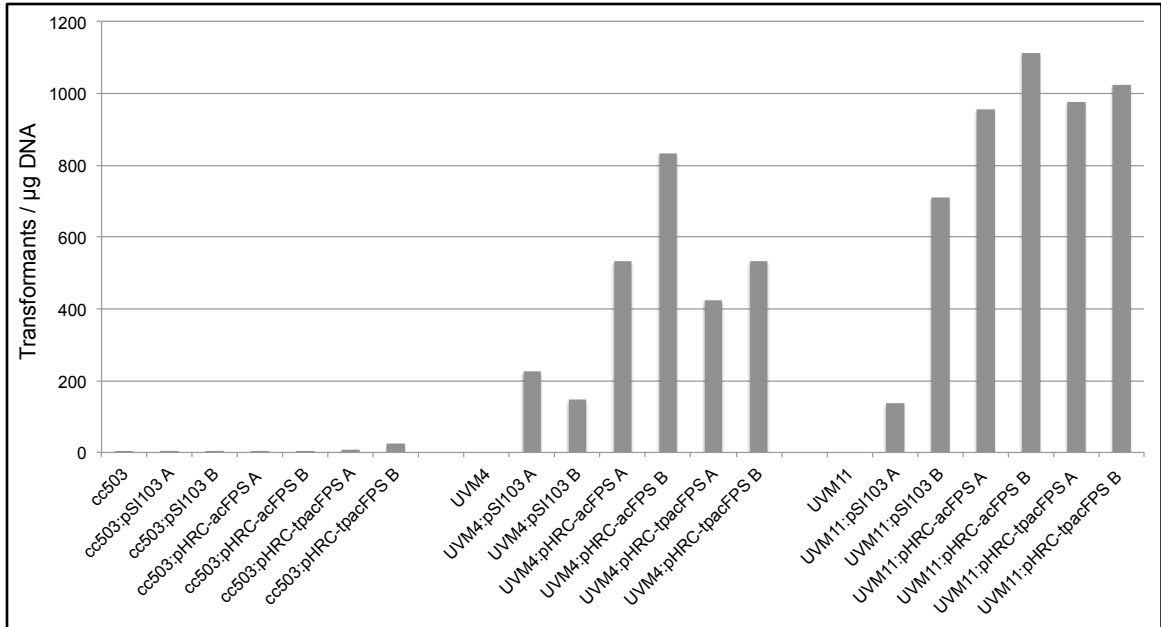
*Efficiency of nuclear transformation with pHRC1 vectors.* Nuclear transformation efficiencies were obtained by transforming vectors into three *C. reinhardtii* strains: CC-503 cw92 (shorthand cc503), UVM4, and UVM11. cc503 is a cell wall deficient strain originally isolated in a mutant study geared toward dissection of cell-wall components,<sup>125</sup> but more recently repurposed for facilitating isolation of genomic DNA for sequencing and easier transformation with foreign DNA. UVM4 and UVM11 were isolated in an effort to circumvent the transgene expression problems in *C. reinhardtii*.<sup>110</sup> Electroporation was chosen as the method of transformation because it was shown to yield better transformation efficiency when compared to glass bead beating mediated transformation.<sup>126</sup> The negative control for these experiments was cells transformed with no plasmid DNA. For the positive control, the pSI103<sup>124</sup> vector (chlamycollection.org) was used because the pHRC1 vector was not fully constructed at the time. The pSI103 vector contained the paromomycin resistance enzyme similar to the pHRC vectors, but was under control of an Hsp70A/RbcS2 chimeric promoter and RbcS2 3' UTR instead of the  $\beta$ 2Tub promoter and COP 3' UTR. Robust transformation efficiency was demonstrated



**Figure 4.3** First-generation *C. reinhardtii* nuclear transformation vectors. Vectors were derived from pHsp70A/RbcS2-Chlmy and pKS-aphVIII-lox and meant to target the acFPS enzyme to either the cytosol (pHRC1-acFPS) or chloroplast (pHRC1-tpacFPS) compartments for an assessment of farnesol accumulation.  $\beta$ 2Tub, *C. reinhardtii* beta2-tubulin promoter; in1, *C. reinhardtii* RbcS2 intron1; aphVIII, *Streptomyces rimosus* aminoglycoside 3'-phosphotransferase VIII; COP, *C. reinhardtii* chlamyopsin1 3' UTR; Hsp70A/RbcS2, chimeric promoter; MCS1 (XhoI-AatII-AclI-AgeI-ApeI-BamHI); RbcS2, *C. reinhardtii* RbcS2 3' UTR; acFPS, *Gallus gallus* FPS gene codon optimized for *C. reinhardtii* expression; tp, *C. reinhardtii* RbcS2 chloroplast signal sequence.

in this experiment, particularly for the UVM4 and UVM11 strains with the pHRC1 vectors (**Figure 4.4**). Much lower transformation efficiencies were observed for the cc503 strain and pSI103 vector. While not a direct comparison between vectors because of differences in plasmid backbones, it seemed that the aphVIII gene under control of the  $\beta$ 2Tub promoter (pHRC1 vectors) was either more efficient at nuclear integration or served as a better expression construct than the aphVIII gene with the chimeric promoter (pSI103). Nonetheless, these experiments provided evidence that robust nuclear transformations could be accomplished, a pivotal accomplishment for future experiments where many independent lines would be needed for chemical profiling to identify high producers.

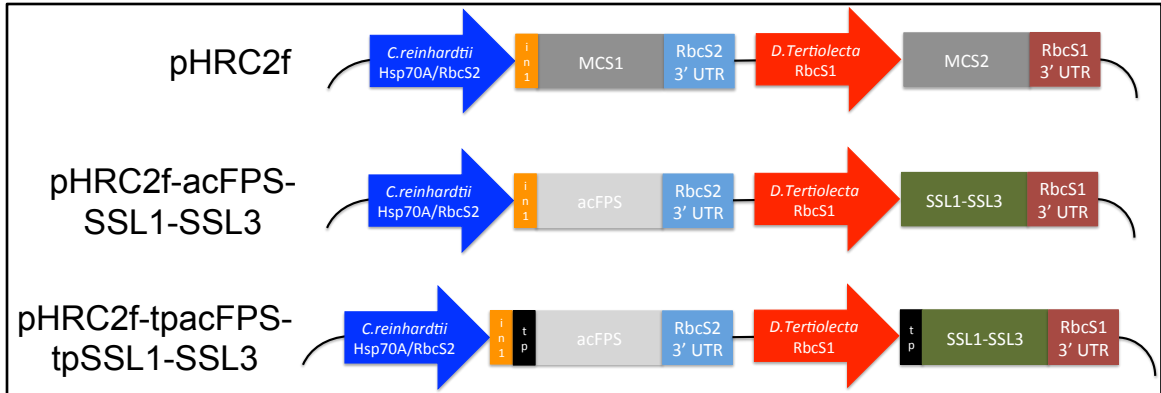




**Figure 4.4** Nuclear transformation efficiency in *C. reinhardtii*. cc503, UVM4, and UVM11 *C. reinhardtii* strains were transformed with no DNA, pSI103, pHRC1-acFPS, or pHRC1-tpacFPS vectors via electroporation with 5 µg of plasmid DNA. Two independent transformations (designated by A or B after plasmid names) were performed for each assessment. After 24 hours of recovery, cells were plated on TAP media with 30 mg/L paromomycin and grown under 23:1 (light:dark) fluorescent lights for seven days until colonies were distinctly visible for scoring.

*Verification of transgene insertion and farnesol accumulation measurements.* After excellent transformation efficiencies were established with the UVM algae strains and pHRC1 vectors, independent transgenic lines were subjected to a genomic DNA PCR analysis to check for the presence of the acFPS transgene. Good genomic DNA preparations achieved with the “QUICK AND EASY” method were verified by primers that amplified a small section of the light harvesting complex I (LHCa) gene; all preps showed amplification of a distinct product with the proper size (~500bp). The percentage of paromomycin resistant lines that tested PCR positive for the presence of the acFPS gene were: UVM4:pHRC1-acFPS = 40.6% (13 of 32); UVM4:pHRC1-tpacFPS = 12.3% (14 of 114); UVM11:pHRC1-acFPS = 43.8% (14 of 32); and UVM11:pHRC1-tpacFPS = 26.6% (8 of 30). PCR positive transgenic lines were grown and tested for the presence of farnesol. None of the transgenic lines that were extracted and assessed *via* GC-FID had detectable levels (>1 ng) of FOH.

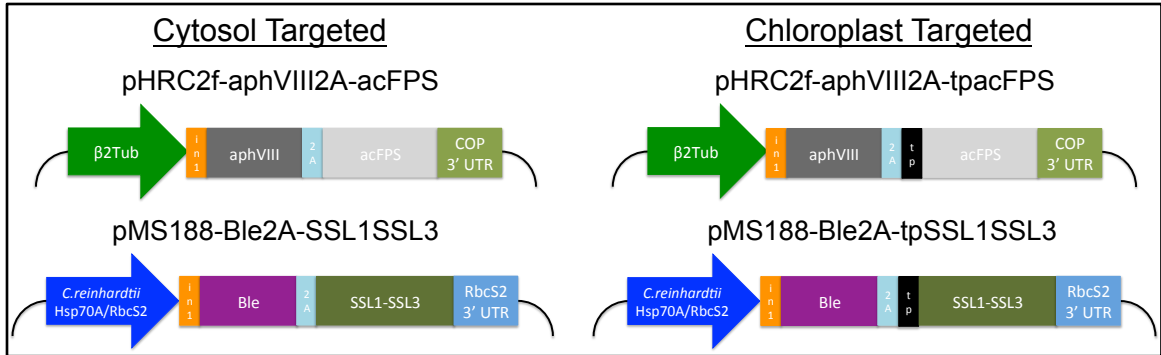
*Second-generation nuclear transformation vectors.* Because engineering botryococcene accumulation in *C. reinhardtii* rather than FPP alone might be a more benign impact on cellular metabolism, we were stimulated to develop a second generation of gene constructs. The enzymes responsible for botryococcene production in *B. braunii* have been characterized and shown to be functional in *Saccharomyces cerevisiae* not only as independently expressed enzymes, but also as a fusion enzyme.<sup>36</sup> The latter was important because not only did the fusion enzyme lead to higher product titers in yeast, but it also only required one additional promoter/terminator for gene expression instead of two. The additional promoter/terminator chosen to drive expression of the SSL1-SSL3 fusion enzyme was from the *Dunaliella tertiolecta* RbcS1 gene that was shown to support expression of the Ble gene in *C. reinhardtii*.<sup>127</sup> Using these additional sequences, a second series of vectors were constructed from the pHRC1 vector. First, pHRC2f was constructed by introducing the *D. tertiolecta* RbcS1 regulatory elements with a multiple cloning site (MCS2) in-between into the KpnI site of pHRC1. This vector was created for use as an empty vector control or testing additional constructs in the future. The acFPS and SSL1-SSL3 coding sequences were sequentially introduced to produce pHRC2f-acFPS-SSL1-SSL3; this vector was then used to introduce the tp sequence upstream of the acFPS and SSL1-SSL3 coding sequences to produce pHRC2f-tpacFPS-tpSSL1-SSL3 (**Figure 4.5**). As before, all pHRC2f vectors contained the previously described aphVIII paromomycin resistance marker. These vectors were transformed into the UVM strains. Many independent lines that were resistant to paromomycin were obtained. Unfortunately, these transgenic lines were not chemically profiled or assessed in any other fashion.



**Figure 4.5** Second-generation *C. reinhardtii* nuclear transformation vectors. Vectors were derived from pHRC1 and meant to target acFPS and SSL1-SSL3 enzymes to either the cytosol (pHRC2f-acFPS-SSL1-SSL3) or chloroplast (pHRC2f-tpacFPS-tpSSL1-SSL3) compartments to afford an assessment of botryococcene accumulation in *C. reinhardtii*. Not shown is the aphVIII paromomycin resistance construct under control of the *C. reinhardtii* beta2-tubulin ( $\beta$ 2Tub) promoter and *C. reinhardtii* chlamyopsin1 (COP) 3' UTR. MCS1 (XhoI-AatII-AclI-AgeI-ApeI-BamHI); MCS2 (XmaI-ClaI-NsiI-PciI-SpeI-AvrII); RbcS2 3' UTR, *C. reinhardtii* RbcS2 3' UTR; acFPS, *Gallus gallus* FPS gene codon optimized for expression in *C. reinhardtii*; SSL1-SSL3, *B. braunii* botryococcene synthase; RbcS1 3' UTR, *D. tertiolecta* RbcS1 3' UTR; tp, *C. reinhardtii* RbcS2 chloroplast signal sequence.

*Third-generation nuclear transformation vectors.* With the emergence of the FMDV 2A self-cleavage peptide technology<sup>105,108</sup> that was demonstrated to be functional in *C. reinhardtii* and another study that provided evidence for averting gene silencing mechanisms in *C. reinhardtii*,<sup>106</sup> a third series of vectors were constructed. The pHRC2f vector was used to append the FMDV 2A sequence downstream of the aphVIII coding sequence followed by the acFPS coding sequence with and without the tp sequence (**Figure 4.6**). This setup was used because the aphVIII resistance enzyme under control of the *C. reinhardtii*  $\beta$ 2Tub and COP regulatory elements was demonstrated as functional in previous experiments. For exploitation of the FMDV 2A technology in expressing the SSL1-SSL3 fusion enzyme, another antibiotic resistance enzyme was necessary. The pMS188<sup>106</sup> vector was chosen because it contained an Hsp70A/RbcS2 chimeric promoter that was shown to be most effective at combating silencing of the Ble transgene when compared to a series of other Hsp70A/RbcS2 chimeric promoters. This vector was obtained from chlamycollection.org and similar to construction of the pHRC2f-aphVIII2A-(tp)acFPS vectors generated, the FMDV 2A sequence was introduced downstream of the Ble gene followed by the SSL1-SSL3 fusion enzyme sequence with and without the tp sequence (**Figure 4.6**).

These vectors were co-transformed into the UVM4 strain and selected on the suggested antibiotic concentrations (15 mg/L zeocin and 30 mg/L paromomycin). When UVM4 was co-transformed with the pHRC2f and pMS188 empty vectors, numerous colonies were obtained which suggested that the double antibiotic selection in *C. reinhardtii* was possible. However, no colonies were ever obtained when the experimental constructs were co-transformed. Numerous colonies were obtained when only one of the experimental vectors was transformed followed by selection on the appropriate single antibiotic. This outcome occurred for two independent experiments. These results suggested two possibilities: either the experimental constructs that were tested produced inactive resistance enzymes due to the remaining FMDV 2A peptide sequence at their C-terminal ends or the double antibiotic selection in concert with the transgenes expressed was too much metabolic burden for the cells to tolerate. The former seems unlikely as colonies were obtained on single antibiotic selection when the FMDV 2A vectors were transformed into UVM4 individually. A follow up experiment should be conducted to test if colonies appear with lower antibiotic concentrations for the double selection after co-transformation of both FMDV 2A vectors. If colonies are obtained, they could then be tested on increasing concentrations of antibiotics to provide additional confidence transgenic lines harbor both selection marker genes and that they are being significantly expressed.



**Figure 4.6** Third-generation *C. reinhardtii* nuclear transformation vectors. Vectors were derived from pHRC2f and pMS188. Genes-of-interest were placed downstream of an antibiotic resistance gene and the FMDV 2A self-cleavage sequence which produces a polycistronic mRNA. When translated, the mRNA results in two separate protein products because a peptide bond is skipped by the ribosome in the FMDV 2A sequence.  $\beta$ 2Tub, *C. reinhardtii* beta2-tubulin promoter; in1, *C. reinhardtii* RbcS2 intron1; aphVIII, *Streptomyces rimosus* aminoglycoside 3'-phosphotransferase VIII; 2A; foot-and-mouth-disease-virus polyprotein self-cleavage sequence; acFPS, *Gallus gallus* FPS codon optimized for expression in *C. reinhardtii*; COP, *C. reinhardtii* chlamyopsin1 3' UTR; Hsp70A/RbcS2, chimeric promoter; Ble, bleomycin/zeocin resistance enzyme; RbcS2 3' UTR, *C. reinhardtii* RbcS2 3' UTR; tp, *C. reinhardtii* RbcS2 chloroplast signal sequence.

## Discussion

Lackluster reports for *Chlamydomonas reinhardtii* terpenoid engineering and more general metabolic engineering insinuates underlying issues that have not yet been fully appreciated.<sup>113–118</sup> One area that has been investigated extensively is construct design for nuclear transformation while another has been host strain development geared towards increasing efficiency of transgene expression. For the former, the FMDV 2A self-cleavage technology developed by Rasala *et al.* (2012) coupled with expression controlled by the Psad promoter described by Kumar *et al.* (2013) might be the best available option for high-level transgene expression.<sup>104,105,108</sup> The latter, tackled with the creation of the *C. reinhardtii* UVM strains, offers the best avenue for host choice to achieve successful transgene expression in this alga.<sup>110</sup> Using these technologies together may be the superior route for assessing metabolic engineering efforts in *C. reinhardtii*. Overcoming the nuclear transgene expression problem in *C. reinhardtii* and presumably other green algae is just one of many technological developments that are needed before complex genetic engineering efforts such as expressing the entire MVA pathway in the cytosol of a green alga or adding extra copies of the MEP pathway enzymes can be reasonably assessed.

High titers of naturally occurring algal terpenoids are clearly possible as observed for *Dunaliella salina*<sup>128</sup> and *Botryococcus braunii* race B.<sup>62</sup> But how these algae accomplish this and what the feasibility is that these mechanisms can be captured and deployed in other algae species remains to be determined. *B. braunii* race B accumulates 30% of its mass in triterpene oils, albeit at the cost of growth rate.<sup>62</sup> Increased growth rates can be achieved under various culture conditions while maintaining high levels of hydrocarbon production<sup>95</sup> – engineered strains that accomplish this have not been reported. Moreover, *B. braunii* hydrocarbon production and secretion into the extracellular matrix seems to be intimately linked with the cell cycle<sup>129,130</sup> making prospects for deployment of this mechanism into other algae quite daunting. Thus, using *B. braunii* as a design template for terpenoid metabolic engineering approaches may not be as fruitful as other algae such as *Dunaliella salina*. This halophilic, marine green alga can accumulate 10-15% of its mass in  $\beta$ -carotene that is stored in plastid globules, and it can also form cytoplasmic globules of TAGs produced during stress inducing conditions.<sup>128,131–134</sup> If the carotenogenic biosynthetic enzymes were knocked down/out, IPP in the chloroplast might be available for diversion towards the production of a designer isoprenoid. Such efforts could be bolstered if extra copies of DXS (a putative rate-limiting MEP enzyme) were introduced into the nuclear genome with chloroplast signal sequences. Nuclear copies of isopentenyl diphosphate isomerase (idi), farnesyl diphosphate synthase (FPS) and botryococcene synthase (BS) could then be introduced and targeted to the cytoplasm. In effect, substrates originally found in the chloroplast (pyruvate and G3P) might be channeled into IPP, which could then be drawn into the cytoplasm and channeled into botryococcene biosynthesis. These considerations are simply speculation until the technological means for facile genetic engineering in algae advances.

Another key area for successful metabolic engineering efforts in the future will be to understand algae-specific oleaginous traits. New 'omics' studies that have emerged suggest these play key roles in the ability of certain species to accumulate their respective lipid. For example, in the heterokont *Nannochloropsis* sp., gene dosage (13 copies of DGAT vs. 4 in *C. reinhardtii*) seems to be extremely important for accumulating high levels of TAGs.<sup>25</sup> Moreover, *B. braunii* race B constitutively expresses three distinct genes for DXS,<sup>137</sup> presumably to promote high carbon flux through the MEP pathway to support the biosynthesis and accumulation of the abundant triterpene molecules. Such *in silico* predictions that are allowed by thorough, 'omics' information may help to improve the success rate for synthetic biology designs aimed at engineering metabolism in intractable organisms like algae.<sup>138</sup>

Currently, green algae are blank slates for testing new strategies to engineer terpenoid accumulations. Emerging information about specialized green algae metabolism will buttress systems/synthetic biology approaches where carbon fluxes are optimized among various cellular pathways in order to achieve maximum growth and metabolite production rates. Coupled with improved genetic engineering methodology, thorough bioinformatics, machine learning, and network flux analyses, new strategies will appear for the successful manipulation of terpenoid metabolism in green algae.

## Materials and Methods

*Construction of first-generation vectors.* All primers used for constructs used in algae work can be found in **Table 4.1**. A synthetic FPS gene based on the avian (*Gallus gallus*) FPS protein sequence was codon optimized and synthesized by Genscript with the *C. reinhardtii* RbcS2 chloroplast targeting sequence (tp) appended to the 5' end and flanked on both side by MfeI sites; this sequence (tpacFPS) was received in pUC57. The tpacFPS sequence was PCR amplified with 5' XhoI MfeI tp and 3' BamHI acFPS primers; this PCR product was then XhoI/BamHI double digested and ligated into pHsp70A/RbcS2-Chlamy<sup>123</sup> (pHRC) obtained from chlamycollection.org that had been digested with XhoI and BamHI to release the crluc coding sequence. Individual clones were then isolated and sequence verified; the new vector was titled pHRC-tpacFPS. pHRC-tpacFPS was digested with MfeI to release the tp sequence then allowed to re-ligate. Individual clones were isolated and sequence verified; the new vector was titled pHRC-acFPS. The aphVIII gene cassette from pKS-aphVIII-lox was then introduced into pHRC-acFPS and pHRC-tpacFPS EcoRV and XbaI sites by first amplifying it with 5' EcoRV  $\beta$ 2Tub Pro and 3' XbaI COP 3' UTR primers, digesting the PCR product with EcoRV and XbaI, then ligating it into the appropriately digested pHRC vectors. Independent clones were isolated and sequence verified. These vectors were titled pHRC1-acFPS and pHRC1-tpacFPS. pHRC1 empty vector was created by digesting pHRC1-acFPS with XhoI and BamHI to release acFPS then ligating in 5' pHRC MCS1 and 3' pHRC MCS1 primers designed to anneal such that they had XhoI and BamHI sticky ends. Independent clones were isolated and sequence verified.

*Construction of second-generation vectors.* Genomic DNA isolated from *D. tertiolecta* was used as template to PCR amplify 236bp of the RbcS1 promoter with 5' KpnI D.tert RbcS1 Pro -236 and 3' KpnI/NheI/AvrII/XmaI D.tert RbcS1 Pro primers; the RbcS1 terminator was PCR amplified with 5' AvrII D.tert RbcS1 Term and 3' NheI D.tert RbcS1 Term primers. These PCR products were ligated into pGEM-T Easy, independent clones were isolated, then sequence verified. The pGEM-T Easy plasmid with *D. tertiolecta* RbcS1 terminator flanked by 5' AvrII and 3' NheI sites was digested to release the AvrII/NheI fragment which was subcloned into the pGEM-T Easy plasmid with the RbcS1 promoter containing AvrII and NheI sites at the 3' end of the introduced fragment. After sequence verification, this new vector with the *D. tertiolecta* RbcS1 promoter and terminator was digested with AvrII and XmaI so that the 5' pHRC MCS2 and 3' pHRC MCS2 primers designed to anneal with AvrII and XmaI sticky ends could be ligated in. Independent clones were isolated and sequence verified then used in a KpnI digestion to release the entire cassette for subcloning into pHRC1 that was KpnI digested; clones sequence verified with the cassette in the forward direction were kept and titled pHRC2f. Next, pHRC2f-tpacFPS was created by subcloning tpacFPS as a XhoI/BamHI fragment from pHRC1-tpacFPS; a sequence verified pHRC2f-tpacFPS clone was then digested with MfeI to release the tp sequence. The vector backbone was then allowed to self-ligate creating pHRC2f-acFPS. SSL1-SSL3 fusion enzyme was PCR amplified from plasmid



**Table 4.1** Primers used for cloning and screening algae constructs.

Primer	5' → 3' Sequence
5' XhoI/MfeI tp	tcccctcgagcaattgatggccgccgtcattgcc
5' XhoI/MfeI acFPS	tccccccgggcaattgatgcacaagttcaccggcggtgaacg
3' BamHI acFPS	cgcgatccttacttctggcgctttagatcttctgcg
5' EcoRV β2Tub Pro	ccggatatctcttcttgcgctatgacacttcc
3' XbaI COP 3' UTR	ctagtctagagtaccatcaactgacgttacattc
5' pHRC MCS1	tcgaggacgtcaacgttaccgggtggcccg
3' pHRC MCS1	gatccgggcccaccggtaacgttgacgtcc
5' KpnI D.tert RbcS1 Pro -236	cggggtaccgcacttgacattcttcttgcattcc
3' KpnI/NheI/AvrII/XmaI D.tert RbcS1 Pro	ggtaccgctagccctaggcccgggtgtgtgtttgtgaagggtgtgtgg
5' AvrII D.tert RbcS1 Term	ctagcctagggctgttagcacagcacttgcc
3' NheI D.tert RbcS1 Term	ctagctagcccattcttctgtactgcttatacag
5' pHRC MCS2	ccgggatcgatatgcatacatgtactagtc
3' pHRC MCS2	ctaggactagtacatgtatgcatatcgatc
5' SpeI/EcoRI SSL-1	ggactagtgaattcatgactatgcaccaagaccacg
3' AvrII SSL-3	ctagcctaggtaagcacccttagctgaaacc
3' AphVIII-2A	gcttcagcaggtcgaagttcagggctgcttcaccggggcgaa gaactcgtccaacagcc
5' 1st 2A-MfeI-acFPS	ctggcgggacgtggagagcaaccgggcccccaattgat gcacaagttcaccggcgtg
5' 2nd 2A-MfeI-acFPS	ggtgaagcagaccctgaacttcgacctgctgaagctggcggg cgacgtggagagcaacc
3' acFPS-G8-flag	tcatccttgaatcgccaccgccaccgccaccgccacccttctg gcgctttagatcttc
5' G8-flag-COP	ggtggcgggtggcgattacaaggatgacgacgataagtgagg gacctgatgggtgtgtgtg
5' SphI-pMS188-Ble	acatgcatgcagcccctggagcgggtgccctcctg
3' 1st BamHI EcoRI 2ABLE	gctctccacgtcggccgagcttcagcaggtcgaagttcagg
3' 2nd BamHI EcoRI 2ABLE	cgggatccgaattcggggcccgggtgtctccacgtcggccc ccagc
5' EcoRI-SSL1	cggtgaagcagaccctgaacttcgacctgctgaagctggcgg gcgacgtggagagcaacc
3' 1st SSL1-G8-myc-Linker	tcaactctgttcgccaccgccaccgccaccgccacccttggtg ggagtggggctgcgc
3' 2 <sup>nd</sup> SSL1-G8-myc-Linker	ccagaaccaccaccagaaccaccggaggtcttctcggaatc aacttctgttcgccaccg
5' Linker-SSL3 v2	gttctggtggtgttctggtggtggttctggtatgaaacttcggga agtcttcgacacc
3' BamHI-pMS188-SSL3	cgggatcctctagagtcgggtcgacgtcggtaagcaccctta gctgaaaccttcc
5' SSL3 (-BamHI 663bp) IF	ggaggatgtcttgaggaccctcctcgcatctggtgg
3' SSL3 (-BamHI 663bp) IR	ccaccagatgcgaggagggtcctccaagacatcctcc
5' SSL3 (-BamHI 1047bp) IF	ccaagtgtgagggaggacccaacttgcactcacag

**Table 4.1** (continued)

3' SSL3 (-BamHI 1047bp) IR	ctgtgagtgcaaagttggggtcctccctcacacacttg
5' LHCa 3'UTR	gcgctctctcttgagttgagtg
3' LHCa 3'UTR	cccatggtacgttacaggaaagg

DNA obtained from Niehaus *et al.* (2011) using primers 5' SpeI/EcoRI SSL-1 and 3' AvrII SSL-3. This PCR product was digested with SpeI and AvrII then ligated into pHRC2f-acFPS and pHRC2f-tpacFPS SpeI and AvrII sites in MCS2. Sequence verified clones were titled pHRC2f-acFPS-SSL1-SSL3 and pHRC2f-tpacFPS-SSL1-SSL3. pHRC2f-tpacFPS-SSL1-SSL3 was then digested with EcoRI and ligated with a EcoRI/EcoRI tp fragment to afford pHRC2f-tpacFPS-tpSSL1-SSL3. The correct direction of the newly inserted tp sequence was sequence verified.

*Construction of third-generation vectors.* To construct pHRC2f-aphVIII2A-acFPS, a sequence overlap extension PCR strategy was employed. Four different PCR reactions were performed as follows: 1) pHRC2f as template with 5' EcoRV  $\beta$ 2Tub Pro and 3' aphVIII-2A primers to amplify the  $\beta$ 2Tub promoter and aphVIII with 2A appended to the 3' end of aphVIII, 2) pHRC1-acFPS as template with 5' 1<sup>st</sup> 2A-MfeI-acFPS and 3' acFPS-G8-flag primers to append a portion of 2A to the 5' end and a FLAG epitope spaced by an octa-glycine (G8) linker<sup>139</sup> to the 3' end of acFPS, 3) PCR product from reaction 2 as template with 5' 2<sup>nd</sup> 2A-MfeI-acFPS and 3' acFPS-G8-flag primers to finish appending the 2A sequence to the 5' end of acFPS, and 4) pHRC2f as template with 5' G8-flag-COP and 3' XbaI COP 3' UTR primers to add overlapping sequence to the COP 3' UTR. In a fifth reaction, PCR products from reactions 1, 3, and 4 were used as template with 5' EcoRV  $\beta$ 2Tub Pro and 3' XbaI COP 3' UTR primers to assemble the entire construct. The PCR product was double digested with EcoRV/XbaI and ligated into purified pHRC2f vector backbone that was digested with EcoRV and XbaI. Positive clones were identified and sequence verified; this vector was titled pHRC2f-aphVIII2A-acFPS. MfeI was used to linearize pHRC2f-aphVIII2A-acFPS so that an MfeI/MfeI tp fragment could be introduced in-frame at the 5' end of acFPS. Independent clones were isolated and sequenced to verify the correct orientation of the tp sequence. This new vector was titled pHRC2f-aphVIII2A-tpacFPS.

Another SOE PCR strategy was employed to construct the pMS188-Ble2A-SSL1-SSL3 and pMS188-Ble2A-tpSSL1-SSL3 with a few additional steps. First, a pMS188-Ble2A plasmid was generated that removed the TAA for the Ble gene and introduced the 2A sequence and an EcoRI site. This was accomplished by first PCR amplifying Ble from pMS188 with the 5' SphI pMS188 Ble and 3' 1<sup>st</sup> BamHI EcoRI 2ABle primers then using this PCR product as template in a second reaction with 5' SphI pMS188 Ble and 3' 2<sup>nd</sup> BamHI EcoRI 2ABle primers. The PCR product of the second reaction was SphI/BamHI double digested and ligated into pMS188 backbone digested with SphI/BamHI. Second, to afford use of the introduced EcoRI site and a BamHI site already present in the pMS188-Ble2A plasmid to clone in SSL1-SSL3, two BamHI sites were synonymously mutated in SSL-3 and a G8 linker-myc epitope was introduced to give the following: EcoRI-SSL1-G8-myc-linker-SSL3-BamHI. This was accomplished using SOE PCR as follows: 1) pET28a:SSL-1 was used as template with 5' EcoRI SSL1 and 3' 1<sup>st</sup> SSL1-G8-myc-Linker primers, 2) PCR product from reaction 1 was used as template with 5' EcoRI SSL1 and 3' 2<sup>nd</sup> SSL1-G8-myc-Linker primers, 3) pET28a:SSL-3 was used as template with 5' Linker-SSL3 v2

and 3' SSL3 (-BamHI 663bp) IR primers, 4) pET28a:SSL-3 was used as template with 5' SSL3 (-BamHI 663bp) IF and 3' SSL3 (-BamHI 1047bp) IR primers, 5) pET28a was used as template with 5' SSL3 (-BamHI 1047bp) IF and 3' BamHI-pMS188-SSL3 primers, and 6) PCR products from reactions 3, 4, and 5 were used as template with 5' Linker-SSL3 v2 and 3' BamHI-pMS188-SSL3 primers. In a seventh reaction, PCR products from reactions 2 and 6 were used as template with 5' EcoRI SSL1 and 3' BamHI-pMS188-SSL3 primers to assemble the entire EcoRI-SSL1-G8-myc-linker-SSL3-BamHI construct. This PCR product was EcoRI/BamHI double digested and ligated into the EcoRI/BamHI double digested pMS188-Ble2A vector. Once sequence verified, this new vector (pMS188-Ble2A-SSL1-SSL3) was linearized with EcoRI so that the tp sequence flanked by EcoRI sites could be ligated in. Positive clones with the tp insert were identified and sequenced to verify the correct orientation of the tp sequence. The new vector was titled pMS188-Ble2A-tpSSL1-SSL3.

*C. reinhardtii* and *D. tertiolecta* maintenance and media. All *C. reinhardtii* strains (CC-503 cw92, UVM4, and UVM11) used in this work were maintained on TAP solid media at room temperature; cultures were restreaked monthly. Dr. Wayne Curtis (Penn State) provided strain CC-503, while UVM4 and UVM11 were obtained from Neupert *et al.* (2009). TAP media recipe was used as described by Gorman and Levine (1965).<sup>140</sup> Liquid TAP cultures were typically grown in shake flasks under 23:1 (light:dark) fluorescent lighting with 100 rpm orbital shaking at room temperature. *Dunaliella tertiolecta* (ATCC 30929) was cultured in liquid f/2 media as described by Guillard and Ryther (1962) and Guillard (1975) but with artificial seawater replacing seawater.<sup>141,55</sup> *D. tertiolecta* cultures were grown in shake flasks with under 23:1 (light:dark) fluorescent lighting with 100 rpm orbital shaking at room temperature. Genomic DNA was isolated from *D. tertiolecta* using a plant genomic DNA isolation kit from Promega.

*C. reinhardtii* electroporation protocol for nuclear transformation. To begin, a starter culture was generated by inoculating 10 mL of TAP media in a 50 mL flask with cells from a slant culture and grown under 23:1 (light:dark) fluorescent lighting with 100 rpm orbital shaking for 3 days at room temperature. 500 mL of TAP media in a 2 L flask was inoculated with 5 mL of the starter culture and grown to OD<sub>550</sub> 0.4-06 (~2-3 x 10<sup>8</sup> cells/mL). Cells from 50 mL of culture were pelleted by centrifugation then resuspended in 700 µL TAP + 40 mM sucrose. Cells were transferred to 4 mm sterile cuvette with 5 µg supercoiled plasmid DNA, incubated on ice for 1 minute immediately prior to the electroporation event, making sure to wipe outside of cuvette dry before electroporation to prevent electrical arcing. Electroporation conditions: exponential decay, 800 V, 25 µF, ∞ Ω. Cells were recovered by transferring from cuvette to a 50 mL flask with 10 mL of TAP + 40 mM sucrose then grown under 23:1 (light:dark) fluorescent lighting with 100 rpm orbital shaking at room temperature for 24 hours. Cells were collected by centrifugation, resuspended in TAP media, and plated on solid TAP media with the appropriate antibiotic (paromomycin = 30 mg/L; zeocin = 15 mg/L). Colonies were typically visible around 5 days and big enough to be restreaked around 7 days.

*PCR screening for acFPS transgene insertion.* Initial colonies obtained on selection were named and restreaked on fresh selection media. Once enough biological material was available, the "QUICK AND EASY" method described by Steve Pollock, Louisiana State University found at [chlamy.org](http://chlamy.org) was used to prepare genomic DNA for PCR screening to check for the presence of acFPS in paromomycin resistant lines.

*Chemical profiling of hydrocarbon products in transgenic lines.* Using only PCR verified transgenic lines, a 10 mL TAP media culture was grown in a 50 mL flask under 23:1 (light:dark) fluorescent lighting with 100 rpm orbital shaking at room temperature for 5 days. 1 mL of the 10 mL culture was transferred to a 4 mL glass screw cap vial with 1 mL of acetone and vortexed vigorously periodically for 30 minutes then 1 mL of hexane was added followed by vigorous vortexing periodically for 30 minutes. After phase separation, 900  $\mu$ L of the hexane phase was transferred to a 2 mL glass vial and dried to completion. The sample was resuspended in 50  $\mu$ L of hexane followed by GC-FID analysis of 1  $\mu$ L using an HP-5 column with the following oven program: initial oven temperature held at 150°C for 1 minute followed by a 10°C/min ramp to 280°C then a 5°C/min ramp to 310°C which was held for 1 minute. Authentic farnesol, squalene, and botryococcene standards were analyzed with each sample set for comparison.

## Chapter Five: Concluding Remarks

Deposited in this dissertation are new layers of enzymology knowledge for triterpene synthases (Chapter 2) and the foundation for a directed-evolution, selection platform (Chapter 3) in yeast that should allow for unbiased studies of squalene synthase-like (SSL) enzymes. Highlights of Chapter 2 include the identification of N171 and G207 in SSL-3 that control product specificity for botryococcene or squalene. This mutant was exploited in Chapter 3 to develop a yeast selection platform that led to the identification of additional substitutions in positions N171 and G207 of SSL-3 that affect product specificity. Follow up work should include kinetic assessments of the mutant enzymes described in Chapter 2 and construction of SSL mutant enzyme libraries for screening with the selection platform described in Chapter 3 to identify SSL mutants converted to fully functional squalene synthases.

A very stimulating observation inadvertently stumbled upon in Chapter 3 was that under certain conditions, *Botryococcus braunii* squalene synthase (BbSS) could complement the squalene synthase knockout (*erg9Δ*) in *Saccharomyces cerevisiae*. Previous studies have suggested that non-fungal squalene synthases are unable to complement because they cannot associate with a metabolon involved in ergosterol formation.<sup>84–86</sup> While this idea of metabolon formation is still very possible, an interesting notion to entertain is that the carboxy-terminal region of squalene synthase may be more heavily involved in sterol homeostasis than previously thought. As discussed in Chapter 3, the effects of BbSS and ScSS that were observed under the different conditions tested were suggestive of a ‘toxic effect’ when too much squalene was fluxed into the ergosterol biosynthetic pathway. In a recent review by Wriessnegger and Pichler (2013), the effects of altering membrane sterol profiles in *S. cerevisiae* were discussed in relation to proper membrane function.<sup>142</sup> Several studies were pointed out that suggested perturbations to the sterol profile as a result of mutations in ergosterol biosynthetic machinery (*erg6Δ* and *erg2Δ*) caused negative consequences for membrane permeability, bilayer lipid packing, and osmotic stresses.<sup>143–145</sup> These observations are intriguing to consider if it is the case that too much squalene fluxed into the ergosterol biosynthetic pathway perturbs sterol profiles, thus leading to sub-optimal membrane properties and ultimately death.

Discussed in Chapter 4 is an effort to engineer triterpene metabolism in *Chlamydomonas reinhardtii* through traditional means of introducing biosynthetic enzymes. Mentioned in this chapter are the technological hurdles that need to be addressed before this alga and likely others become viable options for strain optimization as high-level production platforms for valuable renewable chemicals. Although not a solution for the transgene expression problem in *C. reinhardtii*, an approach developed for the production of novel terpenoids in *S. cerevisiae*<sup>89</sup> might be applicable in *C. reinhardtii* to provide a photosynthetic production platform for novel terpenes. While sterols themselves are not dispensable in *S. cerevisiae*, the biosynthetic pathway is because sterols can be supplied exogenously and taken up from the media. By eliminating the biosynthetic

pathway at the point of squalene synthase in *S. cerevisiae*, high levels of FPP can be accumulated which is available for conversion to terpenoids of interest. In *C. reinhardtii*, the main membrane sterol is also ergosterol<sup>146</sup> making the yeast methodology for strain development an interesting angle to consider. While it would not be exactly applicable to the *C. reinhardtii* system, the overall principle is similar and could offer an alternative approach to achieve high levels of FPP in *C. reinhardtii*. If successful, this system might promote expression of terpene synthase transgenes in *C. reinhardtii* to favor homeostasis of prenyl diphosphates. Nevertheless, pathway independent routes to achieve high FPP levels in *C. reinhardtii* will inevitably involve nuclear transgene expression.

## References

- (1) Zhang, Y.; Li, Y.; Guo, Y.; Jiang, H.; Shen, X. *Acta Pharmacol. Sin.* **2009**, *30*, 333.
- (2) Wang, G.; Tang, W.; Bidigare, R. R. In *Natural Products*; 2005; pp. 197–227.
- (3) Peralta-Yahya, P. P.; Ouellet, M.; Chan, R.; Mukhopadhyay, A.; Keasling, J. D.; Lee, T. S. *Nat. Commun.* **2011**, *2*, 483.
- (4) Wollam, J.; Antebi, A. *Annu. Rev. Biochem.* **2011**, *80*, 885.
- (5) Zhang, F. L.; Casey, P. J. *Annu. Rev. Biochem.* **1996**, *65*, 241.
- (6) Schmelz, E. a.; Huffaker, A.; Sims, J. W.; Christensen, S. a.; Lu, X.; Okada, K.; Peters, R. J. *Plant J.* **2014**, *Accepted*.
- (7) Do, R.; Kiss, R. S.; Gaudet, D.; Engert, J. C. *Clin. Genet.* **2009**, *75*, 19.
- (8) Liu, C.-I.; Liu, G. Y.; Song, Y.; Yin, F.; Hensler, M. E.; Jeng, W.-Y.; Nizet, V.; Wang, A. H.-J.; Oldfield, E. *Science* **2008**, *319*, 1391.
- (9) Metzger, P.; Largeau, C. *Appl. Microbiol. Biotechnol.* **2005**, *66*, 486.
- (10) Mastalerz, M.; Hower, J. C. *Org. Geochem.* **1996**, *24*, 301.
- (11) McKirdy, D. M.; Cox, R. E.; Volkman, J. K.; Howell, V. J. *Nature* **1986**, *320*, 57.
- (12) Moldowan, J. M.; Seifert, W. . *J. Chem. Soc. Chem. Commun.* **1980**, 912.
- (13) Hellier, P.; Al-Haj, L.; Talibi, M.; Purton, S.; Ladommatos, N. *Fuel* **2013**, *111*, 670.
- (14) Hillen, L. W.; Pollard, G.; Wake, L. V; White, N. *Biotechnol. Bioeng.* **1982**, *XXIV*, 193.
- (15) Tracy, N. I.; Crunkleton, D. W.; Price, G. L. *Biomass and Bioenergy* **2011**, *35*, 1060.
- (16) Werthemann, L.; Johnson, W. S. *Proc. Natl. Acad. Sci.* **1970**, *67*, 1465.
- (17) White, J. D.; Reddy, G. N.; Spessard, G. O. *J. Am. Chem. Soc.* **1988**, *110*, 1624.
- (18) Rilling, H. C. *J. Biol. Chem.* **1966**, *241*, 3233.
- (19) Radisky, E. S.; Poulter, C. D. *Biochemistry* **2000**, *39*, 1748.
- (20) Bohlmann, J.; Keeling, C. I. *Plant J.* **2008**, *54*, 656.
- (21) Gershenzon, J.; Dudareva, N. *Nat. Chem. Biol.* **2007**, *3*, 408.
- (22) Rilling, H. C.; Epstein, W. W. *J. Am. Chem. Soc.* **1969**, *91*, 1041.
- (23) Blagg, B. S. J.; Jarstfer, M. B.; Rogers, D. H.; Poulter, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 8846.
- (24) Jarstfer, M. B.; Zhang, D.-L.; Poulter, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 8834.
- (25) Liu, C.-I.; Jeng, W.-Y.; Chang, W.-J.; Shih, M.-F.; Ko, T.-P.; Wang, A. H.-J. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2014**, *70*, 231.
- (26) Lin, F.-Y.; Liu, C.-I.; Liu, Y.-L.; Zhang, Y.; Wang, K.; Jeng, W.-Y.; Ko, T.-P.; Cao, R.; Wang, A. H.-J.; Oldfield, E. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 21337.
- (27) Zhang, D.; Jennings, S. M.; Robinson, G. W.; Poulter, C. D. *Arch. Biochem. Biophys.* **1993**, *304*, 133.
- (28) Huang, Z.; Poulter, C. D. *J. Am. Chem. Soc.* **1989**, *111*, 2713.



- (29) Robinson, G. W.; Tsay, Y. H.; Kienzle, B. K.; Smith-Monroy, C. a; Bishop, R. W. *Mol. Cell. Biol.* **1993**, *13*, 2706.
- (30) Pandit, J.; Danley, D. E.; Schulte, G. K.; Mazzalupo, S.; Pauly, T. a; Hayward, C. M.; Hamanaka, E. S.; Thompson, J. F.; Harwood, H. J. *J. Biol. Chem.* **2000**, *275*, 30610.
- (31) Gu, P.; Ishii, Y.; Spencer, T. A.; Shechter, I. *J. Biol. Chem.* **1998**, *273*, 12515.
- (32) Pelz, A.; Wieland, K.-P.; Putzbach, K.; Hentschel, P.; Albert, K.; Götz, F. *J. Biol. Chem.* **2005**, *280*, 32493.
- (33) Wieland, B.; Feil, C.; Gloria-Maercker, E.; Thumm, G.; Lechner, M.; Bravo, J. M.; Poralla, K.; Götz, F. *J. Bacteriol.* **1994**, *176*, 7719.
- (34) Okada, S.; Devarenne, P.; Chappell, J. *Arch. Biochem. Biophys.* **2000**, *373*, 307.
- (35) Okada, S.; Devarenne, T. P.; Murakami, M.; Abe, H.; Chappell, J. *Arch. Biochem. Biophys.* **2004**, *422*, 110.
- (36) Niehaus, T. D.; Okada, S.; Devarenne, T. P.; Watt, D. S.; Sviripa, V.; Chappell, J. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 12260.
- (37) Mookhtiar, K. A.; Kalinowski, S. S.; Zhang, D.; Poulter, C. D. *J. Biol. Chem.* **1994**, *269*, 11201.
- (38) Lin, F.-Y.; Zhang, Y.; Hensler, M.; Liu, Y.-L.; Chow, O. A.; Zhu, W.; Kang, K.; Pang, R.; Thienphrapa, W.; Nizet, V.; Oldfield, E. *ChemMedChem* **2012**, *7*, 561.
- (39) Li, R.; Chou, W. K. W.; Himmelberger, J. a; Litwin, K. M.; Harris, G. G.; Cane, D. E.; Christianson, D. W. *Biochemistry* **2014**, *53*, 1155.
- (40) Nations, U. *World Population Prospects The 2012 Revision Methodology of the United Nations*; 2014.
- (41) Georgianna, D. R.; Mayfield, S. P. *Nature* **2012**, *488*, 329.
- (42) Ashokkumar, V.; Rengasamy, R. *Bioresour. Technol.* **2012**, *104*, 394.
- (43) Gómez, P. I.; Inostroza, I.; Pizarro, M.; Pérez, J. *AoB Plants* **2013**, *5*, plt026.
- (44) Gimpel, J. a; Specht, E. a; Georgianna, D. R.; Mayfield, S. P. *Curr. Opin. Chem. Biol.* **2013**, *17*, 489.
- (45) Disch, A.; Schwender, J.; Müller, C.; Lichtenthaler, H. K.; Rohmer, M. *Biochem. J.* **1998**, *333*, 381.
- (46) Merchant, S. S.; Prochnik, S. E.; Vallon, O.; Harris, E. H.; Karpowicz, S. J.; Witman, G. B.; Terry, A.; Salamov, A.; Fritz-Laylin, L. K.; Maréchal-Drouard, L.; Marshall, W. F.; Qu, L.-H.; Nelson, D. R.; Sanderfoot, A. a; Spalding, M. H.; Kapitonov, V. V; Ren, Q.; Ferris, P.; Lindquist, E.; Shapiro, H.; Lucas, S. M.; Grimwood, J.; Schmutz, J.; Cardol, P.; Cerutti, H.; Chanfreau, G.; Chen, C.-L.; Cognat, V.; Croft, M. T.; Dent, R.; Dutcher, S.; Fernández, E.; Fukuzawa, H.; González-Ballester, D.; González-Halphen, D.; Hallmann, A.; Hanikenne, M.; Hippler, M.; Inwood, W.; Jabbari, K.; Kalanon, M.; Kuras, R.; Lefebvre, P. a; Lemaire, S. D.; Lobanov, A. V; Lohr, M.; Manuell, A.; Meier, I.; Mets, L.; Mittag, M.; Mittelmeier, T.; Moroney, J. V; Moseley, J.; Napoli, C.; Nedelcu, A. M.; Niyogi, K.; Novoselov, S. V; Paulsen, I. T.; Pazour, G.; Purton, S.; Ral, J.-P.; Riaño-

- Pachón, D. M.; Riekhof, W.; Rymarquis, L.; Schroda, M.; Stern, D.; Umen, J.; Willows, R.; Wilson, N.; Zimmer, S. L.; Allmer, J.; Balk, J.; Bisova, K.; Chen, C.-J.; Elias, M.; Gendler, K.; Hauser, C.; Lamb, M. R.; Ledford, H.; Long, J. C.; Minagawa, J.; Page, M. D.; Pan, J.; Pootakham, W.; Roje, S.; Rose, A.; Stahlberg, E.; Terauchi, A. M.; Yang, P.; Ball, S.; Bowler, C.; Dieckmann, C. L.; Gladyshev, V. N.; Green, P.; Jorgensen, R.; Mayfield, S.; Mueller-Roeber, B.; Rajamani, S.; Sayre, R. T.; Brokstein, P.; Dubchak, I.; Goodstein, D.; Hornick, L.; Huang, Y. W.; Jhaveri, J.; Luo, Y.; Martínez, D.; Ngau, W. C. A.; Otilar, B.; Poliakov, A.; Porter, A.; Szajkowski, L.; Werner, G.; Zhou, K.; Grigoriev, I. V.; Rokhsar, D. S.; Grossman, A. R. *Science* **2007**, *318*, 245.
- (47) Grossman, A. R.; Lohr, M.; Im, C. S. *Annu. Rev. Genet.* **2004**, *38*, 119.
- (48) Hemmerlin, A.; Harwood, J. L.; Bach, T. J. *Prog. Lipid Res.* **2012**, *51*, 95.
- (49) Lichtenthaler, H. K. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 47.
- (50) Wu, S.; Jiang, Z.; Kempinski, C.; Eric Nybo, S.; Husodo, S.; Williams, R.; Chappell, J. *Planta* **2012**, *236*, 867.
- (51) Wu, S.; Schalk, M.; Clark, A.; Miles, R. B.; Coates, R.; Chappell, J. *Nat. Biotechnol.* **2006**, *24*, 1441.
- (52) Hannich, J. T.; Umebayashi, K.; Riezman, H. *Cold Spring Harb. Perspect. Biol.* **2011**, *3*.
- (53) Siedenburg, G.; Jendrossek, D. *Appl. Environ. Microbiol.* **2011**, *77*, 3905.
- (54) Jarstfer, M. B.; Blagg, B. S. J.; Rogers, D. H.; Poulter, C. D. *J. Am. Chem. Soc.* **1996**, *118*, 13089.
- (55) Epstein, W. W.; Rilling, H. C. *J. Biol. Chem.* **1970**, *245*, 4597.
- (56) Pan, J.-J.; Bugni, T. S.; Poulter, C. D. *J. Org. Chem.* **2009**, *74*, 7562.
- (57) Liu, C.-I.; Jeng, W.-Y.; Chang, W.-J.; Ko, T.-P.; Wang, A. H.-J. *J. Biol. Chem.* **2012**, *287*, 18750.
- (58) Poulter, C. D.; Marsh, L. L.; Hughes, J. M.; Argyle, J. C.; Satterwhite, D. M.; Goodfellow, R. J.; Moesinger, S. G. *J. Am. Chem. Soc.* **1977**, *99*, 3816.
- (59) Poulter, C. D. *Acc. Chem. Res.* **1990**, *23*, 70.
- (60) Huang, Z.; Poulter, C. D.; Wolf, F. R.; Somers, T. C.; Whites, J. D. *J. Am. Chem. Soc.* **1988**, *110*, 3959.
- (61) Okada, S.; Devarenne, T. P.; Chappell, J. *Arch. Biochem. Biophys.* **2000**, *373*, 307.
- (62) Eroglu, E.; Okada, S.; Melis, A. *J. Appl. Phycol.* **2011**, *23*, 763.
- (63) Sali, A.; Potterton, L.; Yuan, F.; van Vlijmen, H.; Karplus, M. *Proteins* **1995**, *23*, 318.
- (64) Rogers, D. H.; Yi, E. C.; Poulter, C. D. *J. Org. Chem.* **1995**, *60*, 941.
- (65) Liu, C.-I.; Jeng, W.-Y.; Chang, W.-J.; Ko, T.-P.; Wang, A. H.-J. *J. Biol. Chem.* **2012**, *287*, 18750.
- (66) Greenhagen, B. T.; Griggs, P.; Takahashi, S.; Ralston, L.; Chappell, J. *Arch. Biochem. Biophys.* **2003**, *409*, 385.
- (67) Busquets, A.; Keim, V.; Closa, M.; del Arco, A.; Boronat, A.; Arró, M.; Ferrer, A. *Plant Mol. Biol.* **2008**, *67*, 25.
- (68) Jones, G.; Willett, P.; Glen, R. C. *J. Mol. Biol.* **1995**, *245*, 43.

- (69) Schüttelkopf, A. W.; van Aalten, D. M. F. *Acta Crystallogr. D. Biol. Crystallogr.* **2004**, *60*, 1355.
- (70) Higuchi, R.; Krummel, B.; Saiki, R. K. *Nucleic Acids Res.* **1988**, *16*, 7351.
- (71) Kovach, M. E.; Elzer, P. H.; Hill, D. S.; Robertson, G. T.; Farris, M. a; Roop, R. M.; Peterson, K. M. *Gene* **1995**, *166*, 175.
- (72) Ind, A. C.; Porter, S. L.; Brown, M. T.; Byles, E. D.; de Beyer, J. a; Godfrey, S. a; Armitage, J. P. *Appl. Environ. Microbiol.* **2009**, *75*, 6613.
- (73) Shine, J.; Dalgarno, L. *Nature* **1975**, *254*, 34.
- (74) Kuswik-rabiegas, G.; Rillings, H. C. *J. Biol. Chem.* **1987**, *262*, 1505.
- (75) Nakashima, T.; Inoue, T.; Oka, A.; Nishino, T.; Osumi, T.; Hata, S. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, *92*, 2328.
- (76) Karst, F.; Lacroute, F. *Mol. Genet. genomics* **1977**, *154*, 269.
- (77) Tozawa, R.; Ishibashi, S.; Osuga, J.; Yagyu, H.; Oka, T.; Chen, Z.; Ohashi, K.; Perrey, S.; Shionoiri, F.; Yahagi, N.; Harada, K.; Gotoda, T.; Yazaki, Y.; Yamada, N. *J. Biol. Chem.* **1999**, *274*, 30843.
- (78) De Souza, W.; Rodrigues, J. C. F. *Interdiscip. Perspect. Infect. Dis.* **2009**, *2009*, 642502.
- (79) Fox, C. B. *Molecules* **2009**, *14*, 3286.
- (80) Allison, A. C. *Methods* **1999**, *19*, 87.
- (81) Tracy, N. I.; Crunkleton, D. W.; Price, G. L. *Biomass and Bioenergy* **2011**, *35*, 1060.
- (82) Beytia, E.; Qureshi, A. A.; John, W. *J. Biol. Chem.* **1973**, *248*, 1856.
- (83) Jennings, S. M.; Tsay, Y. H.; Fisch, T. M.; Robinson, G. W. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 6038.
- (84) Niehaus, T. D. *Theses Diss. - Plant Soil Sci.* **2011**, *1*.
- (85) Robinson, G. W.; Tsay, Y. H.; Kienzle, B. K.; Smith-Monroy, C. a; Bishop, R. W. *Mol. Cell. Biol.* **1993**, *13*, 2706.
- (86) Kribii, R.; Arro, M.; Del Arco, A.; Gonzalez, V.; Balcells, L.; Delourme, D.; Ferrer, A.; Karst, F.; Boronat, A. *Eur. J. Biochem.* **1997**, *249*, 61.
- (87) Fegueur, M.; Richard, L.; Charles, a D.; Karst, F. *Curr. Genet.* **1991**, *20*, 365.
- (88) Stamellos, K. D.; Shackelford, J. E.; Shechterp, I.; Jiang, G.; Conrad, D.; Keller, G.; Krisans, S. K. *J. Biol. Chem.* **1993**, *268*, 12825.
- (89) Zhuang, X. *Theses Diss. - Plant Soil Sci.* **2013**, *1*.
- (90) Janke, C.; Magiera, M. M.; Rathfelder, N.; Taxis, C.; Reber, S.; Maekawa, H.; Moreno-Borchart, A.; Doenges, G.; Schwob, E.; Schiebel, E.; Knop, M. *Yeast* **2004**, *21*, 947.
- (91) Meyer, V.; Arentshorst, M.; Flitter, S. J.; Nitsche, B. M.; Kwon, M. J.; Reynaga-Peña, C. G.; Bartnicki-Garcia, S.; van den Hondel, C. a M. J. J.; Ram, A. F. *J. Eukaryot. Cell* **2009**, *8*, 1677.
- (92) Madeo, F.; Fröhlich, E.; Fröhlich, K. *J. Cell Biol.* **1997**, *139*, 729.
- (93) Goldsmith, M.; Tawfik, D. S. *Curr. Opin. Struct. Biol.* **2012**, *22*, 406.
- (94) Guan, X. L.; Riezman, I.; Wenk, M. R.; Riezman, H. *Methods Enzymol.* **2010**, *470*, 369.
- (95) Khatri, W.; Hendrix, R.; Niehaus, T.; Chappell, J.; Curtis, W. R. *Biotechnol. Bioeng.* **2014**, *111*, 493.

- (96) Potvin, G.; Zhang, Z. *Biotechnol. Adv.* **2010**, *28*, 910.
- (97) Wang, B.; Wang, J.; Zhang, W.; Meldrum, D. R. *Front. Microbiol.* **2012**, *3*, 344.
- (98) Tam, L.-W.; Lefebvre, P. A. *Genetics* **1993**, *135*, 375.
- (99) Kindle, K. L.; Schnell, R. A.; Fernandez, E.; Lefebvre, P. A. *J. Cell Biol.* **1989**, *109*, 2589.
- (100) Sodeinde, O. A.; Kindle, K. L. *Proc. Natl. Acad. Sci.* **1993**, *90*, 9199.
- (101) Purton, S.; Szaub, J. B.; Wannathong, T.; Young, R.; Economou, C. K. *Russ. J. Plant Physiol.* **2013**, *60*, 491.
- (102) Baroli, I.; Do, A. D.; Yamane, T.; Niyogi, K. K. *Plant Cell* **2003**, *15*, 992.
- (103) Mayfield, S. P.; Manuell, A. L.; Chen, S.; Wu, J.; Tran, M.; Siefker, D.; Muto, M.; Marin-navarro, J. *Curr. Opin. Biotechnol.* **2007**, *18*, 1.
- (104) Kumar, A.; Falcao, V. R.; Sayre, R. T. *Algal Res.* **2013**, *2*, 321.
- (105) Rasala, B. a; Barrera, D. J.; Ng, J.; Plucinak, T. M.; Rosenberg, J. N.; Weeks, D. P.; Oyler, G. a; Peterson, T. C.; Haerizadeh, F.; Mayfield, S. P. *Plant J.* **2013**, *74*, 545.
- (106) Strenkert, D.; Schmollinger, S.; Schroda, M. *Nucleic Acids Res.* **2013**, *41*, 5273.
- (107) Fischer, N.; Rochaix, J.-D. *Mol. Genet. Genomics* **2001**, *265*, 888.
- (108) Rasala, B. a; Lee, P. a; Shen, Z.; Briggs, S. P.; Mendez, M.; Mayfield, S. P. *PLoS One* **2012**, *7*, e43349.
- (109) Ryan, M. D.; King, a M.; Thomas, G. P. *J. Gen. Virol.* **1991**, *72*, 2727.
- (110) Neupert, J.; Karcher, D.; Bock, R. *Plant J.* **2009**, *57*, 1140.
- (111) Kong, F.; Yamasaki, T.; Ohama, T. *J. Biosci. Bioeng.* **2013**, *xx*, 1.
- (112) Hyka, P.; Lickova, S.; Přibyl, P.; Melzoch, K.; Kovar, K. *Biotechnol. Adv.* **2012**, *31*, 2.
- (113) Steinbrenner, J.; Sandmann, G. *Appl. Environ. Microbiol.* **2006**, *72*, 7477.
- (114) León, R.; Couso, I.; Fernández, E. *J. Biotechnol.* **2007**, *130*, 143.
- (115) Cordero, B. F.; Couso, I.; León, R.; Rodríguez, H.; Vargas, M. A. *Appl. Microbiol. Biotechnol.* **2011**, *91*, 341.
- (116) Couso, I.; Vila, M.; Rodríguez, H.; Vargas, M. a; León, R. *Biotechnol. Prog.* **2011**, *27*, 54.
- (117) Couso, I.; Cordero, B. F.; Vargas, M. Á.; Rodríguez, H. *Mar. Drugs* **2012**, *10*, 1955.
- (118) Liu, J.; Gerken, H.; Huang, J.; Chen, F. *Process Biochem.* **2013**, *48*, 788.
- (119) Wu, S.; Schalk, M.; Clark, A.; Miles, R. B.; Coates, R.; Chappell, J. *Nat. Biotechnol.* **2006**, *24*, 1441.
- (120) Wu, S.; Jiang, Z.; Kempinski, C.; Nybo, S. E.; Husodo, S.; Williams, R.; Chappell, J. *Planta* **2012**, *236*.
- (121) Lohr, M.; Schwender, J.; Polle, J. E. W. *Plant Sci.* **2012**, *185-186*, 9.
- (122) Grossman, A. R.; Lohr, M.; Im, C. S. *Annu. Rev. Genet.* **2004**, *38*, 119.
- (123) Heitzer, M.; Zschoernig, B. *Biotechniques* **2007**, *43*, 324.
- (124) Sizova, I.; Fuhrmann, M.; Hegemann, P. *Gene* **2001**, *277*, 221.
- (125) Hyams, J.; Davies, D. R. *Mutat. Res.* **1972**, *14*, 381.
- (126) Shimogawara, K.; Fujiwara, S.; Grossman, A.; Usuda, H. *Genetics* **1998**, *148*, 1821.

- (127) Walker, T. L.; Becker, D. K.; Collet, C. *Plant Cell Rep.* **2005**, *23*, 727.
- (128) Ramos, A. a.; Polle, J.; Tran, D.; Cushman, J. C.; Jin, E.-S.; Varela, J. C. *Algae* **2011**, *26*, 3.
- (129) Suzuki, R.; Ito, N.; Uno, Y.; Nishii, I.; Kagiwada, S.; Okada, S.; Noguchi, T. *PLoS One* **2013**, *8*, e81626.
- (130) Weiss, T. L.; Roth, R.; Goodson, C.; Vitha, S.; Black, I.; Azadi, P.; Rusch, J.; Holzenburg, A.; Devarenne, T. P.; Goodenough, U. *Eukaryot. Cell* **2012**, *11*, 1424.
- (131) Cifuentes, A. S.; Gonzalez, M.; Conejeros, M.; Dellarossa, V.; Parra, O. *J. Appl. Phycol.* **1992**, *4*, 111.
- (132) Lamers, P. P.; van de Laak, C. C. W.; Kaasenbrood, P. S.; Lorier, J.; Janssen, M.; De Vos, R. C. H.; Bino, R. J.; Wijffels, R. H. *Biotechnol. Bioeng.* **2010**, *106*, 638.
- (133) Rabbani, S.; Beyer, P.; Lintig, J.; Hugueney, P.; Kleinig, H. *Plant Physiol.* **1998**, *116*, 1239.
- (134) Davidi, L.; Katz, A.; Pick, U. *Planta* **2012**, *236*, 19.
- (135) Davies, F. K.; Jinkerson, R. E.; Posewitz, M. C. *Photosynth. Res.* **2014**.
- (136) Wang, D.; Ning, K.; Li, J.; Hu, J.; Han, D.; Wang, H.; Zeng, X.; Jing, X.; Zhou, Q.; Su, X.; Chang, X.; Wang, A.; Wang, W.; Jia, J.; Wei, L.; Xin, Y.; Qiao, Y.; Huang, R.; Chen, J.; Han, B.; Yoon, K.; Hill, R. T.; Zohar, Y.; Chen, F.; Hu, Q.; Xu, J. *PLoS Genet.* **2014**, *10*, e1004094.
- (137) Matsushima, D.; Jenke-Kodama, H.; Sato, Y.; Fukunaga, Y.; Sumimoto, K.; Kuzuyama, T.; Matsunaga, S.; Okada, S. *Plant Sci.* **2012**, *185-186*, 309.
- (138) Mayfield, S. *Genome* **2013**, *56*, 551.
- (139) Sabourin, M.; Tuzon, C. T.; Fisher, T. S.; Zakian, V. A. *Yeast* **2007**, *24*, 39.
- (140) Gorman, D. S.; Levine, R. P. *Proc. Natl. Acad. Sci.* **1965**, *54*, 1665.
- (141) Guillard, R. R. L.; Ryther, J. H. *Can. J. Microbiol.* **1962**, *8*, 229.
- (142) Wriessnegger, T.; Pichler, H. *Prog. Lipid Res.* **2013**, *52*, 277.
- (143) Sharma, S. C. *FEMS Yeast Res.* **2006**, *6*, 1047.
- (144) Abe, F.; Hiraki, T. *Biochim. Biophys. Acta* **2009**, *1788*, 743.
- (145) Dupont, S.; Beney, L.; Ferreira, T.; Gervais, P. *Biochim. Biophys. Acta* **2011**, *1808*, 1520.
- (146) Gealt, M. A.; Adler, J. H.; Nes, W. R. *Lipids* **1981**, *16*, 133.

## Vita

1. Place of Birth: Owensboro, KY
2. Education: Bachelor of Science in Biology at the University of Kentucky, December, 2008
3. Professional Positions: None held
4. Scholastic Honors: Departmental Honors in Biology at the University of Kentucky; University Honors, *Magna Cum Laude*, at the University of Kentucky
5. Publications:

Niehaus, T.D., Kinison, S., Okada, S., Yeo, Y., Bell, S.A., Cui, P., Devarenne, T.P., Chappell, J. Functional identification of triterpene methyltransferases from *Botryococcus braunii* race B. J. Biol. Chem. 287(11):8163-8173.

Bell, S.A. and Hunt, A.G. The Arabidopsis ortholog of the 77 kDa subunit of the cleavage stimulatory factor (AtCstF-77) involved in mRNA polyadenylation is an RNA-binding protein. FEBS Lett. 584(8):1449-1454.

6. Stephen A. Bell