UNIVERSITY OF CALIFORNIA, SAN DIEGO

Investigating fatty acid biosynthesis within the algal chloroplast using *Chlamydomonas reinhardtii* as a model

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Chemistry

by

Jillian L. Blatti

Committee in charge:

Professor Michael D. Burkart, Chair
Professor Elizabeth Komives
Professor Stephen P. Mayfield
Professor Emmanuel Theodorakis
Professor Jerry Yang

2012

The Dissertation of Jillian L. Blatti is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2012

DEDICATION

I dedicate this dissertation to my love, Jeff.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

KS, ketosynthase
TE, thioesterase
ACP, acyl carrier protein
DH, dehydratase
ER, enoyl reductase
KR, ketoreductase
FAD, fatty acid desaturase
MAT, malonyl transferase
AT, acyl transferase
ACCase, acetyl CoA carboxylase
PPTase, phosphopantetheinyl transferase
CoA, coenzyme A
FAS, fatty acid synthase
Cr, *Chlamydomonas reinhardtii*
Pt, *Phaeodactylum tricornutum*
Hp*, Haematococcus pluvialis*
At, *Arabidopsis thaliana*
Uc*, Umbellularia californica*
Ch, *Cuphea hookeriana*
Cc, *Cinnamomum camphorum*
Rc, *Ricinus communis*
Ec, *Escherichia coli*
Vh, *Vibrio harveyi*
TAG, triacylglycerol
FA, fatty acid
FAME, fatty acid methyl ester
DMF, dimethyl formamide
DMSO, dimethyl sulfoxide
EtOAc, ethyl acetate
MeOH, methanol
GC/MS, gas chromatography/ mass spectrometry
NMR, nuclear magnetic resonance
MS, mass spectrometry

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Mike Burkart for the opportunity to work in his laboratory – I truly appreciate the freedom he gave me to pursue my ideas and interests. I must also express my gratitude to Steve Mayfield for all of his help, guidance, and expertise in working with Chlamy. Barbara Sawrey has been an excellent mentor and friend, and I am grateful for all of our conversations about science education. Bill Baker first introduced me to natural products organic chemistry and invited me to work in his laboratory, and for that I am forever indebted. Sheryl Tsai has been very helpful and supportive in a collaborative project that started at the beginning of this year, along with Grace Festin. Joris Beld was often a scientific comrade in the laboratory, and many discussions and experiments ensued. He was always helpful around the lab and his assistance is always appreciated.

I thank my friends and family for all of their love.

I dedicate this dissertation to Jeff, for all of his love, patience, and support. Jeff has been by my side throughout – he is my best friend and my soulmate.

Chapter 1 is adapted from a review article currently being prepared to be published in Current Opinion in Chemical Biology as part of an upcoming Bioenergy series. Professors Stephen Mayfield and Michael Burkart have contributed words, thoughts, expert ideas and opinions to the article.

Chapter 2 is adapted from a recently submitted article to Plos One, on which I

was the primary author. Joris Beld assisted with the *in silico* modeling and *in vitro* crosslinking experiments, including protein expression and purification and synthesis of chemical probes. Craig Behnke and Mike Mendez assisted with engineering of TEs into the *C. reinhardtii* chloroplast. Yongzuan Su (UCSD Mass Spectrometry Facility) helped with mass spectrometric analsysis.

Chapter 3 is adapted from a manuscript currently being prepared for publication in Algal Research, on which I am the primary author. Joris Beld assisted with the *in silico* work and *in vitro* crosslinking experiments. Craig Behnke and Mike Mendez assisted with engineering of KS into the *C. reinhardtii* chloroplast. Professor John Cronan provided *E. coli* strain C1877 and Professor Byers supplied the *Vibrio harveyi* ACP construct.

Chapter 4 is adapted from an article in the Journal of Chemical Education, on which I was the primary and corresponding author. The curriculum was piloted and implemented at Abraham Lincoln High School in San Diego in the classroom of Janelle Javier, and the results published as a laboratory experiment. The project was supported by an NSF-GK12 fellowship (NSF Grant 0742551, Principal Investigator Maarten J. Chrispeels), the Department of Chemistry and Biochemistry at the University of California San Diego, and SD-CAB. Stephen P. Mayfield (University of California, San Diego) provided *C. reinhardtii* strain *csc*137c.

VITA

EDUCATIONAL BACKGROUND

2007-2012	Ph.D. in Chemistry University of California, San Diego Principle Investigator: Michael D. Burkart
2007-2009	M.S. in Chemistry University of California, San Diego Principle Investigator: Michael D. Burkart
2001-2004	B.S. in Biomedical Sciences University of South Florida, Tampa, Florida Principle Investigator: Bill J. Baker

TEACHING EXPERIENCE

2010-2011	Biochemistry Curriculum Development National Science Foundation Graduate Fellowship San Diego Unified School District, California Curriculum available in *J. Chem. Educ.* 2012, *89* (2), 239-242, *SI*
2009-2010	Preparing Professional Faculty Program University of California, San Diego
2009 and 2010 (W)	UC San Diego Lecturer in the Renewable Energy Series Chemistry 15 [UC San Diego science series for non-majors]
2007-present	Chemistry and Biochemistry Tutor Department of Chemistry & Biochemistry, UC San Diego
2007-2011	Teaching Assistant Department of Chemistry & Biochemistry, UC San Diego Organic Chemistry; Organic Chemistry Laboratory; Chemistry for non-science majors; Biochemistry for non-science majors; Biochemistry
2006-2007	Mathematics Education at Harvard-Westlake Preparatory School Los Angeles, California [Teacher/Tutor]
2006-2007	Physical Education at the Buckley School Los Angeles, California
2006-2007	Mathematics Education at West Coast Nursing University Los Angeles, California [Teacher/Tutor]
2004-2006	KAPLAN Test Preparation Lecturer/Tutor

PUBLICATIONS

- Blatti, J.L., Mayfield, S. P. and Burkart, M.D. Engineering fatty acid biosynthesis in microalgae for sustainable biodiesel. (2012)
- Blatti, J.L., Beld, J., Behnke, C., Mendez, M., Mayfield, S. P. and Burkart, M.D. Evolution of fatty acid synthase enzymes dictates interchangeability. (2012)
- Blatti, J.L., Beld, J., Behnke, C., Mendez, M., Mayfield, S. P. and Burkart, M.D. Manipulation of fatty acid biosynthesis in microalgae for biofuel through protein-protein interactions. (2012)
- Blatti, J.L. and Burkart, M.D. Releasing stored solar energy within pond scum: Biodiesel from algal lipids. *J. Chem. Educ.* **2012**, *89* (2), 239-242.
-

SELECTED PRESENTATIONS

- **Algae Biofuel Symposium: Food and Fuel for the 21st Century**. San Diego, California. May 11, 2012. *Research poster presentation*: "Evolution of fatty acid synthase enzymes dictates interchangeability towards manipulating algal fatty acid biosynthesis."

- **American Chemical Society 243rd National Meeting**. San Diego, California. March 29, 2012. CHED Division of Chemical Education. *Oral chemical education research presentation*: "Releasing stored solar energy within pond scum: biodiesel from algal lipids."

- **American Chemical Society 43rd Western Regional Meeting**. Pasadena, California. November 11, 2011. BIOT001 Biotechnological Approaches to Energy. Biofuel Feedstocks and Production Processes. *Oral research presentation*: "Manipulation of fatty acid biosynthesis in microalgae through protein-protein interactions."

- **UC San Diego Forum on Energy Research.** *Research poster presentation*: 'San Diego Center for Algal Biotechnology (SD-CAB).' May 25, 2011

- **SD-CAB Student and Post-doctoral Symposium.** Scripps Institute of Oceanography, La Jolla, California. *Oral research presentation*: 'Manipulating fatty acid biosynthesis in eukaryotic microalgae.' February 20, 2011

- **Algae Biofuels Symposium Invited Lecture.** Southwestern Community College, California 'Microalgae as biofuel cell factories.' November 7, 2009

- **UC San Diego Howard Hughes Medical Institute Invited Lecture.** 'Genetic engineering of algae for renewable energy.' July 3, 2009

OUTREACH AND EDUCATION

05/10 - 05/11	Scientific Advisor at the Greater San Diego Science and Engineering Fair
11/09/09	Guest Speaker at Valhalla High School, El Cajon, California
08/13/10	Lab tour and demonstration for the Ocean Discovery Institute, San Diego
06/12/10	Career Day at Correia Middle School, Ocean Beach, CA
03/13/11 – 03/15/11	National Science Teachers Association (NSTA) Conference, San Francisco
03/02/10 – 03/09/10	UC San Diego BioBridge 'Scientist in Residency' with Sapphire Energy
03/06/10	Expanding Your Horizons: Women in Science, University of San Diego
01/23/10	Biofuel Outreach Lecture on Algal Culturing & Metabolism, BAAN, UCSD
12/10/09	NSF-GK12 Guest Lecture / Demonstration: 'Algae biofuels'
05/09 – 07/09	UC San Diego BioBridge Lecture for the 'Scientific Leadership Institute'

ABSTRACT OF THE DISSERTATION

Investigating fatty acid biosynthesis within the algal chloroplast using *Chlamydomonas reinhardtii* as a model

by

Jillian L. Blatti

Doctor of Philosophy in Chemistry

University of California, San Diego, 2012

Professor Michael D. Burkart, Chair

As finite petroleum reserves run their course and combustion-related CO_2 emissions rise concerns about global warming, humanity is faced with the challenge of finding new sources of energy that are carbon-neutral, renewable and sustainable to meet the growing demand. Photosynthetic organisms convert solar energy and CO_2 directly into metabolic products that can serve as fungible biofuels. Microalgae are particularly attractive as a biodiesel feedstock, as they produce oil in high yields, grow at fast rates in habitats not suitable for conventional agriculture, and do not compete

with the food supply. However, oil accumulation occurs under environmental stress, which compromises biomass productivity, and algal fatty acids are not ideal for biodiesel quality. The ability to manipulate algal fatty acid biosynthesis would thus be a significant stride towards developing algae as a biodiesel feedstock.

In fatty acid biosynthesis within an algal chloroplast, an acyl carrier protein (ACP) tethers the growing fatty acid as it undergoes iterative cycles of elongation, and a thioesterase (TE) domain catalyzes the release of a mature fatty acid from the ACP. As plant TEs specific for certain chain length fatty acids have altered the fatty acid profile of transgenic plants and bacteria, this has emerged as a promising strategy to modify algal fatty acid content to fashion an optimized biodiesel feedstock.

The work outlined in this thesis aims to investigate intermolecular interactions in algal fatty acid biosynthesis to facilitate engineering. A novel strategy was employed in which a chemical probe inspired by the enzymatic activity of the algal TE was synthesized, attached to the algal ACP chemoenzymatically, and used to trap algal ACP-TE interactions *in vitro*. No protein-protein interactions were detected between plant TEs and the algal ACP *in vitro*, and thus plant TEs did not elicit the desired phenotype when engineered into the algal chloroplast. Using protein-protein interactions as a means to control product identity may shift the paradigm towards rationally designed engineering approaches to optimize algae as a bioenergy source.

Renewable energy outreach and education has been an indispensable facet of this work to generate awareness and instill passion for sustainable energy in our future scientists.

Chapter 1 : Engineering fatty acid biosynthesis in microalgae for sustainable biodiesel

Abstract

With dwindling petroleum reserves, and combustion-related CO_2 emissions infiltrating the atmosphere causing a greenhouse effect, we are faced with the challenge of creating new sources of fuel that are renewable, sustainable, and carbon-neutral. Microalgae are emerging as a promising feedstock for renewable biofuels due to their high photosynthetic efficiency, fast growth rate, vast lipid yields, and ability to grow in a broad range of environments. Microalgae sequester CO_2 and recycle nutrients in wastewater, both attractive features in a biofuel feedstock. Furthermore, microalgae do not compete for arable land or the food supply, the main drawbacks of plant-based biofuels. Of the range of algal metabolites that are being pursued for bioenergy (i.e. bio-H_2, bioethanol, and biodiesel), biodiesel from algal lipids is of the most promising due to its high energy density and compliance with the current diesel infrastructure. Although microalgae are known to accumulate lipids, this is often due to environmental stress, which hinders growth and reproduction. Indeed, there is a tradeoff between lipid production and overall productivity. Additionally, algal lipids are composed of fatty acids that are not ideal for biodiesel properties. Metabolic engineering of algal fatty acid biosynthesis promises to create strains capable of fungible fuels on scale in a sustainable fashion.

Fatty acid biosynthesis proceeds through catalytic activity of seven discrete domains, all of which must interact with an acyl carrier protein (ACP) as it tethers the growing fatty acid intermediate. Engineering algal fatty acid metabolism necessitates a thorough characterization of the algal fatty acid synthase (FAS), including the timing and regulation of catalytic domains and their interactions with ACP. Compatibility

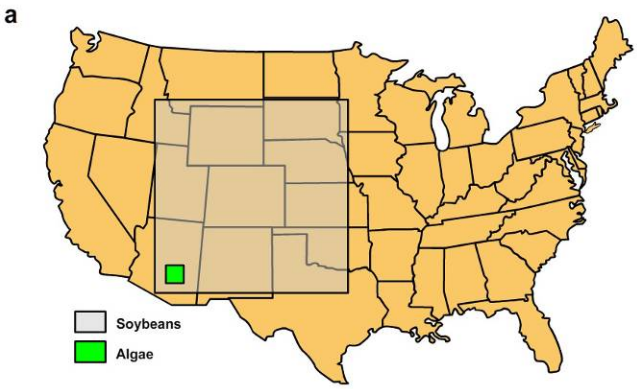
between heterologously expressed FAS enzymes and the native algal machinery must be considered. Here, we focus on efforts to engineer algal FAS and include a brief review of metabolic engineering technologies currently available for algae.

Introduction

Since the Industrial Revolution in the 1800s, human beings have relied on cheap, readily available sources of energy to fuel our growing society, which has drastically improved the quality of our lives [1]. Petroleum, derived from ancient fossilized algae [2], has served as this energy source, abundant at the time of its discovery in the mid-1800s. However, the increasing world population coupled to the desire of countries to become affluent (directly related the amount of energy they possess) is depleting petroleum reserves faster than anyone could have envisioned, and it is predicted that they will be consumed within the next 100 years [3]. Therefore, it is necessary to fashion renewable fuels to replace these traditional fossil fuels. We now know that combustion-related CO_2 emissions are contributing to the warming of the planet [4], and this adds a requisite property to the new fuel source that we create – it must sequester CO_2 from the environment. Photosynthetic organisms, such as plants, algae, and cyanobacteria, are at the forefront of research aimed at creating new sources of renewable energy because they efficiently convert solar energy and CO_2 into biofuels.

There are challenges that exist for each photosynthetic biofuel feedstock, summarized in Table 1-1. For instance, oleaginous plant species such as the physic nut (*Jatropha curcas*) have high oil content (up to 40%), but they have a slow growth rate

(germinate seeds after two years), long perennial life cycles, and genetic manipulations are difficult, creating an economic barrier [5]. Biotechnology companies such as SG Biofuels are breeding *Jatropha* to solve these concerns. Photosynthetic prokaryotes, such as cyanobacteria (i.e. *Synechococcus elongatus*), are easier to manipulate genetically [6] [7] and can be engineered to secrete oils [8] alleviating costs associated with biomass harvesting, but their low oil content hampers overall productivity and perhaps economic viability [9]. Companies such as Joule Unlimited and Synthetic Genomics are engineering cyanobacteria to overcome some of these challenges.



b

Crop Oil Yield	(Gallons/Acre/Yr)
Soybean	48
Camelina	62
Sunflower	102
Jatropha	202
Oil palm	635
Algae	1,000-4,000

Figure 1-1 Land requirement and yields of biomass feedstocks

a) Amount of land required to meet 50% of the current petroleum consumption using soybean (gray) and algae (green). b) Oil yields of leading biomass feedstocks. (Adapted from the National Algal Biofuels Technology Roadmap, U.S. Department of Energy) [79].

Eukaryotic microalgae are unicellular photosynthetic eukaryotes that are of interest as a biofuel source because they reproduce quickly, are highly efficient at photosynthesis, and can tolerate a wide range of environments, particularly salt and wastewater [10] (Figure 1-1). Microalgae naturally accumulate energy-dense

metabolites such as lipids (Figure 1-2), some species producing oil up to 60% of their biomass [11], and techniques exist for robust chloroplast engineering [12]. However, although microalgae are capable of accumulating large quantities of oil, this is often associated with stressors that compromise biomass productivity [13] [14]. One of the main challenges is to find the metabolic 'trigger' that will allow microalgae to overproduce lipids while continuing to grow quickly and robustly.

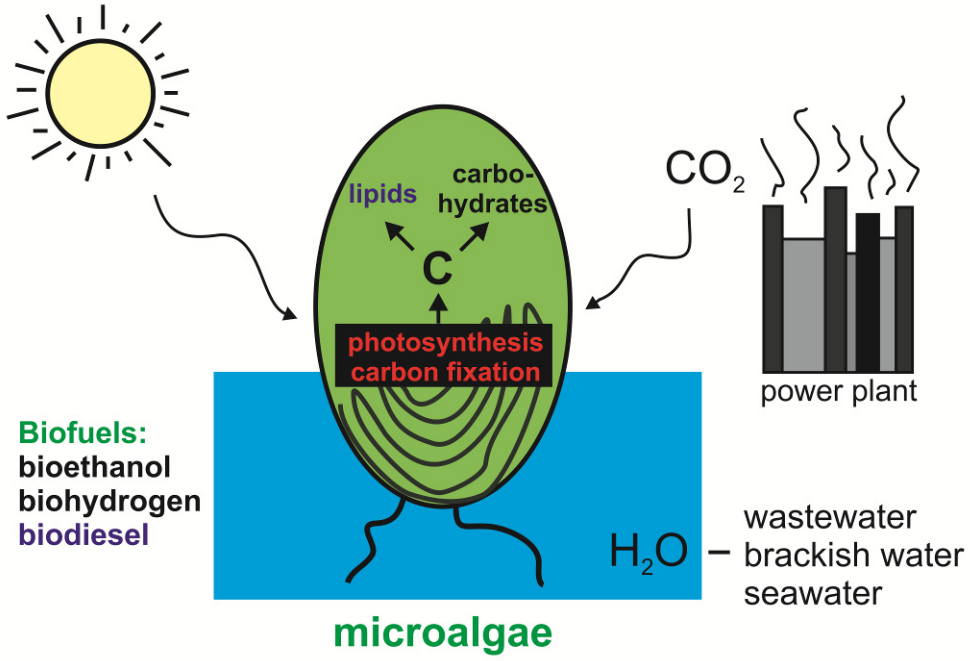


Figure 1-2 Microalgal metabolites of interest for biofuels

Microalgae store solar energy as lipids or carbohydrates. Biofuels are derived through photosynthesis and carbon fixation. Carbon dioxide (from coal-fired power plants) serves as the carbon source and nutrients in wastewater are recycled as algae biomass is cultivated in an ideal system. Biodiesel can be synthesized chemically from algal lipids and carbohydrates fermented to bioethanol or converted to hydrogen [11].

Traditional plant agriculture uses breeding and selection to arrive at crops with improved traits [15]. Mutagenesis and genetic engineering are more recent techniques

aimed at enhancing productivity and optimizing plant phenotypes. Wildtype microalgae do not have innate metabolic pathways optimized for oil biosynthesis, and the fatty acids that comprise algal lipids are not of ideal length and saturation for biodiesel properties [16]. As genetic engineering has successfully altered these properties in plants [17], engineering microalgal fatty acid biosynthesis is a promising strategy to create strains with increased oil yields and improved oil quality. Many groups have recognized the potential of microalgae as a biofuel source, and there are currently widespread efforts to develop genetic transformation technologies and screens for industrially relevant microalgal strains [18]. Various algal metabolites are being explored as potential biofuel candidates, including biohydrogen, bioethanol, biodiesel, and hydrocarbons (Table 1-1) [19], and the development of each has challenges to overcome.

Of the most promising is biodiesel, a high-energy, lightweight fuel derived from fatty acids that is a drop-in replacement of petroleum diesel, and if fashioned from oil-rich, photosynthetic organisms such as microalgae, the potential exists to robustly produce liquid biofuels via the sun using microalgae as a biocatalyst [17] [20] [21]. Indeed, genetically modified algal strains raise concern about the introduction of transgenes into the environment [22] and there is ongoing research to identify effective methods of biocontainment. Here, we address metabolic engineering of fatty acid biosynthesis in microalgae for the optimized production of biodiesel. We also outline techniques developed for algal metabolic engineering beginning with the Aquatic Species Program in 1976 [23] to current algal biofuel technology.

Table 1-1 Metabolic engineering strategies aimed at optimizing eukaryotic algae for biofuel production

Biofuel	Biodiesel	Bioethanol	Biohydrogen	Hydrocarbons
Metabolic Target	Lipids	Carbohydrates	Hydrogen	Isoprenoids Non-native products
Goal	Higher lipid yields Improve biofuel quality	Enhance starch yields in algae	Increase H_2 evolution in algae	Produce hydrocarbons in algae
Biochemical pathway	Fatty acid biosynthesis ACCase TE	Carbohydrate biosynthesis AGPase Isoamylase	Hydrogen production in algae H_2ase Isoamylase	Isoprenoid biosynthesis
Engineering solution	Upregulate ACCase (acetyl-CoA carboxylase) Inhibit Beta-oxidation Engineer plant TEs modify algal FA chain length	Engineer AGPases (ADP-glucose pyrophosphorylase) to increase starch production Inhibit starch degradation	H_2ase evolution to enhance its enzymatic efficiency to increase H_2 production in algae Decrease light-harvesting antennae size	Introduce non-native biosynthetic pathway to produce hydrocarbons Limonene synthase DMAP pathway Isoprene synthase
Results	ACCase did not effect algal lipid production Plant TEs caused a 6% and 12% increase in shorter chain fatty acids in diatoms, but no effect in green algae	Enhanced starch production in green algae Mutants were found to accumulate TAG during nitrogen deprivation	Engineering increased H_2 yields in green algae Alginate film – algae evolved up to 67% H_2 gas produced anaerobically	Hydrocarbons have been successfully produced in algae by non-native enzymes (i.e. limonene synthase) LS9 engineered heptadecene biosynthesis in algae
Reference	[145] [116] [75]	[14] [146]	[147] [148]	[86] Unpublished results

Model Organisms in Algae Biofuel Research

In 1939, a German researcher named Hans Goffron at the University of Chicago discovered that *Chlamydomonas reinhardtii* produces hydrogen gas [24]. For the next sixty years, *C. reinhardtii* would become an important model organism for biohydrogen production [25]. The photosynthetic carbon reduction cycle was mapped using radiolabeled $^{14}CO_2$ in experiments between 1946 and 1953 by Professor Calvin

and colleagues at Berkeley using green microalga *Chlorella pyrenoidosa* [26]. At that time, algae were emerging as an important model system to study mechanisms of photosynthesis, but research was mainly conducted with intact *Chlorella*, *Scenedesmus*, *Anacystis* and *Euglena* cells [27]. Researchers were beginning to realize that microalgae have unique traits that can be exploited, such as in the development of an *in vivo* assay using red alga *Porphyridium cruentum* to demonstrate photoreduction and photooxidation of cytochrome *f* by PSII and PSI, respectively, in 1961 [28]. Until this time, most experiments on microalgae were physical measurements, such as gas exchange and fluorescence, or analyzing the production of $^{14}CO_2$ incorporation [27].

Alternatively, a new technique was emerging from the laboratories of R. P. Levine at Harvard University in the 1960s which consisted of altering the genetic constitution of photosynthetic organisms and study the phenotypic output, generating so-called 'photosynthetic mutants' [29]. While Levine created mutants of *C. reinhardtii*, another prominent algae researcher, N. Bishop, pioneered mutant work on *Scenedesmus obliquus* [30] [31]. Analysis of photosynthetic algal mutants revealed new information about the photosynthetic apparatus by the early 1960s. Although *S. obliquus* was a more established model, the potential of *C. reinhardtii* seemed greater because of the ability to control sexual reproduction and conduct detailed genetic analyses [32]. Continuing into the 1970s, tools to study the structure and function of the biochemistry of the algal chloroplast were evolved [33]. The Levine laboratory developed screens for genetic mutants of *C. reinhardtii* [34] and a new liquid growth media called TAP (Tris Acetate Phosphate) for culturing photosynthetic *C. reinhardtii*

mutants, which is still widely used in current algal biofuel research (now sold by Life Technologies) [35].

The Levine laboratories also began to explore the regulation of expression in the chloroplast, and for the first time, mapped chloroplast functions to nuclear genes [36]. Along with a heightened interest in algae as a source of renewable fuels, the field of plant genetic engineering was emerging in the early 1980s, although no method existed to transform plants. *Agrobacterium* was used for plant transformation in 1986 [37], but a major breakthrough came through when Professor Sanford at Cornell University experimented with physical means of introducing DNA into cells, and after a series of unpractical ideas, landed on using gold- or tungsten-coated microprojectiles as carriers of DNA and accelerated the particles at high velocities to penetrate the cell wall. In 1988, the first transformation of algae was accomplished in *Chlamydomonas* using biolistics [38], which directly followed the transformation of maize [39]. With tools to transform the genome of model alga *C. reinhardtii*, the door was opened to algal genetic engineering.

For the heterologous expression of biosynthetic enzymes, *C. reinhardtii* is an ideal host [40]. A eukaryote compartmentalized by three separate genomes, each now fully sequenced, *C. reinhardtii* has a codon bias in the nucleus (62% G-C rich) different from that of the chloroplast (34% G-C rich), and genes can be selectively transformed into each of these subcellular regions [12]. Additionally, proteins are naturally trafficked between the organelles in *C. reinhardtii* by means of well known transit peptides that may be exploited for guiding the sub-cellular localization of the engineered metabolic machinery [41]. The purification of recombinant proteins and

metabolites is also greatly simplified relative to terrestrial plants, as the cellular population is uniform in size and type. Hence, there is no gradient of protein or metabolite distribution, which simplifies purification and reduces the amount of biomass that goes toward non-productive ends [33] [42]. Another valuable benefit of *C. reinhardtii* is that it has an exceptionally high capacity for economical production (1.2 million liters/ acre in 12 weeks). *C. reinhardtii* is thus emerging as a useful and important model organism in algae biofuels research.

Genetic Engineering of Eukaryotic Algae for Biodiesel

Rudolph Diesel showcased his newly fashioned diesel engine at the World Fair in Paris in 1912, running the engine from peanut oil.

> *The diesel engine can be fed with vegetable oils and would help considerably in the development of agriculture of the countries which use it. The use of vegetable oils for engine fuels may seem insignificant today. But such oils may become in course of time as important as petroleum and the coal tar products of the present time.*
>
> *-Rudolph Diesel, 1912*

Current biodiesel is a derivative of vegetable oils, defined as fatty acid methyl esters [43]. As oleaginous microalgae are an order of magnitude more productive than even the highest yielding oil crops (Figure 1-1), microalgae are currently under investigation as a biodiesel feedstock so that renewable biodiesel can compete economically with petroleum.

Using microalgae as cell factories to produce biofuels is a concept that dates back several decades. Algae were first considered by Meier in 1955 [44] and the Oswald and Golueke laboratories at UC Berkeley in 1957 to produce methane via

anaerobic digestion [45] and to recycle nutrients in wastewater and CO_2 [46]. In 1973, OPEC (Organization of Arab Petroleum Exporting Companies) pronounced an oil embargo on Western nations. This energy crisis sparked a renewed interest in alternative fuels [47], which included algal biofuels [48], but the focus remained on methane and hydrogen. Major efforts to conserve oil were undertaken in the U.S., including the creation of the Department of Energy (DOE) in 1977, but dependence on foreign oil was exposed.

In 1978, with the foresight of energy independence, President Carter established the Aquatic Species Program (ASP), a widespread research effort funded by the U.S. DOE Office of Renewable Fuels aimed at developing renewable transportation fuels from algae. ASP took the concept of algal biofuels to the next level, shifting the focus from gaseous fuels to oils, lipids, and diesel replacements [23]. During the years of ASP (1978 – 1996), 3000 strains of unique algal species were identified, those capable of producing high amounts of oil and growing in extreme environments, and they were stored in a single culture collection at the University of Hawaii. Algal cultivation in outdoor ponds using CO_2 from coal-fired power plants was first demonstrated [23], which illuminated challenges such as crop protection. Lipids within algal biomass were chemically transformed into biodiesel using transesterification and extraction methods were explored. Importantly, the ASP noted the most promising biofuel feedstock organisms as diatoms and green algae, which dominated the culture collection at UH because they are robust, abundant species that grow fast and contain large amounts of oil [49]. The ASP showed the potential of algae as a biodiesel feedstock, and key enzymes in algal lipid metabolism

were characterized [50], but the cost of algal biodiesel production remained a major obstacle for economic viability at a large scale [23]. In the 1980s, petroleum prices plummeted and President Reagan cut funding to many of the ASP programs. Then in 1996, President Clinton terminated the ASP to shift the greatly diminished budget to bioethanol production. The work conducted during the ASP was summarized in a report by Sheehan et al. in 1998, which left the community with the overall conclusion that 'algae biofuel' has many technical barriers to overcome for economical production, and at best, with aggressive biological breeding, algae could yield biodiesel at two times the cost of petroleum with current technology [23]. However, scientists involved in the ASP remained optimistic about the future of algae biofuels:

> *This report should be seen not as an ending, but as a beginning. When the time is right, we fully expect to see renewed interest in algae as a source of fuels and other chemicals. The highlights presented here should serve as a foundation for these future efforts.*
>
> *-John Sheehan, 1998*

We are now in the era of the renaissance of algae biofuel research. With fluctuating oil prices creating an unstable world economy, the desire for renewable energy has resurged, and the potential of algae as an energy feedstock cannot be overlooked. Although the shutdown of the ASP led to major downsizing in algae biofuels research for roughly ten years, during the nineties there were still laboratories focused on engineering algae for high-value co-products, such as therapeutic proteins [51], or to study facets of algal metabolism [52] [53] and evolution [54]. These groups sought to develop efficient algal transformation technologies [55] [56], characterize genes involved in algal lipid metabolism [57], and study the molecular genetics of

sexual reproduction in algae [58]. Many key advances in algal biotechnology and metabolic engineering occurred in model eukaryotic green alga *C. reinhardtii* [32].

Microalgae are unicellular, photosynthetic eukaryotes that contain separate genomes in the nucleus, mitochondria, and chloroplast, due in part to evolutionary endosymbioses [59] [60]. The algal chloroplast genome is comprised mainly of genes that encode proteins involved in photosynthesis [61]. The energy derived from photosynthetic electron transport is used to drive other biosynthetic processes, such as fatty acid biosynthesis, which also occurs in the chloroplast [57]. However, FAS enzymes are encoded in the nuclear genome of green algae and targeted to the chloroplast by means of transit peptides [62]. *C. reinhardtii* contains a large chloroplast that occupies 40% of its volume, a hydroxyproline-rich cell wall, mitochondria, a nucleus, a large pyrenoid, an eyespot, and swims with two flagella. *C. reinhardtii* contains ~20% oil and is a haploid species [27]. Thus, changes to recessive alleles can be observed immediately, an attractive feature for breeding 'crops'.

To accelerate the domestication process for microalgae, in combination with selection and breeding [63], genetic approaches are essential in studying lipid metabolism and understanding carbon partitioning during reduced nutrient conditions. Genetic techniques can be used to engineer algal lipid metabolism and select for robust lipid producers that retain high biomass productivity, which will eventually facilitate large-scale production of algae [63]. Among these approaches are genomics, genetic engineering, directed evolution and screening, and sexual crossing. Initial genetic engineering attempts have unveiled regulatory networks in microalgae that are

not well understood [64]. To evolve efficient genetic manipulation strategies for a variety of algal species, the major goals are to develop selectable markers, universal transformation vectors, tools to control gene expression, and gene-tagging protocols to monitor protein expression and localization [65].

Algal metabolic engineering has evolved over the past several decades, and certain discoveries expedited this process. The contributions of R.P. Levine and others in the 1960s and 1970s provided the foundation of tools to manipulate algal genetics, using *C. reinhardtii* to discover many aspects of photosynthesis [32]. A significant stride forward in algal metabolic engineering came in 1988 when stable chloroplast transformation using biolistics was successfully applied to *C. reinhardtii* [38], and soon thereafter to model diatom *Phaeodactylum tricornutum* [50]. The following year, the first high efficiency nuclear transformation protocol was described for *C. reinhardtii*, in which a cell wall-deficient *C. reinhardtii* strain, agitated using glass beads in the presence of DNA and polyethylene glycol, exhibits nuclear transformation rates of ~10^3 transformants per micrograms of plasmid DNA [66]. In the following decade, groups sought to transform *C. reinhardtii* by electroporation, silicon carbide whiskers, glass beads, and biolistics [67].

With tools to selectively transform its chloroplastic and nuclear genomes, *C. reinhardtii* was used as a model organism to study a wide range of metabolic processes [40]. For example, mechanisms behind the translational apparatus of *C. reinhardtii* chloroplast were becoming clear [68]. In 2002, GFP was used as a reporter for chloroplast gene expression, but high auto-fluorescence required high levels of GFP accumulation for visualization *in vivo* [69]. Notably, this study

revealed that codon-optimizing genes for the *C. reinhardtii* chloroplast results in gene expression levels that are 80-fold higher than unoptimized genes [69]. A more-sensitive and versatile luciferase reporter was developed to visualize gene-expression in the chloroplast in living algal cells [70]. In this work, a codon-optimized luciferase gene resulted in expression levels that allowed real-time imaging of chloroplast gene expression [70]. In groundbreaking work, *C. reinhardtii* was engineered to express a fully active human antibody in the chloroplast [71], which demonstrated to the algae biofuel community that algae can be optimized to express high value coproducts, critical for economical algae biofuel production. The discovery of non-coding RNAs in *C. reinhardtii* in 2004 allowed for the downregulation of genes using RNAi technology [72]. Recently, as an improved strategy, a high-throughput artificial miRNA (amiRNA) technique was developed for targeted, highly specific and stable gene knockdowns in *C. reinhardtii* [73], emerging as a useful strategy to elucidate metabolic pathways in algae.

In 2007, the genome of *C. reinhardtii* was published in *Science* [61]. This provided the first genetic blueprint of a model green alga, which would facilitate algal metabolic engineering. Although this was a major achievement towards identifying genes involved in algal metabolism, many of the pathways of interest for biofuels had yet to be characterized, in particular the enzymes involved in fatty acid biosynthesis. R. P. Levine used *C. reinhardtii* mutants to study algal fatty acid biosynthesis in 1971, and arrived at the important observation that FAS is ACP-dependent [74]. The genome sequence of *C. reinhardtii* has allowed the cloning and characterization of both of its ACPs [75], an important step towards manipulating algal FAS.

There are still many unanswered questions in algal metabolism, and there are processes in algal genetic engineering that must be optimized. Many groups have been working to enhance algal chloroplast transformation efficiency, improve upon algal nuclear transformation and expression, and develop strategies to control gene expression. Notably, riboswitches are being explored as a means of controlling algal gene expression and regulation [76]. A chemically-inducible regulation system was developed for *C. reinhardtii* that allowed for temporally-controlled regulation of foreign genes [77]. The light-inducible promoter *psbA* has been widely used to achieve high levels of protein expression in the *C. reinhardtii* chloroplast [78]. The development of inducible systems and ways to control gene expression are important in engineering algal fatty acid biosynthesis, and strategies must be applicable to a wide range of algal species.

Desire for oil-independence, national security, and mitigation of CO_2-related climate change lead to a renewed interest in sustainable algal biofuels by the U.S. DOE, which created the National Algal Biofuel Roadmap in 2008 [79]. As academic research gravitated towards algae biofuels, Venture Capitalists began to invest in algae biofuel companies. Sapphire Energy (La Jolla, CA), Solazyme (San Francisco, CA), Aurora Biofuels (Hayward, CA), and Synthetic Genomics (La Jolla, CA) are examples of algae biofuel companies funded by the U.S. DOE or the private sector in recent years, and companies with similar goals are emerging globally. Petroleum companies have also began to invest in algae biofuel companies. As examples, Chevron has partnered with the NREL and Solazyme, Exxon Mobil with Synthetic Genomics, Shell with HR Biopetroleum and BP with Martek Biosciences. The U.S. DOE has awarded

funds to Solazyme and Sapphire Energy to provide algal biofuels to the U.S. NAVY and U.S. Department of Defense.

The ASP taught us that widespread efforts are necessary to address all of the technical barriers involved in developing a sustainable algae biofuels industry. There have been important recent efforts undertaken to facilitate this idea. Consortia have been established between industry, government, and academia to foster dissemination of ideas, research, and collaboration. One example is the San Diego Center for Algal Biotechnology (SD-CAB), a collaborative algae biofuel research community in San Diego, and its nation-wide research unit CAB-COMM (Consortium for Algae Biofuels Commercialization), funded by the U.S. DOE. In another important effort, biotechnology companies have made a commitment to providing laboratories with tools to engineer algae and to study algal metabolism. For example, Life Technologies has been actively communicating with scientists involved in algal biofuel research, collecting ideas and suggestions to generate 'algae kits' for growth and transformation so that new information about algal biochemistry and metabolism can be derived quickly.

DNA sequencing is now more efficient and accessible, and genomes of industrially relevant algae are rapidly becoming available [56]. Biologists and biochemists are working to develop tools to select for algal strains with desirable characteristics [63] and chemists are optimizing analytical tools to evaluate biofuel-producing algal strains, such as a recently developed high-throughput GC/MS approach to analyze fatty acid-derived algal biofuels [80]. This round of algae biofuel

research necessitates a concerted effort, merging a wide-array of disciplines in order to make algae biofuels a reality.

Attractive Metabolic Targets for Algae Biofuel Engineering

Essentially, all of the algal metabolites of interest for biofuels are derived from the sun and CO_2 through photosynthesis and carbon-fixation [81] (Figure 1-2). Water and nutrients are required for the survival of the organism, but a major advantage of algae over plants is their ability to grow on land not suitable for conventional agriculture [11]. While growing at a fast rate in these habitats, algae efficiently store energy from photons in energy-dense metabolites such as lipids and carbohydrates (starch), two major targets of algae biofuel research (Figure 1-2). The fatty acids that comprise algal lipids can be converted into biodiesel using transesterification to create a 'drop-in' petroleum replacement [16]. Alternately, starch can be fermented into ethanol, a less energy-dense product, or metabolized to a variety of biofuels, such as alcohols, lipids, hydrogen and methane [82]. Algae naturally produce hydrogen in an anaerobic environment replete with sulfur, which is coupled to photoassimilation and the catalytic oxidation of water [25], and many research efforts are aimed at optimizing this process. Engineering algae for hydrogen production is focused on upregulating or evolving hydrogenases to enhance catalytic efficiency, decreasing the light-harvesting antennae size, and abolishing state transitions [83] [84] [19]. These initial studies have added to our understanding of hydrogen metabolism and enzyme evolution. However, even advocates of a 'hydrogen-economy' are aware that hydrogen storage and transport are major obstacles [85], as well as the inevitability of

building a multi-trillion dollar hydrogen infrastructure *de novo*. Biodiesel and hydrocarbons are compatible with current petroleum infrastructure and are ideal for energy transport and storage.

Along with the aforementioned goals to exploit native metabolic pathways to produce biofuels in algae (Table 1-1), there are aims in synthetic biology to introduce entire non-native biosynthetic pathways to coax microalgae into producing a variety of products. Some companies and research groups have engineered green algae with genes encoding isoprenoid biosynthesis to create 'green crude,' a term that describes the direct synthesis of petrol fuels from algae. For example, we have engineered *C. reinhardtii* to express limonene synthase in the chloroplast, which resulted in high levels of limonene production (unpublished results), and scientists at LS9 have engineered algae with biosynthetic enzymes to produce heptadecene, a C17 hydrocarbon product [86]. Although these are promising strategies towards algal biofuels and coproducts, we focus on the optimization of an innate process – one in which algae are quite efficient – rather than introducing a non-native biosynthetic pathway [87]. Here we review engineering algal fatty acid biosynthesis to increase oil content and enhance fuel quality.

Engineering algal fatty acid biosynthesis

Biodiesel is a high-energy, lightweight fuel derived from fatty acids via transesterification [43]. Engineering fatty acid biosynthesis in oil-rich biodiesel feedstock organisms, such as microalgae, is therefore of considerable interest. Remarkable successes have been achieved in engineering fatty acid biosynthesis in

plants to enhance productivity and create designer oils [17]. Reasoning that relatively high homology exists between plants and algae, the initial hypothesis amongst the algae biofuel community followed that these same plant enzymes could be engineered into algae with similar results. However, experiments in diatoms and green algae reveal that this strategy only marginally works [88] [75]. Here, we describe pioneering efforts to engineer fatty acid biosynthesis in eukaryotic algae and recent discoveries of the algal fatty acid biosynthetic pathway that will facilitate future engineering.

A thorough understanding of algal fatty acid synthase (FAS) is essential for successful engineering. The past five years has seen an explosion in algae biofuel research, and accordingly, there are currently ~35 algal genomes sequenced (and rising), transformation technologies exist for at least twelve different algal species, and 'omics' research has acquired a substantial amount of data for algal metabolism [89] [90]. Fatty acid and lipid biosynthetic genes are now annotated in several algal species, BLAST analysis continues to find homologues in newly sequenced algae, and several FAS genes have been cloned from algae and characterized [91]. Biochemical characterization of algal FAS enzymes has been carried out on the acetyl CoA carboxylase (ACCase), which forms malonyl CoA from acetyl CoA [50], thioesterase (TE) domain, which catalyzes the hydrolysis of fatty acids from the ACP, and the ACP, the metabolic template for fatty acid synthesis [75] (Figure 1-3). Heterologous FAS enzymes from plants were recently introduced into diatoms (*P. tricornutum*) and green algae (*C. reinhardtii*) with the goal of increasing oil content and modifying the identity of the algal fatty acids, and the relatively low success of these engineering

attempts highlights the requirement for detailed understanding of the algal FAS biosynthetic machinery. Biochemical characterization of algal FAS enzymes, including catalytic efficiency, timing, regulation, and protein-protein interactions, is essential for successful FAS engineering in microalgae.

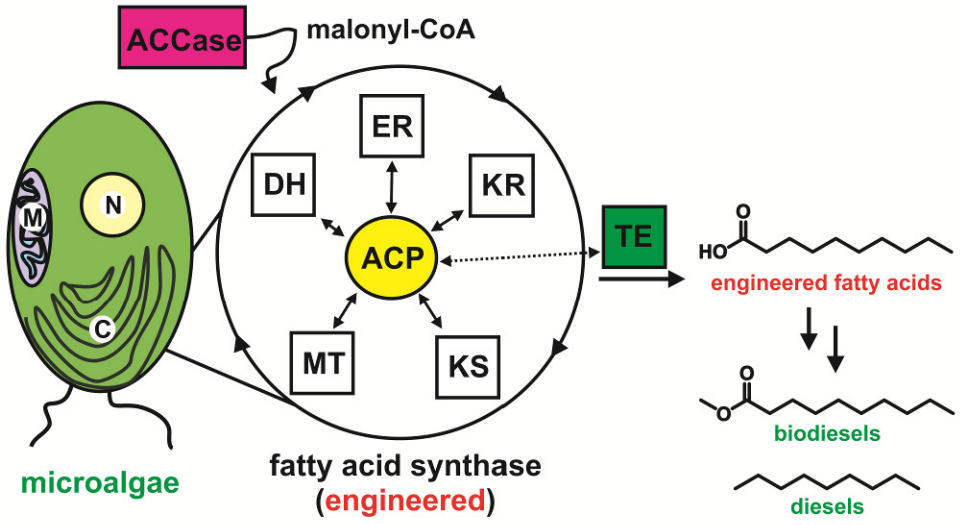


Figure 1-3 Engineering algal fatty acid biosynthesis

In microalgae, the first committed step in the synthesis of a fatty acid occurs with the carboxylation of acetyl-CoA to form malonyl-CoA via an acetyl-CoA-carboxylase enzyme (ACCase). This is the rate-limiting reaction in the fatty acid biosynthetic process, and thus efforts to increase oil content are aimed at overexpressing ACCase (magenta). Once malonate is attached to the acyl carrier protein (ACP), an iterative cycle of chain elongation commences by action of ketosynthase (KS), ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) domains, each turn resulting in a net addition of two carbons to the growing fatty acid. Once a fatty acid has reached a mature length on ACP (yellow), a thioesterase (TE) domain catalyzes hydrolysis of the fatty acid and release from ACP, which diffuses through the plastid membrane, esterifies to CoA and incorporates into cellular lipids. In plants, TEs demonstrate high substrate specificity and can alter the fatty acid profile when engineered into transgenic plants. Thus, in efforts to improve biodiesel quality by engineering fatty acid chain length, TEs (green) are a major focus. Fatty acid can be converted into their methyl ester derivatives to synthesize biodiesel or hydrotreated to produce conventional diesel products [57].

Overview of fatty acid biosynthesis and TAG accumulation in microalgae

Fatty acid biosynthesis has been well characterized in bacteria and plants [93] [92], and the algal pathway has been theoretically deduced from homology. In sequenced algae, such as *C. reinhardtii*, all of the enzymes of FAS have been accounted for [61]. In the algal chloroplast, fatty acids are biosynthesized by a type II FAS, a group of dissociated enzymes that catalytically elongate and reduce a growing fatty acid as it iteratively grows by 2 carbon units [57]. This process begins as ACCase carries out the carboxylation of acetyl-CoA to form malonyl-CoA, considered the rate-limiting step in fatty acid biosynthesis in plants [93] [94]. Therefore, ACCase is an enzymatic target of manipulation to increase oil content (Figure 1-3).

The central player in FAS in the acyl carrier protein (ACP), which must first get post-translationally modified by a phosphopantetheinyl transferase (PPTase), which converts *apo*-ACP to its *holo* form by transferring a 4'-phosphopantetheine flexible prosthetic group from Coenzyme A to a conserved serine on ACP. This 'arm' on *holo*-ACP tethers the growing fatty acid via thioester linkage throughout fatty acid biosynthesis. Protein-protein interactions with each of the FAS enzymes cause a conformational change in ACP that elicits the fatty acid from its hydrophobic core for catalytic elongation and reduction in a 'switchblade mechanism' [95]. KASIII, the initiating ketosynthase domain, condenses acetyl-CoA with malonyl-ACP to form a β-ketone extended 2 carbon unit on acyl-ACP, which then undergoes reduction to a fully-saturated methylene unit by ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) enzymes (Figure 1-3). This cycle repeats seven more times with KASI (elongates C4-C16) and KASII (forms C18 from C16), and when a mature fatty acid

reaches its destined length, it is either retained in the prokaryotic pathway and transferred to glycerol-3-phosphate by an acyl transferase (AT) for incorporation into membrane lipids, or hydrolyzed by action of a thioesterase domain (TE), which releases the free fatty acid from the FAS for export out of the plastid and incorporation into cellular lipids. Once hydrolyzed from the ACP by the TE, fatty acids diffuse into the plastid envelope, are esterified to Coenzyme A by a CoA ligase enzyme, and the fatty acyl-CoA departs from the chloroplast for incorporation into cellular lipids, such as membrane phospholipids and storage triglycerides (TAGs) (Figure 1-4). In plants, the TE determines the carbon flux between the prokaryotic and eukaryotic lipid termination pathways, serving as a metabolic gatekeeper in lipid biosynthesis [96]. The TE functionally determines the identity of the end fatty acid product [97], and as such, this enzyme is of interest to modify algal fatty acid chain length for improved biodiesel quality in feedstock algal strains (Figure 1-3).

Fatty acids are biosynthesized in the chloroplast, and must be supplied to the plastid membrane lipids (prokaryotic) and cellular lipids (eukaryotic), akin to plants [98]. The primary products of algal FAS are C16:0 and C18:0, which can be hydrolyzed by a TE or desaturated by action of a soluble delta-9-acyl-ACP desaturase. In plants, chloroplastic glycerol-3-phosphate acyl transferase (GPAT) and lysophosphatidic acid acyl transferase (LPAT) attach fatty acids directly to glycerol-3-phosphate, forming lysophosphatidic acid and phosphatidic acid. Phosphatidic acid, at the center of many signaling cascades in higher organisms [99], and its dephosphorylated product diacylglycerol (DAG), serve as precursors for membrane lipids. GPAT and LPAT isoforms are present in the ER. Putative genes for GPAT

and LPAT have also found in the *C. reinhardtii* genome [61]. The gene encoding plastid phosphatidylglycerolphosphate phosphatase (phosphatidic acid phosphatase), necessary for the biosynthesis of the diacylglycerol precursors, has remained elusive in algae, but recently candidates have been identified [100]. Disruption of this gene in cyanobacteria has been correlated with deficient photosynthesis [101]. Detailed PSI-Blast analysis (unpublished results) illuminates homologues in green algae, including *C. reinhardtii*. Plastid lipids of *C. reinhardtii* consist of mono- and digalactosyldiacylglycerol (MGDG and DGDG), sulfoquinovosyldiacylglycerol (SQDG) and phosphatidylglycerol (PG), from which the galactoglycerol lipids are the major membrane lipids [102] [103] [104] (Figure 1-4). Most of the genes involved in the biosynthesis of prokaryotic membrane lipids have been found in the genome of *C. reinhardtii* [105].

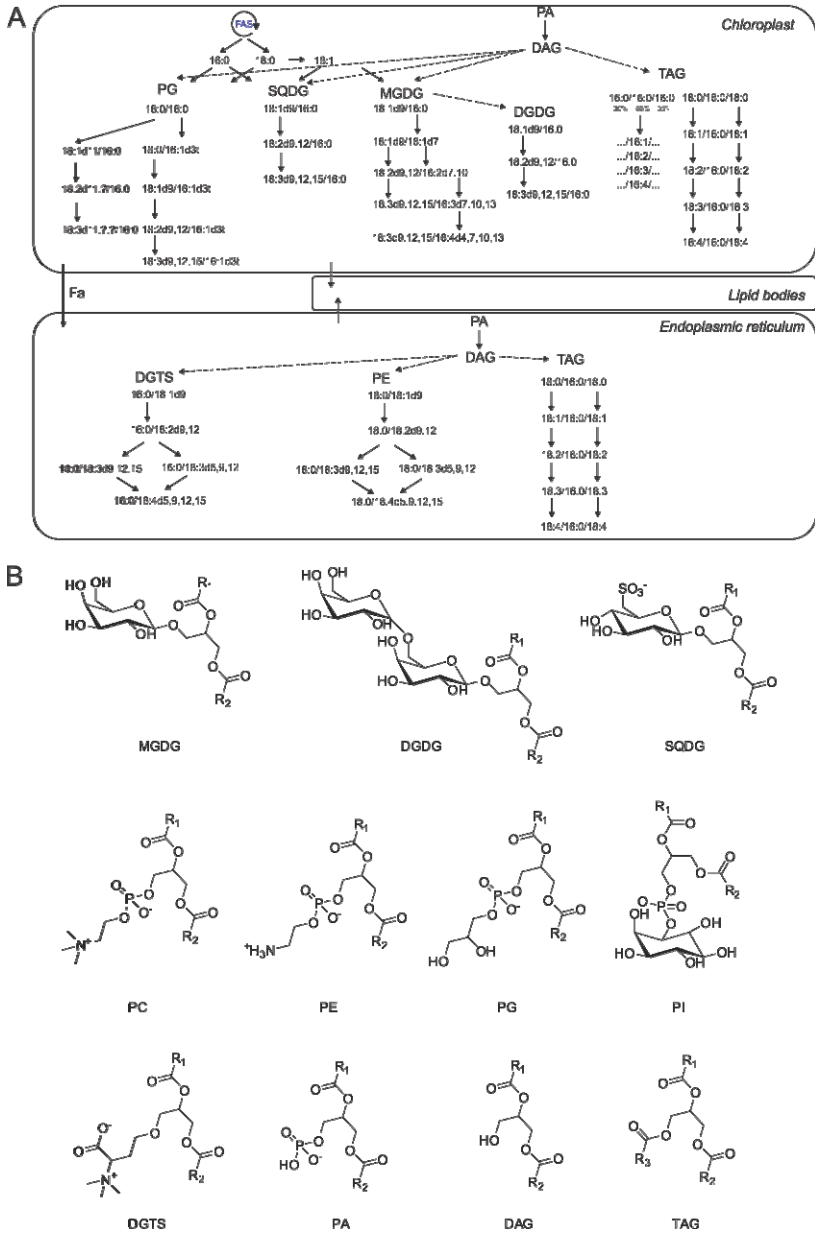


Figure 1-4 Algal lipid biosynthesis

(a) Polar and neutral (membrane) lipid biosynthesis in the chloroplast and endoplasmic reticulum of *C. reinhardtii* [49] [107]. FAS, fatty acid biosynthesis. Fa, fatty acids. PG, phosphatidylglycerol. SQDG, sulfoquinovosyldiacylglycerol. MGDG, monogalactosyldiacylglycerol. DGDG, digalactosyldiacylglycerol. TAG, triacylglycerol. DGTS, diacylglyceryl(N,N,N-trimethyl)homoserine. PE, phosphatidylethanolamine. PI, phosphatidylinositol. PC, phosphatidylcholine. PA, phosphatidic acid. DAG, diacylglycerol. Dashed lines present lipid headgroup biosynthesis, whereas solid lines represent fatty acid modifications on the fatty acyl groups attached to the lipids. (b) Structures of the major lipids; R1 and R2 are fatty acid acyl chains.

In the eukaryotic pathway, fatty acids are hydrolyzed from the ACP by the TE and exported into the cytosol and used for the assembly of cellular membrane lipids or neutral storage lipids (TAGs) [49]. Eukaryotic microalgae are quite diverse in lipid species [106] [107] (Figure 1-4). Algae generally contain more unsaturated fatty acids than cyanobacteria, yet the major fatty acid species present in both organisms is palmitic acid (16:0). The distribution of fatty acids is distinct for each species and controlled tightly by the organism, most likely relating to the function of fatty acids in acyl membrane lipids [108]. Glycosylglycerides, monogalactosyldiacylglycerol (MGDG), digalactosyldiacylglycerol (DGDG), sulphoquinovosyldiacylglycerol (SQDG) are the major lipid species present in eukaryotic algae [49]. Most algae contain ether lipids, usually diacylglyceryl-(*N*,*N*,*N*-trimethyl) homoserine (DGTS). Interestingly, seed plants and microalgae utilize similar pathways to biosynthesize cellular lipids, but *C. reinhardtii* uses DGTS, a structural analogue of phosphatidylcholine (PC) (an ether bond is in place of the phosphodiester bond in PC), thus thought to serve the same function [105] (Figure 1-4). Also, lipid precursors do not traffic from the ER to the plastid in *C. reinhardtii*, whereas in seed plants, thylakoid lipid biosynthesis requires both plastid and ER pathways [98].

Microalgae are attractive as a biodiesel feedstock due to their ability to accumulate neutral lipids, such as triglycerides (TAGs), under environmental stress conditions, which they store in dedicated organelles called lipid bodies [14]. The biochemistry of these structures is currently being deduced through proteomics, transcriptomics, and genomics, and these efforts have successfully identified a lipid-

droplet protein in *C. reinhardtii* [109] [110], which challenged the notion that lipid bodies were solely comprised of TAGs and fatty acids, their major constituents.

The biosynthesis of neutral lipids is not well studied in microalgae, but substantial research efforts are aimed at elucidating this pathway so that oil production can be manipulated. It has been widely assumed that TAGs are synthesized in algae via the ER-localized Kennedy Pathway similar to seed plants [98]. In this route, fatty acids are added stepwise to glycerol-3-phosphate, which is converted to DAG by phosphatidic acid phosphatase (PAP). Diacylglycerol acyltransferase (DGAT) acylates DAG to append the third acyl moiety, forming TAG. This is supported by the observation that putative DGAT genes have been identified in *C. reinhardtii* [100]. Under nitrogen or phosphate replete conditions or high salt conditions, *C. reinhardtii* accumulates TAGs in lipid bodies in the cytosol (~ 2.5% to 40%) [11] [14].

C. reinhardtii contains alpha-cytosolic lipid bodies, which under nitrogen deprivation grow into beta-cytosolic lipid bodies [111]. Goodson et al. observe a close association between one face of the beta-cytosolic lipid bodies and the ER and the opposite face associates with the outer membrane of the chloroplast envelope [111]. These results have lead to the formation of a new model of TAG assembly in which fatty acids are biosynthesized in the chloroplast, then diffuse into the lipid body (shares a membrane with both the chloroplast and ER) and eventually reach acyl transferases (DAGAT) in the membrane of the ER, which catalyze the esterification of the acyl moiety of the fatty acid to DAG, forming TAG in the growing lipid body.

An alternative pathway was recently postulated in which TAGs are also biosynthesized in the chloroplast [112]. However, chloroplastic lipid bodies were only

observed in the *sta6* strain, a starchless *C. reinhardtii* mutant [111]. One possible explanation is that carbon flux normally allocated for starch biosynthesis is redirected to a constitutive TAG pathway under N-replete conditions. Another is that *C. reinhardtii* contains a silent TAG biosynthetic cluster in its genome that is induced upon nitrogen limitation when starch biosynthesis is blocked. It appears initially that *C. reinhardtii*, and perhaps other eukaryotic algae, contain two TAG biosynthetic machineries present in both the chloroplast and ER. For successful manipulation of algal oil content, these pathways must be further elucidated in detail.

Plant engineering sets precedent for engineering of algal fatty acid biosynthesis

Plants have been engineered with heterologous plant FAS enzymes to successfully modify oil and fatty acid content [17]. In one striking example, engineering a 12:0-specific plant TE from *Umbellularia californica* into transgenic *Arabidopsis thaliana* achieved a large increase in laurate (C12:0) in its storage oil (TAG) [113]. With similar engineering goals aimed at optimizing microalgae as a biodiesel feedstock, namely increasing the amount of oil and increasing the percentage of short-chain fatty acids in the oil, plant TEs initially attracted a lot of attention. However, recent reports of engineering plant TEs into green algae [75] and diatoms [88] resulted in quite lower yields. Structure-based sequence analysis of plant TEs (FatA and FatB) and comparison to algal TEs (Fat1) has unveiled distinct features of algal enzymes [75]. Furthermore, recent work in green algal FAS suggests protein-protein interactions between engineered FAS domains and endogenous algal biosynthetic machinery are required for successful biocatalysis.

Increasing oil content

Eukaryotic microalgae are attractive as a biodiesel feedstock due to their ability to biosynthesize large amounts of oil [11]. However, oil-hyperaccumulation is triggered by environmental stress, such as nitrogen depletion, which hinders growth and reproduction of the algae [14]. A major goal in algae biofuels research since the ASP has been to find the elusive 'lipid trigger'. Alternately, there are considerable metabolic engineering efforts to enhance oil production in algae while maintaining robustness and high growth rates.

In plants, it is postulated that the ACCase is the rate-limiting enzymatic step in the biosynthesis of fatty acids [108]. In experiments aimed at studying algal lipid metabolism under nutrient deprived conditions in unicellular algae *Phaeodactylum tricornutum*, ACCase activity was enhanced up to fourfold in silicon-starved strains [114]. Kinetic characterization of ACCases purified from *P. tricornutum* [51] [117] and *Isochrysis galbana* [115] demonstrated the similarity between algal ACCases and those present in plants. Thus, transformation was accomplished using biolistics to introduce ACCase genes into diatoms *P. tricornutum* and *Navicula saprophila*. Surprisingly, ACCase overexpression did not result in increased oil production in diatoms [50].

Recently, in efforts to engineer fatty acid chain length in diatoms, plant thioesterases specific for 12:0-ACP (*U. californica*) and 14:0-ACP (*Cinnamomum camphora*) were engineered into *P. tricornutum*, which only slightly altered the fatty acid profile, but overall lipid content increased up to twofold in strains expressing plant TEs and native PtTE [116]. In contrast, overexpression of the endogenous TE in

C. reinhardtii did not result in an increase in overall fatty acid content, nor did expression of plant TEs from *U. californica* and *Cuphea hookeriana* [75].

These initial attempts to introduce FAS enzymes into microalgae have brought forth the realization that algal fatty acid biosynthesis is not as similar to plants as originally suspected – even in green microalgae such as *C. reinhardtii*. Metabolic engineering strategies that have been successful in plants and prokaryotes may not be amenable to microalgae. To effectively engineer fatty acid biosynthesis in microalgae, it is critical that we conduct detailed biochemical investigations of algal FAS and elucidate its regulatory networks so that we can devise an optimized and controllable system to facilitate economical large-scale algae biofuel production.

Modification of lipid composition/ fatty acid identity

In addition to increasing oil content and productivity of algal strains, it is also desirable to improve the quality of biodiesel produced. Biodiesel is composed of methyl esters of fatty acids, ideally those relatively short in length (8-14 carbons) and either saturated or monounsaturated [43]. However, most microalgae biosynthesize fatty acids composed of 16-22 carbons with varying degrees of unsaturation [107]. Because thioesterases terminate chain elongation in fatty acid biosynthesis, they functionally determine the identity of the fatty acid end product [97]. Therefore, engineering TEs that select for C8:0-C14:0 fatty acids into microalgae is a promising strategy to alter fatty acid composition (Figure 1-3).

Plant TEs are highly specific and are classified according to their substrate preference. FatA TEs hydrolyze 18:1-ACP and FatB TEs siphon a range of saturated

acyl-ACPs (C8-C14) [117]. Certain plant species accumulate medium chain fatty acids in their seed storage oil, and accordingly, there exists FatB TEs with specificities corresponding to the medium chain phenotype in high abundance in their seeds [118]. For example, the Mexican cigar shrub (*Cuphea hookeriana*) hordes up to 75% caprylate (C8:0) and caprate (C10:0) in its seed storage oil [119], and 8:0- and 10:0-specific FatB TEs are expressed at high levels in its seeds [120]. The proof-of-principle of introducing a plant FatB TE into a heterologous host to alter its fatty acid phenotype was demonstrated in 1992 by Voelker and coworkers at Calgene by engineering at 12:0-specific FatB TE from the California bay laurel plant (*U. californica*) into *Arabidopsis thaliana*, which resulted in a large increase in 12:0 fatty acids incorporated into its storage oil (TAG) [113]. Following this initial success, a variety of plant TEs were engineered into oleaginous plants, effectively altering the oil composition [17]

In contrast to plants, which contain highly specific TEs (FatA, FatB1, FatB2), distinct TEs have been identified in microalgae (*C. reinhardtii* and *P. tricornutum*) that are more promiscuous in the substrates they act upon [116] [75]. Phylogenetic analysis of TEs in all sequenced microalgal genomes illuminates a new class of TEs termed ‘Fat1’ [75]. Sequencing of the *C. reinhardtii* genome revealed a fatty acid TE in its nuclear genome [61] annotated as a FatA TE based on homology to plant FatA TEs. However, it is the only fatty acid TE detected in the *C. reinhardtii* genome, and the fatty acid profile of wildtype *C. reinhardtii* contains fatty acids comprised of a variety of chain lengths (C14-C18) and degrees of unsaturation [121], suggesting that the CrTE can siphon a range of acyl-ACP chains. Overexpression of the native TE in

C. reinhardtii resulted in a short-circuiting of fatty acids to increase myristic acid production (C14:0), which we propose is due to a stoichiometric imbalance of TE to ACP in the *C. reinhardtii* chloroplast and thus premature hydrolysis of fatty acids from the ACP [75] (see Chapter 2). In contrast, overexpression of PtTE in *P. tricornutum* did not result in altered fatty acid content [116].

Plant TEs have recently been engineered into eukaryotic microalgae to alter the fatty acid content for biodiesel production [88] [75]. Short-chain FatB TEs from *U. californica* (12:0-specific) [113] and *Cinnamomum camphora* (14:0-specific) [122] were genetically engineered into the diatom *P. tricornutum*, which redirected fatty acid production to the desired short-chain phenotype, albeit at low yield (6.2% increase in C12:0 and 15% increase in C14:0 of total fatty acids by weight) [88]. In a more recent example, plant TEs from *U. californica* and *C. hookeriana* were engineered in the chloroplast of *C. reinhardtii*, and no change in fatty acid profile was observed [75]. Additionally, a panel of 11 homologous plant FatB TEs were codon-optimized and engineered into the *C. reinhardtii* chloroplast, and no transgenic strain resulted in an altered fatty acid profile (unpublished results, Sapphire Energy), whereas in diatoms, a marginal effect on the fatty acid profile was observed using these plant TEs [88].

Diatoms and green algae are both promising as a photosynthetic biodiesel feedstock, but they contain different metabolic networks originating from different evolutionary histories [59]. Recent examples of engineering FAS enzymes into green algae highlight the difference in phenotypic output as compared to diatoms, perhaps pointing towards evolution as a tool to predict compatibility (see Chapter 3). This

contrast between green algal and diatoms highlights the diversity of algae, and the realization that as we discover new algal strains, we should aim to develop universal engineering strategies for a wide range of algal species.

Engineering fatty acid biosynthesis is a promising strategy to modify the fatty acid content of algal oils. As genomes of oleaginous algae emerge, characterization of algal FAS domains and their catalytic timing, interactions, and regulation should be a major research focus. Understanding protein-protein interactions between engineered FAS domains and the ACP is essential to functionally engineer fatty acid biosynthesis. Deciphering interactions between metabolic enzymes and uncovering regulation networks and ways to manipulate them will be necessary to achieve titers necessary for sustainable and economical algae biofuel production.

Engineering cyanobacterial FAS

Cyanobacteria are ancient photosynthetic prokaryotes that are evolutionary predecessors of the chloroplast [123]. They are naturally transformable, thus easily manipulated genetically, and can be engineered to secrete free fatty acids into the culture medium, offering a continuous algal production-secretion system that alleviates costs associated with harvesting, dewatering, and extracting algal lipids [124]. As such, there are many research groups interested in engineering fatty acid production and secretion in cyanobacteria [125]. In December of 2009, the first example of engineering fatty acid biosynthetic enzymes from plants into cyanobacteria was reported in *PNAS*, which reported that mutant cyanobacterial strains (*Synechocystis* sp. 6803) could secrete fatty acids into the culture medium at an

efficiency of 133 +/- 12 mg/ L at cell densities of 1.5 X 10^8 cells/ mL [126]. The authors from the Biodesign Institute and Laboratory for Algal Research and Biotechnology (LARB) at Arizona State University created five generations of mutant *Synechocystis* 6803 strains in which genetic insertions of FAS enzymes (aimed at increasing lipid yield and altering fatty acid chain length) replaced negative or competing genetic pathways for FFA production [126].

Extending the work performed over a decade ago by John Cronan at the University of Illinois using modified thioesterases to cause FA secretion in prokaryotes (*Escherichia coli*) [127], Curtiss et al. began with engineering the *E. coli* TE gene *tesA* to facilitate FA secretion in *Synechocystis* sp. 6803 Sun Devil (SD) strains, which replaced, and thus knocked out, the endogenous fatty acid-activating gene *aas* encoding ACP-synthetase, an enzyme recently found to activate exogenous fatty acids for reincorporation into membranes [128]. Both Ni^{2+}-inducible and constitutively expressed *tesA* SD strains were created. In Generation 2 SD strains, an artificial operon containing the *acc* gene was introduced into SD strains (knocking out two poly-3-hydroxybutyrate genes) under the control of a strong inducible promoter to overproduce ACCase subunits, hypothesizing an increased flux of carbon through the fatty acid biosynthetic pathway. To modify FA chain length, plant TEs from *U. californica* (12:0-specific) and *C. hookeriana* (8:0/10:0-specific) effectively knocked out proteins involved in the *S* layer of the cell wall, which increased fatty acids by three-fold. In Generations 4 and 5 SD strains, the *chfatB2* TE gene was engineered in place of the cyanophycin synthetase gene and the *ccfatB1* TE gene from *C. camphora* was inserted to replace the gene encoding the penicillin binding protein.

The authors claimed that SD strains could theoretically produce 6500 gallons of biodiesel per acre during a one year period at their predicted cell densities [126].

Due to a disagreement between authors, the article was retracted from *PNAS* [129]. One year after the initial publication, the work was republished in *PNAS* with the title, "Fatty Acid Production in Genetically Modified Cyanobacteria" [8]. Changes from the first study include the yield reported for fatty acid secretion, which authors revised, claiming improved SD strains could achieve a secretion efficiency of 197 +/- 14 mg/ L at cell densities of 1.0 X 10^9 cells/ mL [8]. Also, in this study, thioesterase gene sequences were codon-optimized for *Synechocystis* 6803 prior to insertion and a sixth genetic mutation was added to the five previously described. In Generation 6, an artificial operon encoding a *tesA137* gene was inserted to knock out a *pta* gene encoding a phosphotransacetylase [130], important for the survival of *Synechocystis* 6803 when grown in high light and a regulator of genes involved in energy metabolism. Here, the more standard method of analyzing fatty acids using GC/MS (rather than ESI-MS) was employed to characterize the fatty acid composition of transgenic cyanobacterial strains.

Taken together, the strategy of simultaneously knocking-out of genes in competing pathways coupled to the introduction of fatty acid biosynthetic enzymes, such as ACCase and plant TEs, appears to be a promising strategy to enhance fatty acid secretion in cyanobacteria, but one cannot extract from the data presented whether these changes are due to the genetic knockouts or FAS replacements. The authors conclude with a pivotal lesson learned in this work, that industrial algal strains

need high, durable biofuel productivity, robust cell growth, and cell rigidity, and that one should avoid genetic mutations that sacrifice growth and rigidity of algal cells.

Owing to their ease of genetic manipulation, cyanobacteria have also been engineered with heterologous enzymes from a range of biosynthetic pathways to produce a variety of biofuels [7], including hydrogen, alcohols (ethanol, butanol, isobutanol and isobutyraldehyde), hydrocarbons (ethylene, isoprene, and n-alkanes), and free fatty acids [131]. For example, the Kudzu (*Puereria montana*) isoprene synthase gene (*IspS*) was engineered into cyanobacteria for the photosynthetic production of the volatile hydrocarbon isoprene [132]. This creative strategy exploits the observation that cyanobacteria do not contain an isoprene synthase yet synthesize the penultimate precursor molecules DMAPP and IPP [132]. However, even though Melis et al. were able to achieve high (IspS) protein expression levels using the light-driven *psbA* promoter, they found that the overall carbon partitioning from photosynthetic carbon fixation was biased heavily towards sugars (~80-85%) and fatty acids (~10%), rather than terpenoids (~3-5%) [133], highlighting a major obstacle in engineering these types of secondary metabolic pathways. Nonetheless, there are many research groups and start-up companies (Joule Unlimited, Synthetic Genomics, Algenol, and Targeted Growth) that are working to overcome challenges involved in sustainable and economical development of cyanobacterial biofuels. For a review about cyanobacterial biofuels, please see reference [124].

Progress towards engineering algal fatty acid biosynthesis

Elucidating fatty acid biosynthesis in oil-accumulating species of microalgae

Screening algal strains to discover high lipid-producers is an important effort in algae biofuel research. It was heavily employed during the Aquatic Species Program (which generated a unique collection of 3000 species that accumulate oil and grow in extreme habitats), and currently, there is extensive research to screen algae for lipid content [134]. Metabolic engineering is an alternate approach to optimize oil production and robustness of algal strains, and this strategy can also be applied to oil-rich species.

However, fatty acid and lipid metabolism remain largely uncharacterized in algae, especially in oleaginous species, and the successful manipulation of these pathways requires a detailed understanding. As new oleaginous algal species emerge and their genomes become sequenced, high-throughput strategies must be developed to quickly annotate and characterize genes (and proteins) involved in fatty acid biosynthesis and lipid metabolism. Genomic, proteomic, and metabolomic techniques have been implemented in FAS elucidation efforts, and it is essential that we uncover the *interactome* of protein-protein interaction networks that occur inside algal cells on which many biological processes rely, in particular fatty acid biosynthesis.

In a recent study from the National Bioenergy Center at the National Renewable Energy Laboratory (Golden, Colorado), a comprehensive proteomic analysis of lipid accumulation in an unsequenced microalga was conducted using a *de novo* assembled transcriptome as a template for proteomic analysis [135], bypassing

genomics altogether. However, although proteomic profiling adds to our understanding of algal lipid biosynthesis, it is also important to link genetics with expression of enzymes involved in lipid metabolism to facilitate engineering.

A major breakthrough in algal biofuel biotechnology occurred in December of 2011, when a high-efficiency transformation protocol was described for oil-rich, fast-growing microalga *Nannochloropsis sp*. [136]. This represents a significant advancement towards elucidating metabolic networks and regulation in an industrially relevant model algal species that accumulates large amounts of oil [136]. Kilian et al. demonstrate that *Nannochloropsis* sp. exhibits efficient homologous recombination allowing for rapid and targeted genetic manipulation [136]. Additionally, three different selectable markers were developed (hygromycin, zeocin, and blasticidin resistance genes) and used to show high transformation rates. Notably, an 'activation tagging' protocol was developed, whereby Kilian et al. use a bidirectional promoter fused to the gene of interest in which the selectable marker gene is driven by one side of the promoter and the gene of interest is driven by the distal side of the promoter [136]. Also, *Nannochloropsis* sp. was found to be haploid, a positive attribute in a model genetic organism. This work advances the algal biofuel industry in establishing a new model species that is industrially relevant, as it biosynthesizes large amounts of oil and grows very quickly. It has also provided genetic tools to generate algal genetic knockout strains to study the largely uncharacterized metabolic networks that we aim to manipulate, such as fatty acid biosynthesis.

Using a traditional genetic approach, fatty acid biosynthesis of two microalgal species under stress conditions was recently examined. The first study, conducted at

the Pacific Northwest National Laboratories, explored the effects of light and temperature on fatty acid production in oil-rich microalga *Nannochloropsis salina*, which confirmed that in this species, total fatty acid content can be increased by exposing the algal culture to photooxidative stress and extreme temperatures, although this slows the growth of the organism [137]. In the second article, the expression of fatty acid synthesis genes and fatty acid accumulation in green microalga *Haematococcus pluvialis* was examined under different stressors. Five genes involved in fatty acid biosynthesis (ACP, KAS, TE, FAD – fatty acid desaturase, and MAT – malonyl ACP transacylase) were cloned from *H. pluvialis* and expression levels were correlated with fatty acid accumulation under a variety of stress conditions [138]. From their results, Lei et al. propose that the key rate-determining genes in the biosynthesis of fatty acids as those encoding the ACP, KS, and 'FatA' TE domains, as their expression displayed a linear relationship with fatty acid synthesis in *H. pluvialis* [138].

The authors likely annotated the HpTE as a 'FatA' TE due to homology with plant FatA TEs, specific for 18:1-ACP. However, closer inspection of the primary sequence of HpTE reveals a much higher degree of similarity to other microalgal TEs (*Chlorella variabilis*, *Chlamydomonas reinhardtii*, and *Volvox carteri*). As previously described, there now exists a distinct class of algal TEs called 'Fat1' which appear to be plant FatA / FatB hybrid TEs, thus more promiscuous in the substrates they act upon [75]. It is not surprising that HpTE was misannotated as a FatA TE, as this misnomer was also applied to PtTE and CrTE. Sequence alignment shows only dispersed regions of similarity with FatA TEs, and phylogenetic analysis depicts algal

TEs in a unique clade [75]. However, it is important to note that PtTE, a diatom TE, is different in sequence and structure than TEs present in green algae [116]. The discovery of this novel class of Fat1 TEs in microalgae highlights a critical, but often-overlooked fact, that microalgae are not plants. Although similar to plants, microalgae contain a more primitive metabolism than plants containing enzymes that may exhibit dual functions. Therefore, characterization of enzymes involved in algal fatty acid and lipid metabolism is essential for efficient engineering.

Protein-protein interactions

Using green microalga *C. reinhardtii* as a model, it was recently demonstrated that protein-protein interactions between the fatty acid acyl carrier protein (ACP) and thioesterase (TE) govern fatty acid hydrolysis within the algal chloroplast [75] (see Chapter 2). A structural docking simulation was conducted between the CrACP binding to CrTE, which clearly illustrates a protein-protein recognition surface between the two proteins, and an electrostatic potential map shows compatibility between the CrACP and CrTE [75]. A virtual screen of homologous plant TEs was used to identify those with similar *in silico* binding to CrACP. Two plant TE candidates were selected, a 12:0-specific FatB TE from *U. californica* [113] and an 8:0/10:0-selective FatB TE from *C. hookeriana* [120], and purified TE proteins were obtained. Employing an activity-based crosslinking probe designed to selectively trap transient protein-protein interactions between the TE and ACP [139] [140], it was demonstrated *in vitro* that CrTE must functionally interact with CrACP to release fatty acids, while vascular plant TEs did not show mechanistic crosslinking to CrACP.

This phenomenon was also observed *in vivo* in the *C. reinhardtii* chloroplast. Whereas overproduction of the endogenous CrTE increased levels of short-chain fatty acids, engineering plant TEs into the *C. reinhardtii* chloroplast did not significantly alter the fatty acid profile [75]. These findings highlight the critical role of protein-protein interactions in manipulating fatty acid biosynthesis for algae biofuel engineering. Using activity-based chemical probes inspired by the enzymatic activity of FAS domains is a creative approach towards elucidating functional protein-protein interactions in the algal chloroplast.

Compatibility of FAS enzymes across species

Recent engineering attempts have highlighted both a lack of fundamental understanding of algal fatty acid biosynthesis and the interactions between catalytic domains. As the ACP must catalytically interact with all FAS enzymes in the biosynthesis of a fatty acid, for successful engineering of FAS, one must consider protein-protein interactions between the target FAS enzyme and the ACP of the host organism. Initial studies have shown that introducing heterologous enzymes from plants into eukaryotic microalgae is not as straightforward as in other organisms. For example, it was not clear at the onset that plant TEs would not function in green algae or diatoms, nor that overexpression of native ACCase would not increase flux through the FAS pathway in diatoms. Similar strategies have been successful in prokaryotes, such as TE engineering in cyanobacteria [141], which altered the fatty acid content, and in eukaryotes, such as ACCase overexpression in potatoes (*Solanum tuberosum*), which increased total fatty acid content fivefold [142]. Taken together, these FAS

engineering have illuminated an apparent directionality of FAS enzymes to interact with ACPs in organisms of related species that is dependent on the FAS domain of interest and its evolutionary origin.

Our laboratory has recently shown that protein-protein interactions between the ACP and TE mediate the hydrolysis of fatty acids in green algae [75]. We have also shown that ACP-KS interactions are required in bacterial fatty acid biosynthesis [139]. In combination with the results obtained when engineering plant TEs into algae and phylogenetic analyses, we hypothesized that ACP-KS interactions are more interchangeable across species than ACP-TE interactions, as KS domains all derive from a common thiolase ancestor [143] (see Chapter 3). We also sought to answer the question of the evolutionary origin of the algal TE and found evidence that algal Fat1 TEs are derived from an ancient anaerobic bacterium, rather than a cyanobacterium as previously thought [144]. From this, we postulate that ACP-TE interactions are more selective and unidirectional, a consequence of divergent evolution.

We investigated the compatibility between FAS enzymes and ACPs across species using phylogenetic analyses, *in silico* protein-protein docking, *in vitro* activity-based crosslinking between FAS enzymes and ACPs, and *in vivo* engineering of FAS enzymes. *In vitro* crosslinking was used to screen for ACP-KS interactions using a probe designed to inhibit the enzymatic activity of the KS [139], which successfully trapped interactions between the algal ACP and bacterial KS [144]. Interactions between the algal ACP and plant KS from *Ricinus communis* were demonstrated using *in vivo* engineering [144], providing evidence that ACP-KS interactions are interchangeable in both evolutionary directions. *In vitro* crosslinking between the

bacterial ACP loaded with a TE-specific α-bromopalmitamide probe shows limited crosslinking with the algal CrTE, again illustrating directionality, as plant TEs did not crosslink with the algal ACP [144]. This work has shown that ACP-KS interactions are permissive across species and that ACP-TE interactions are selective, a proposal based on the evolutionary origin of each distinct FAS enzyme [144]. The goal of this research was to use evolution as a tool to predict successful protein-protein interactions between FAS enzymes and ACPs, and ultimately to enable the metabolic engineering of FAS in any species.

Conclusions

Microalgae are reemerging as a viable biofuel-producing system due to their vast lipid yields and flexible habitats. However, microalgae accumulate lipids under stress conditions, which hinders growth. Metabolic engineering promises a better level of control and predictability of the system, and is more ideal than nutrient deprivation techniques, which slow algal growth rates [19]. To optimize microalgae as a biodiesel feedstock, there are two main goals. The first major aim is to increase oil content, and the second goal is to engineer fatty acid chain length to fashion an oxidatively stable algal biodiesel fuel that is compliant with the current infrastructure. Engineering algal fatty acid biosynthesis promises to create microalgal strains capable of producing fungible biofuels on scale in a sustainable fashion (Figure 1-3).

Recent efforts to engineer fatty acid biosynthesis have been focused on ACCases to enhance lipid yields and thioesterases to alter fatty acid chain length (Figure 1-3). As diatoms and green algae were identified by the ASP as the most

promising biodiesel feedstock organisms, plant TEs were engineered into model diatom *P. tricornutum* and model green alga *C. reinhardtii* to create strains that overproduce short-chain fatty acids in their oil [88] [75]. However, this did not result in high yields of short-chain fatty acids, as plant TEs did not functionally interact with CrACP *in vivo* as well as *in vitro*. Furthermore, overexpression of ACCase did not result in increased oil production in diatoms (Figure 1-5). Taken together, results suggest that new strategies must be envisioned to enhance fatty acid content and oil yield.

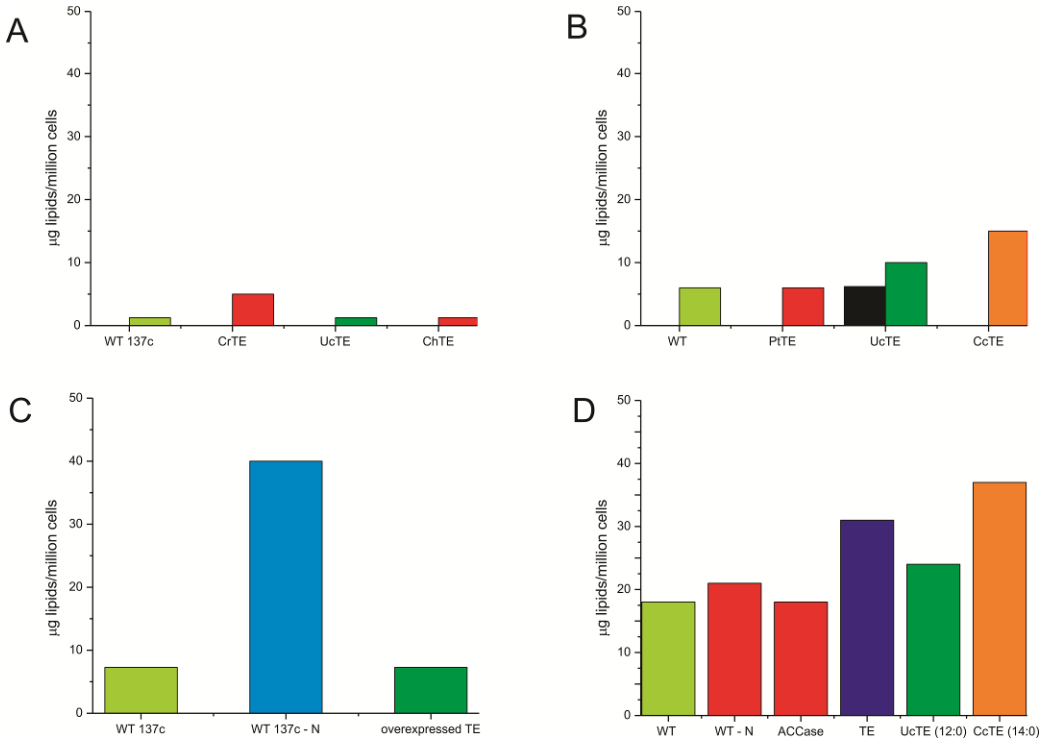


Figure 1-5 Current yields in efforts to engineer fatty acid biosynthetic enzymes into green algae and diatoms for enhanced oil production and to modify oil composition

(a) Yields achieved in engineering efforts aimed at increasing amount of oil in green algae (*C. reinhardtii*). (b) Yields achieved in engineering efforts aimed at modifying fatty acid chain length in green algae (*C. reinhardtii*). (c) Yields achieved in engineering efforts aimed at increasing amount of oil in diatoms (*P. tricornutum*). (d) Yields achieved in engineering efforts aimed at modifying fatty acid chain length in diatoms (*P. tricornutum*).

Biochemical characterization of algal FAS is essential for successful engineering. Recent discoveries illuminate the importance of interactions between FAS domains and ACP in algal FAS. Protein-protein interactions between the ACP and TE control the identity of the fatty acid end product, which was demonstrated in green algae using activity-based *in vitro* crosslinking and *in vivo* engineering. Creative tools have been developed, such as the use of activity-based chemical probes to trap FAS enzymes and study their interactions with endogenous biosynthetic machinery in algal chloroplasts. Evolution can also be used as a tool to predict compatibility between algal ACPs and FAS enzymes of interest for successful engineering. Initial characterization has begun to shift the paradigm from treating algae as a 'black box' in which genes are engineered without a thorough understanding of algal metabolism and regulation. The output is often unpredictable and yields are less than desirable. With new model organisms, such as oil-rich microalga *Nannochloropsis*, and ways to genetically alter them, we can begin to understand mechanisms underlying oil accumulation so that we can better manipulate these pathways for economical algal biofuel production.

Outlook

The second wave of algae biofuel research has brought new sequencing and transformation technologies, a plethora of bioinformatics data about algal metabolism and its responses to stress, and some ingenious strategies aimed at optimizing algae as a feedstock have been envisioned.

The Aquatic Species Program laid the foundation for algae research, but there is still a lot to be determined, as current research has highlighted unforeseen challenges. For example, efforts to engineer microalgae with heterologous FAS enzymes have only been marginally successful (Figure 1-5), whereas much higher yields were observed when these same enzymes were engineered into plants and prokaryotes. Studies from our laboratory and others have illuminated that protein-protein interactions and metabolic regulation should be considered in engineering FAS. Functional interaction with the endogenous algal ACP is required for successful FAS engineering, and as genomes of industrially significant algal species emerge, characterization of algal ACPs will be essential. However, the question remains as to whether algal FAS is rate limiting. As researchers from a wide range of disciplines apply their expertise to solving current bottlenecks in algae biodiesel production, we get closer to realizing a sustainable future with economical algal biofuel production.

References

1. Vitousek PM, Mooney HA, Lubchenco J, Melillo JM (1997) Human domination of Earth's ecosystems. Science 277: 494-499.

2. Cooper JE (1962) Fatty acids in recent and ancient sediments and petroleum reservoir waters. Nature 193: 744-746.

3. Dresselhaus MS, Thomas IL (2001) Alternative energy technologies. Nature 414: 332-337.

4. Sheldon KS, Yang S, Tewksbury JJ (2011) Climate change and community disassembly: impacts of warming on tropical and temperate montane community structure. Ecology Letters 14: 1191-1200.

5. Rommens CM (2010) Barriers and paths to market for genetically engineered crops. Plant Biotechnology Journal 8: 101-111.

6. Golden SS, Brusslan J, Haselkorn R (1987) Genetic engineering of the cyanobacterial chromosome. Methods in Enzymology 153: 215-231.

7. Ruffing AM (2011) Engineered cyanobacteria: Teaching an old bug new tricks. Bioengineered Bugs 2: 136-149.

8. Liu X, Sheng J, Curtiss III R (2011) Fatty acid production in genetically modified cyanobacteria. Proceedings of the National Academy of Sciences, USA 108: 6899-6904.

9. Sharma NK, Tiwari SP, Tripathi K, Rai AK (2011) Sustainability and cyanobacteria (blue-green algae): Facts and challenges. Journal of Applied Phycology 23: 1059-1081.

10. Dismukes GC, Carrieri D, Bennette N, Ananyev GM, Posewitz MC (2008) Aquatic phototrophs: Efficient alternatives to land-based crops for biofuels. Current Opinion in Biotechnology 19: 235-240.

11. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A (2008) Microalgal triacylglycerols as feedstocks for biofuel production: Perspectives and advances. Plant Journal 54: 621-639.

12. Specht E, Miyake-Stoner S, Mayfield S (2010) Micro-algae come of age as a platform for recombinant protein production. Biotechnology Letters 32: 1373-1383.

13. Siaut M, Cuine S, Cagnon C, Fessler B, Nguyen M, Carrier P, Beyly A, Beisson F, Triantaphylides C, Li-Beisson Y, Peltier G (2011) Oil accumulation in the model green alga *Chlamydomonas reinhardtii*: Characterization, variability between common laboratory strains and relationship with starch reserves. BMC Biotechnology 11: 7.

14. Wang ZT, Ullrich N, Joo S, Waffenschmidt S, Goodenough U (2009) Algal lipid bodies: Stress induction, purification, and biochemical characterization in wild-type and starchless *Chlamydomonas reinhardtii*. Eukaryotic Cell 8: 1856-1868.

15. Phillips SL, Wolfe MS (2005) Evolutionary plant breeding for low input systems. Journal of Agricultural Science 143: 245-254.

16. Chisti Y (2007) Biodiesel from microalgae. Biotechnology Advances 25: 294-306.

17. Thelen JJ, Ohlrogge JB (2002) Metabolic engineering of fatty acid biosynthesis in plants. Metabolic Engineering 4: 12-21.

18. Rosenberg JN, Oyler GA, Wilkinson L, Betenbaugh MJ (2008) A green light for engineered algae: Redirecting metabolism to fuel a biotechnology revolution. Current Opinion in Biotechnology 19: 430-436.

19. Radakovits R, Jinkerson RE, Darzins A, Posewitz MC (2010) Genetic engineering of algae for enhanced biofuel production. Eukaryotic Cell 9: 486-501.

20. Sommerfeld MR, Hu Q (2009) Microalgae-based biofuel: A perspective on the state of the science. Journal of Phycology 45: 18-18.

21. Mayfield S (2011) Microalgae for the production of biofuels and bio-products. Journal of Phycology 47: S5-S6.

22. Henley WJ, Duke CS, Litaker WR, Quemada H, Sayre RT, Shore S (2011) Ecological risks of genetically modified algae in commodity-scale cultivation. Journal of Phycology 47: S57-S57.

23. Sheehan J, Dunahay T, Benemann J, Roessler P (1998) A Look Back at the U.S. Department of Energy's Aquatic Species Program - Biodiesel from Algae. DOE: National Renewable Energy Laboratory.

24. Melis A, Happe T (2004) Trails of green alga hydrogen research - from Hans Gaffron to new frontiers. Photosynthesis Research 80: 401-409.

25. Melis A, Happe T (2001) Hydrogen production. Green algae as a source of energy. Plant Physiology 127: 740-748.

26. Calvin M, Benson AA (1948) The path of carbon in photosynthesis. Science 107: 476-480.

27. Rochaix JD, Goldschmidt-Clermont M, Merchant SS (1998) The Molecular Biology of Chloroplasts and Mitochondria in Chlamydomonas. Kluwer Academic Publishers: Dordrecht, The Netherlands.

28. Duysens LNM, Sweers HE (1963) Mechanism of two photochemical reactions in algae as studied by means of fluorescence. In: Japanese Society of Plant Physiologists (ed) Studies on Microalgae and Photosynthetic Bacteria. University of Tokyo Press: Tokyo, pp. 353-372.

29. Levine RP, Volkmann D (1961) Mutants with impaired photosynthesis in *Chlamydomonas reinhardi*. Biochemical and Biophysical Research Communications 6: 264-269.

30. Pratt LH, Bishop NI (1968) Chloroplast reactions of photosynthetic mutants of *Scenedesmus obliquus*. Biochimica et Biophysica Acta 153: 664-674.

31. Bishop NI (1962) Mechanism of oxygen production in photosynthesis - pigment and physiological mutants of green alga, *Scenedesmus obliquus*. Journal of General Physiology 45: A592.

32. Davies JP, Grossman AR (1998) The use of Chlamydomonas (Chlorophyta : Volvocales) as a model algal system for genome studies and the elucidation of photosynthetic processes. Journal of Phycology 34: 907-917.

33. Harris EH (1989) The Chlamydomonas sourcebook: A comprehensive guide to biology and laboratory use. Academic Press: San Diego, CA.

34. Bennoun P, Levine RP (1967) Detecting mutants that have impaired photosynthesis by their increased level of fluorescence. Plant Physiology 42: 1284-1287.

35. Gorman DS, Levine RP (1965) Cytochrome F and plastocyanin - their sequence in photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. Proceedings of the National Academy of Sciences, USA 54: 1665-1669.

36. Ebersold WT, Levine RP, Levine EE, Olmsted MA (1962) Linkage maps in *Chlamydomonas reinhard*. Genetics 47: 531-543.

37. Ow DW, Wood KV, Deluca M, Dewet JR, Helinski DR, Howell SH, Wood KV, Deluca M (1986) Transient and stable expression of the firefly luciferase gene in plant cells and transgenic plants. Science 234: 856-859.

38. Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM, Jones AR, Randolph-Anderson BL, Robertson D, Klein TM, Shark KB, et al. (1988) Chloroplast transformation in Chlamydomonas with high velocity microprojectiles. Science 240: 1534-1538.

39. Klein TM, Kornstein L, Sanford JC, Fromm ME (1989) Genetic transformation of maize cells by particle bombardment. Plant Physiology 91: 440-444.

40. Harris EH (2001) Chlamydomonas as a model organism. Annual Review of Plant Physiology and Plant Molecular Biology 52: 363-406.

41. Mayfield SP, Manuell AL, Chen S, Wu J, Tran M, Siefker D, Muto M, Marin-Navarro J (2007) *Chlamydomonas reinhardtii* chloroplasts as protein factories. Current Opinion in Biotechnology 18: 126-133.

42. Pulz O, Gross W (2004) Valuable products from biotechnology of microalgae. Applied Microbiology and Biotechnology 65: 635-648.

43. Knothe G, Van Gerpen J, J K (2005) The Biodiesel Handbook. AOCS Press: Champaign, Illinois.

44. Meier RL (1955) Biological cycles in the transformation of solar energy into useful fuels. In: Daniels F, Duffie JA (eds). Solar Energy Research. University of Wisconsin Press: Madison, WI, pp. 179-183.

45. Golueke CG, Oswald WJ, Gotaas HB (1957) Anaerobic digestion of algae. Applied Microbiology 5: 47-55.

46. Oswald WJ, Gotaas HB, Golueke CG, Kellen WR (1957) Algae in waste treatment. Sewage and Industrial Wastes 29: 437-455.

47. Pimentel D, Hurd LE, Bellotti AC, Forster MJ, Oka IN, Sholes OD, Whitman RJ (1973) Food production and energy crisis. Science 182: 443-449.

48. Benemann JR, Weissman JC, Koopman BL, Oswald WJ (1977) Energy production by microbial photosynthesis. Nature 268: 19-23.

49. Guschina IA, Harwood JL (2006) Lipids and lipid metabolism in eukaryotic algae. Progress in Lipid Research 45: 160-186.

50. Roessler PG, Brown LM, Dunahay TG, Heacox DA, Jarvis EE, Schneider JC, Talbot SG, Zeiler KG (1994) Genetic engineering approaches for enhanced production of biodiesel fuel from microalgae. Enzymatic Conversion of Biomass for Fuels Production. ACS Symposium Series 566: 255-270.

51. Franklin SE, Mayfield SP (2005) Recent developments in the production of human therapeutic proteins in eukaryotic algae. Expert Opinion on Biological Therapy 5: 225-235.

52. Huppe HC, Turpin DH (1994) Integration of carbon and nitrogen metabolism in plant and algal cells. Annual Review of Plant Physiology and Plant Molecular Biology 45: 577-607.

53. Raven JA, Girard-Bascou J (2001) Algal model systems and the elucidation of photosynthetic metabolism. Journal of Phycology 37: 943-950.

54. Cavalier-Smith T (2000) Membrane heredity and early chloroplast evolution. Trends in Plant Science 5: 174-182.

55. Grossman AR (2005) Paths toward algal genomics. Plant Physiology 137: 410-427.

56. Walker TL, Collet C, Purton S (2005) Algal transgenics in the genomic ERA. Journal of Phycology 41: 1077-1093.

57. Ryall K, Harper JT, Keeling PJ (2003) Plastid-derived type II fatty acid biosynthetic enzymes in chromists. Gene 313: 139-148.

58. Goodenough UW, Armbrust EV, Campbell AM, Ferris PJ (1995) Molecular genetics of sexuality in Chlamydomonas. Annual Review of Plant Physiology and Plant Molecular Biology 46: 21-44.

59. Reyes-Prieto A, Weber APM, Bhattacharya D (2007) The origin and establishment of the plastid in algae and plants. Annual Review of Genetics 41: 147-168.

60. Keeling PJ (2010) The endosymbiotic origin, diversification and fate of plastids. Philosophical Transactions of the Royal Society B - Biological Sciences 365: 729-748.

61. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, et al. (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318: 245-251.

62. Felsner G, Sommer MS, Maier UG (2010) The physical and functional borders of transit peptide-like sequences in secondary endosymbionts. BMC Plant Biology 10: 223.

63. Larkum AWD, Ross IL, Kruse O, Hankamer B (2012) Selection, breeding and engineering of microalgae for bioenergy and biofuel production. Trends in Biotechnology 30: 198-205.

64. Cross FR, Buchler NE, Skotheim JM (2011) Evolution of networks and sequences in eukaryotic cell cycle control. Philosophical Transactions of the Royal Society B - Biological Sciences 366: 3532-3544.

65. Yu WL, Ansari W, Schoepp NG, Hannon MJ, Mayfield SP, Burkart MD (2011) Modifications of the metabolic pathways of lipid and triacylglycerol production in microalgae. Microbial Cell Factories 10: 91.

66. Kindle KL (1990) High-frequency nuclear transformation of *Chlamydomonas reinhardtii*. Proceedings of the National Academy of Sciences, USA 87: 1228-1232.

67. Grossman AR, Harris EE, Hauser C, Lefebvre PA, Martinez D, Rokhsar D, Shrager J, Silflow CD, Stern D, Vallon O, Zhang Z (2003) *Chlamydomonas reinhardtii* at the crossroads of genomics. Eukaryotic Cell 2: 1137-1150.

68. Manuell A, Beligni MV, Yamaguchi K, Mayfield SP (2004) Regulation of chloroplast translation: interactions of RNA elements, RNA-binding proteins and the plastid ribosome. Biochemical Society Transactions 32: 601-605.

69. Franklin S, Ngo B, Efuet E, Mayfield SP (2002) Development of a GFP reporter gene for *Chlamydomonas reinhardtii* chloroplast. Plant Journal 30: 733-744.

70. Mayfield SP, Schultz J (2004) Development of a luciferase reporter gene, luxCt, for *Chlamydomonas reinhardtii* chloroplast. Plant Journal 37: 449-458.

71. Mayfield SP, Franklin SE, Lerner RA (2003) Expression and assembly of a fully active antibody in algae. Proceedings of the National Academy of Sciences, USA 100: 438-442.

72. Rohr J, Sarkar N, Balenger S, Jeong BR, Cerutti H (2004) Tandem inverted repeat system for selection of effective transgenic RNAi strains in Chlamydomonas. Plant Journal 40: 611-621.

73. Molnar A, Bassett A, Thuenemann E, Schwach F, Karkare S, Ossowski S, Weigel D, Baulcombe D (2009) Highly specific gene silencing by artificial microRNAs in the unicellular alga *Chlamydomonas reinhardtii*. Plant Journal 58: 165-174.

74. Sirevag R, Levine RP (1972) Fatty acid synthetase from *Chlamydomonas reinhardi* - sites of transcription and translation. Journal of Biological Chemistry 247: 2586-2591.

75. Blatti JL, Beld J, Behnke C, Mendez M, Mayfield S, Burkart MD (2012) Manipulating fatty acid biosynthesis in microalgae for biofuel through protein-protein interactions. Submitted.

76. Verhounig A, Karcher D, Bock R (2010) Inducible gene expression from the plastid genome by a synthetic riboswitch. Proceedings of the National Academy of Sciences, USA 107: 6204-6209.

77. Ferrante P, Catalanotti C, Bonente G, Giuliano G (2008) An optimized, chemically regulated gene expression system for Chlamydomonas. Plos One 3: e3200.

78. Rasala BA, Muto M, Sullivan J, Mayfield SP (2011) Improved heterologous protein expression in the chloroplast of *Chlamydomonas reinhardtii* through promoter and 5 ' untranslated region optimization. Plant Biotechnology Journal 9: 674-683.

79. US DOE (2008) National Algal Biofuels Technology Roadmap. U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Biomass Program.

80. Guan W, Zhao H, Lu X, Wang C, Yang M, Bai F (2011) Quantitative analysis of fatty-acid-based biofuels produced by wild-type and genetically engineered cyanobacteria by gas chromatography-mass spectrometry. Journal of Chromatography A 1218: 8289-8293.

81. Robertson DE, Jacobson SA, Morgan F, Berry D, Church GM, Afeyan NB (2011) A new dawn for industrial photosynthesis. Photosynthesis Research 107: 269-277.

82. Mata TM, Martins AA, Caetano NS (2010) Microalgae for biodiesel production and other applications: a review. Renewable and Sustainable Energy Reviews 14: 217-232.

83. Melis A, Seibert M, Ghirardi ML (2007) Hydrogen fuel production by transgenic microalgae. In: Leon R, Gavan A, Fernandez E (eds) Transgenic Microalgae as Green Cell Factories. Landes Bioscience: Austin, Texas. Chapter 10, pp. 108-121.

84. Beer LL, Boyd ES, Peters JW, Posewitz MC (2009) Engineering algae for biohydrogen and biofuel production. Current Opinion in Biotechnology 20: 264-271.

85. Eberle U, Felderhoff M, Schueth F (2009) Chemical and Physical Solutions for Hydrogen Storage. Angewandte Chemie - International Edition 48: 6608-6630.

86. Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB (2010) Microbial Biosynthesis of Alkanes. Science 329: 559-562.

87. Alper H, Stephanopoulos G (2009) Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? Nature Reviews Microbiology 7: 715-723.

88. Radakovits R, Eduafo PM, Posewitz MC (2011) Genetic engineering of fatty acid chain length in *Phaeodactylum tricornutum*. Metabolic Engineering 13: 89-95.

89. Jamers A, Blust R, De Coen W (2009) Omics in algae: Paving the way for a systems biological understanding of algal stress phenomena? Aquatic Toxicology 92: 114-121.

90. Yu WL, Ansari W, Schoepp NG, Hannon MJ, Mayfield SP, Burkart MD (2011) Modifications of the metabolic pathways of lipid and triacylglycerol production in microalgae. Microbial Cell Factories 10: 91.

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

92. Harwood JL (1996) Recent advances in the biosynthesis of plant fatty acids. Biochimica et Biophysica Acta - Lipids and Lipid Metabolism 1301: 7-56.

93. Postbeittenmiller D, Jaworski JG, Ohlrogge JB (1991) In vivo pools of free and acylated acyl carrier proteins in spinach - evidence for sites of regulation of fatty acid biosynthesis. Journal of Biological Chemistry 266: 1858-1865.

94. Postbeittenmiller D, Roughan G, Ohlrogge JB (1992) Regulation of plant fatty acid biosynthesis - analysis of acyl-coenzyme A and acyl-acyl-carrier-protein substrate pools in spinach and pea chloroplasts. Plant Physiology 100: 923-930.

95. Zornetzer GA, Fox BG, Markley JL (2006) Solution structures of spinach acyl carrier protein with decanoate and stearate. Biochemistry 45: 5217-5227.

96. Rawsthorne S (2002) Carbon flux and fatty acid synthesis in plants. Progress in Lipid Research 41: 182-196.

97. Voelker T (1996) Plant acyl-ACP-thioesterases: chain length determining enzymes in plant fatty acid biosynthesis. In: Setlow JK (ed) Genetic Engineering, Vol. 18. Plenum Press: New York, pp. 111-133.

98. Ohlrogge J, Browse J (1995) Lipid biosynthesis. Plant Cell 7: 957-970.

99. Jang J-H, Lee CS, Hwang D, Ryu SH (2012) Understanding of the roles of phospholipase D and phosphatidic acid through their binding partners. Progress in Lipid Research 51: 71-81.

100. Lopez D, Casero D, Cokus SJ, Merchant SS, Pellegrini M (2011) Algal functional annotation tool: a web-based analysis suite to functionally interpret large gene lists using integrated annotation and expression data. BMC Bioinformatics 12: 282.

101. Wu F, Yang Z, Kuang T (2006) Impaired photosynthesis in phosphatidylglycerol-deficient mutant of cyanobacterium Anabaena sp PCC7120 with a disrupted gene encoding a putative phosphatidylglycerophosphatase. Plant Physiology 141: 1274-1283.

102. Giroud C, Gerber A, Eichenberger W (1988) Lipids of *Chlamydomonas reinhardtii* - analysis of molecular species and intracellular site(s) of biosynthesis. Plant and Cell Physiology 29: 587-595.

103. Benning C (2010) The Anionic Chloroplast Membrane Lipids: Phosphatidylglycerol and Sulfoquinovosyldiacylglycerol. In: Constantin A, Rebeiz et al. (eds) Chloroplast: Basics and Applications. Springer: Dordrecht, The Netherlands. Chapter 12, pp. 171-183.

104. Moellering ER, Miller R, Benning C (2009) Molecular Genetics of Lipid Metabolism in the Model Green Alga *Chlamydomonas reinhardtii*. In: Wada H, Murata N (eds) Lipids in Photosynthesis: Essential and Regulatory Functions. Springer: Dordrecht, The Netherlands. Chapter 7, pp. 139-155.

105. Riekhof WR, Sears BB, Benning C (2005) Annotation of genes involved in glycerolipid biosynthesis in *Chlamydomonas reinhardtii*: Discovery of the betaine lipid synthase BTA1(Cr). Eukaryotic Cell 4: 242-252.

106. Harwood JL (2011) Algae: diverse lipids with many uses. Journal of Phycology 47: 49-49.

107. Harwood JL, Guschina IA (2009) The versatility of algae and their lipid metabolism. Biochimie 91: 679-684.

108. Ohlrogge JB, Jaworski JG (1997) Regulation of fatty acid synthesis. Annual Review of Plant Physiology and Plant Molecular Biology 48: 109-136.

109. Nguyen HM, Baudet M, Cuine S, Adriano J-M, Barthe D, Billon E, Bruley C, Beisson F, Peltier G, Ferro M, Li-Beisson Y (2011) Proteomic profiling of oil bodies isolated from the unicellular green microalga *Chlamydomonas reinhardtii*: With focus on proteins involved in lipid metabolism. Proteomics 11: 4266-4273.

110. Moellering ER, Benning C (2010) RNA interference silencing of a major lipid droplet protein affects lipid droplet size in *Chlamydomonas reinhardtii*. Eukaryotic Cell 9: 97-106.

111. Goodson C, Roth R, Wang ZT, Goodenough U (2011) Structural correlates of cytoplasmic and chloroplast lipid body synthesis in *Chlamydomonas reinhardtii* and stimulation of lipid body production with acetate boost. Eukaryotic Cell 10: 1592-1606.

112. Fan J, Andre C, Xu C (2011) A chloroplast pathway for the de novo biosynthesis of triacylglycerol in *Chlamydomonas reinhardtii*. FEBS Letters 585: 4029-4029.

113. Voelker TA, Worrell AC, Anderson L, Bleibaum J, Fan C, Hawkins DJ, Radke SE, Davies HM (1992) Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. Science 257: 72-74.

114. Roessler PG (1988) Changes in the activities of various lipid and carbohydrate biosynthetic enzymes in the diatom *Cyclotella cryptica* in response to silicon deficiency. Archives of Biochemistry and Biophysics 267: 521-528.

115. Livne A, Sukenik A (1990) Acetyl coenzyme A carboxylase from the marine Prymnesiophyte *Isochrysis galbana*. Plant and Cell Physiology 31: 851-858.

116. Gong Y, Guo X, Wan X, Liang Z, Jiang M (2011) Characterization of a novel thioesterase (PtTE) from *Phaeodactylum tricornutum*. Journal of Basic Microbiology 51: 666-672.

117. Salas JJ, Ohlrogge JB (2002) Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. Archives of Biochemistry and Biophysics 403: 25-34.

118. Voelker T, Kinney AT (2001) Variations in the biosynthesis of seed-storage lipids. Annual Review of Plant Physiology and Plant Molecular Biology 52: 335-361.

119. Graham SA (1989) Cuphea - a new source of medium-chain fatty acids. Critical Reviews in Food Science and Nutrition 28: 139-173.

120. Dehesh K, Jones A, Knutzon DS, Voelker TA (1996) Production of high levels of 8:0 and 10:0 fatty acids in transgenic canola by overexpression of Ch FatB2, a thioesterase cDNA from *Cuphea hookeriana*. Plant Journal 9: 167-172.

121. James GO, Hocart CH, Hillier W, Chen H, Kordbacheh F, Price GD, Djordjevic MA (2011) Fatty acid profiling of *Chlamydomonas reinhardtii* under nitrogen deprivation. Bioresource Technology 102: 3343-3351.

122. Yuan L, Voelker TA, Hawkins DJ (1995) Modification of the substrate specificity of an acyl-acyl carrier protein thioesterase by protein engineering. Proceedings of the National Academy of Sciences, USA 92: 10639-10643.

123. Linka M, Weber APM (2012) Evolutionary Integration of Chloroplast Metabolism with the Metabolic Networks of the Cells. In: Burnap RL and Vermaas WFJ (eds) Functional Genomics and Evolution of Photosynthetic Systems. Springer: Dordrecht, The Netherlands. Chapter 8, pp. 199-224.

124. Angermayr SA, Hellingwerf KJ, Lindblad P, de Mattos MJT (2009) Energy biotechnology with cyanobacteria. Current Opinion in Biotechnology 20: 257-263.

125. Lu X (2010) A perspective: photosynthetic production of fatty acid-based biofuels in genetically engineered cyanobacteria. Biotechnology Advances 28: 742-746.

126. Liu X, Brune D, Vermaas W, Curtiss III R (2010) Production and secretion of fatty acids in genetically engineered cyanobacteria. Proceedings of the National Academy of Sciences, USA 107: 13189.

127. Cho HS, Cronan JE (1995) Defective export of a periplasmic enzyme disrupts regulation of fatty acid synthesis. Journal of Biological Chemistry 270: 4216-4219.

128. Kaczmarzyk D, Fulda M (2010) Fatty acid activation in cyanobacteria mediated by acyl-acyl carrier protein synthetase enables fatty acid recycling. Plant Physiology 152: 1598-1610.

129. Schekman R (2010) Retraction for Liu et al.: Production and secretion of fatty acids in genetically engineered cyanobacteria. Proceedings of the National Academy of Sciences, USA 107: 13189.

130. Morrison SS, Mullineaux CW, Ashby MK (2005) The influence of acetyl phosphate on DspA signalling in the cyanobacterium Synechocystis sp PCC6803. BMC Microbiology 5: 49.

131. Quintana N, Van Der Kooy F, Van De Rhee MD, Voshol GP, Verpoorte R (2011) Renewable energy from cyanobacteria: Energy production optimization by metabolic pathway engineering. Applied Microbiology and Biotechnology 91: 471-490.

132. Melis A (2012) Photosynthesis-to-fuels: From sunlight to hydrogen, isoprene, and botryococcene production. Energy and Environmental Science 5: 5531-5539.

133. Lindberg P, Park S, Melis A (2010) Engineering a platform for photosynthetic isoprene production in cyanobacteria, using Synechocystis as the model organism. Metabolic Engineering 12: 70-79.

134. Stockenreiter M, Graber A-K, Haupt F, Stibor H (2012) The effect of species diversity on lipid production by micro-algal communities. Journal of Applied Phycology 24: 45-54.

135. Guarnieri MT, Nag A, Smolinski SL, Darzins A, Seibert M, Pienkos PT (2011) Examination of triacylglycerol biosynthetic pathways via de novo transcriptomic and proteomic analyses in an unsequenced microalga. Plos One 6: e25851.

136. Kilian O, Benemann CSE, Niyogi KK, Vick B (2011) High-efficiency homologous recombination in the oil-producing alga Nannochloropsis sp. Proceedings of the National Academy of Sciences, USA 108: 21265-21269.

137. Van Wagenen J, Miller TW, Hobbs S, Hook P, Crowe B, et al. (2012) Effects of light and temperature on fatty acid production in *Nannochloropsis salina*. Energies 5: 731-740.

138. Lei A, Chen H, Shen G, Hu Z, Chen L, Wang J (2012) Expression of fatty acid synthesis genes and fatty acid accumulation in *Haematococcus pluvialis* under different stressors. Biotechnology for Biofuels 5: 18.

139. Worthington AS, Rivera H, Jr., Torpey JW, Alexander MD, Burkart MD (2006) Mechanism-based protein crosslinking probes to investigate carrier protein-mediated biosynthesis. ACS Chemical Biology 1: 687-691.

140. Worthington AS, Porter DF, Burkart MD (2010) Mechanism-based crosslinking as a gauge for functional interaction of modular synthases. Organic and Biomolecular Chemistry 8: 1769-1772.

141. Liu X, Sheng J, Curtiss III R (2011) Fatty acid production in genetically modified cyanobacteria. Proceedings of the National Academy of Sciences, USA 108: 6899-6904.

142. Roesler K, Shintani D, Savage L, Boddupalli S, Ohlrogge J (1997) Targeting of the Arabidopsis homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. Plant Physiology 113: 75-81.

143. Jiang C, Kim SY, Suh D-Y (2008) Divergent evolution of the thiolase superfamily and chalcone synthase family. Molecular Phylogenetics and Evolution 49: 691-701.

144. Blatti JL, Beld J, Behnke C, Mendez M, Mayfield S, Burkart MD (2012) Evolution of fatty acid synthase enzymes dictates interchangeability. Submitted.

145. Dunahay TG, Jarvis EE, Roessler PG (1995) Genetic transformation of the diatoms *Cyclotella cryptica* and *Navicula saprophila*. Journal of Phycology 31: 1004-1012.

146. Zabawinski C, Van den Koornhuyse N, D'Hulst C, Schlichting R, Giersch C, Delrue B, Lacroix JM, Preiss J, Ball S (2001) Starchless mutants of *Chlamydomonas reinhardtii* lack the small subunit of a heterotetrameric ADP-glucose pyrophosphorylase. Journal of Bacteriology 183: 1069-1077.

147. Posewitz MC, Smolinski SL, Kanakagiri S, Melis A, Seibert M, Ghirardi ML (2004) Hydrogen photoproduction is attenuated by disruption of an isoamylase gene in *Chlamydomonas reinhardtii*. Plant Cell 16: 2151-2163.

148. Kosourov SN, Seibert M (2009) Hydrogen photoproduction by nutrient-deprived *Chlamydomonas reinhardtii* cells immobilized within thin alginate films under aerobic and anaerobic conditions. Biotechnology and Bioengineering 102: 50-58.

Chapter 1 is adapted from a review article currently being prepared for publication in Current Opinion in Chemical Biology as part of an upcoming Bioenergy series. Professors Stephen Mayfield and Michael Burkart have contributed words, thoughts, expert ideas and opinions to the article.

Chapter 2 : Manipulating fatty acid biosynthesis in microalgae for biofuel through protein-protein interactions

Abstract

Microalgae are a promising feedstock for renewable fuels, and algal metabolic engineering can lead to crop improvement, thus accelerating the development of commercially viable biodiesel production from algae biomass. We demonstrate that protein-protein interactions between the fatty acid acyl carrier protein (ACP) and thioesterase (TE) govern fatty acid hydrolysis within the algal chloroplast. Using green microalga *Chlamydomonas reinhardtii* (Cr) as a model, a structural simulation of docking CrACP to CrTE identifies a protein-protein recognition surface between the two domains. A virtual screen reveals plant TEs with similar *in silico* binding to CrACP. Employing an activity-based crosslinking probe designed to selectively trap transient protein-protein interactions between the TE and ACP, we demonstrate *in vitro* that CrTE must functionally interact with CrACP to release fatty acids, while TEs of vascular plants show no mechanistic crosslinking to CrACP. This is recapitulated *in vivo*, where overproduction of the endogenous CrTE increased levels of short-chain fatty acids and engineering plant TEs into the *C. reinhardtii* chloroplast did not alter the fatty acid profile. These findings highlight the critical role of protein-protein interactions in manipulating fatty acid biosynthesis for algae biofuel engineering as illuminated by activity-based probes.

Introduction

In our quest to replenish diminishing reserves of fossil fuels with high-energy alternatives while mitigating CO_2 emissions, microalgae have emerged as an attractive option to convert solar energy directly into fungible fuels [1]. However, for microalgal

biofuels to be used at industrial scale, productivity must be enhanced and consistency made tunable [2]. For example, to develop biodiesels to function in existing petroleum-based infrastructure, fatty acids from microalgae must be altered to more closely mimic conventional diesel [3]. Fusion of the powerful tools of systems biology and metabolic engineering could enable us to develop microalgal strains capable of producing commercially viable quantities fatty acids with desired chain lengths [4]. New advances in algal genetic engineering will allow us to fashion viable fuels and commodities from microalgal metabolic pathways [5] [6], yet our knowledge of algal fatty acid biosynthesis remains incomplete. Without a detailed understanding of enzyme activity, timing, and regulation, the engineering of biofuel products from these pathways will struggle to meet our growing energy demands.

Fatty acid biosynthesis has been successfully manipulated in oilseed crops to produce fatty acids with novel compositions [7]. In pioneering work, Voelker and coworkers achieved short-circuiting of fatty acyl chain elongation by expressing a laurate (12:0)-specific thioesterase from the California bay plant (*Umbellularia californica*) in the seeds of *Arabidopsis* and rapeseed *(Brassica napus)* to increase laurate by 24 and 58%, respectively [8]. Since the discovery that heterologous expression of thioesterases can influence the lipid profile of an organism [8], plant TEs have been engineered into a variety of plant species effectively altering their oil content [7].

By terminating fatty acid biosynthesis, the TE functionally determines the length and identity of the fatty acid end product [9]. Plant FatA TEs select for oleoyl (18:1)-ACP substrates and FatB TEs preferentially hydrolyze ACPs loaded with

saturated fatty acids [10]. Some plants have evolved FatB TEs capable of prematurely siphoning short chain fatty acids for incorporation into seed storage oil [11]. Of the range of fatty acids found in Nature, saturated medium chain fatty acids (C_8-C_{14}) are ideal for biodiesel because they have properties that mimic current diesel fuels [4]. Recently, plant FatB TEs were genetically engineered into diatoms (*Phaeodactylum tricornutum*) [12] and cyanobacteria (*Synechocystis* sp. PCC6803) [13] with the goal of creating a superior biodiesel feedstock, but these efforts were met with limited success.

De novo fatty acid biosynthesis occurs within an algal plastid by action of a type II fatty acid synthase (FAS), a modular multi-domain enzymatic complex where each activity is encoded onto a separate protein [14]. Central to FAS, an acyl carrier protein (ACP) acts as a metabolic scaffold, tethering the growing fatty acid as it is shuttled iteratively between catalytic domains of the synthase. Fatty acid biosynthesis begins by post-translational modification of the ACP catalyzed by a phosphopantetheinyl transferase (PPTase), which transfers 4'-phosphopantetheine from coenzyme A to a conserved serine residue on ACP. This converts inactive *apo*-ACP to its active *holo* form bearing a flexible prosthetic arm for attachment of fatty acids via thioester linkage. Once *holo*-ACP is loaded with an acyl starter unit, fatty acid synthesis occurs on the ACP by sequential action of ketosynthase (KS), ketoreductase (KR), dehydratase (DH), and enoyl reductase (ER) enzymes, each cycle resulting in a net addition of two carbons to the growing chain. During chain elongation, ACP buries the growing fatty acid in its hydrophobic core to protect it from hydrolysis [15]. Functional interaction with each FAS domain induces a

conformational change in ACP that draws the acyl chain out of the pocket (‘switchblade’ mechanism) for further processing by downstream enzymes [16]. When a mature fatty acid has been assembled on the ACP, a thioesterase (TE) catalyzes acyl transfer from ACP to its active site cysteine residue, priming the thioester for attack by water. Based on our previous work in bacteria [17] as well as plant studies to manipulate fatty acid chain length [7], we propose that in algal fatty acid biosynthesis, the TE must functionally interact with ACP to receive the acyl chain, and thus significant protein-protein interactions between the ACP and TE are required to mediate the hydrolysis of fatty acids.

We chose to investigate *Chlamydomonas reinhardtii*, as it is one of a small number of algal species for which all three genomes have been sequenced [18]. Most importantly, it is one of only ten algae that can be transformed with current genetic technologies [19], and techniques have been developed that facilitate robust expression of transgenes in both the nuclear and plastid genome of *C. reinhardtii* [20] [21]. Together, these make *C. reinhardtii* an ideal model organism to explore algal fatty acid biosynthesis and decipher ACP-TE interactions. Such dynamic protein-protein interactions within the cell are well known to govern many biological processes [22] [23]. Typical methods to decipher protein-protein interactions include yeast 2-hybrid and TAP-tag systems [24], although these techniques can produce false positive readings. Chemical crosslinking can also be used to identify partner proteins, but this procedure suffers from a lack of specificity [24]. To investigate ACP-TE interactions *in vitro*, we employ a novel approach that circumvents issues associated with traditional methods. Activity-based chemical crosslinking is a powerful means to

capture the transient interactions occurring between FAS domains and ACP [17]. We have used this methodology extensively in prokaryotic systems to study protein-protein interactions, successfully implementing activity-based probes designed for a wide array of activities [25] [26].

Here, we illustrate the use of activity-based probes to investigate ACP-TE interactions towards engineering fatty acid chain length in *C. reinhardtii*. By employing chemical probes specifically designed for the enzymatic activity of TEs, we report the first demonstration of metabolic product control through protein-protein interactions. This work emphasizes the importance of protein-protein recognition between the ACP and catalytic enzymes in mediating fatty acid biosynthesis in algal plastids. We address this hypothesis *in silico*, *in vitro* and *in vivo* using *C. reinhardtii* as a model system. Characterizing algal FAS domains and their interactions will further enable us to optimize heterologous expression of fatty acid biosynthetic enzymes in microalgae and significantly alter the fatty acid profile of transgenic production strains.

Results

Sequence alignment of plant TEs and comparison to CrTE

To critically compare the *C. reinhardtii* TE sequence to activity and its structure to function, we first conducted a multiple sequence alignment between plant FatA and FatB TEs and CrTE (Figure 2.1, Figure 2-6). Structure-based analysis of known TEs and comparison to CrTE places it in thioesterase family TE14 along with FatA and FatB TEs [27]. Sequence alignment with plant TEs shows that CrTE

contains a catalytic cysteine residue as part of a Cys-Asn-His catalytic triad (Figure 2.1). FatB TEs have a conserved hydrophobic 18-amino acid domain [28], while FatA TEs contain distinctive sub-motifs in an N-terminal 60 amino acid transit peptide sequence [29]. Although ChloroP1.1 identifies a 60 amino acid plastid targeting sequence at its N-terminus (Figure 2-6), CrTE is distinct from FatA and FatB TEs in this region, which has been proposed as the source for substrate specificity in TEs [9]. BLAST analysis indicates that CrTE shares the closest homology to predicted TEs from *Volvox carteri* and *Chlorella variabilis*, a multicellular chlorophyte alga and a unicellular green alga, respectively, both closely related to *C. reinhardtii*.

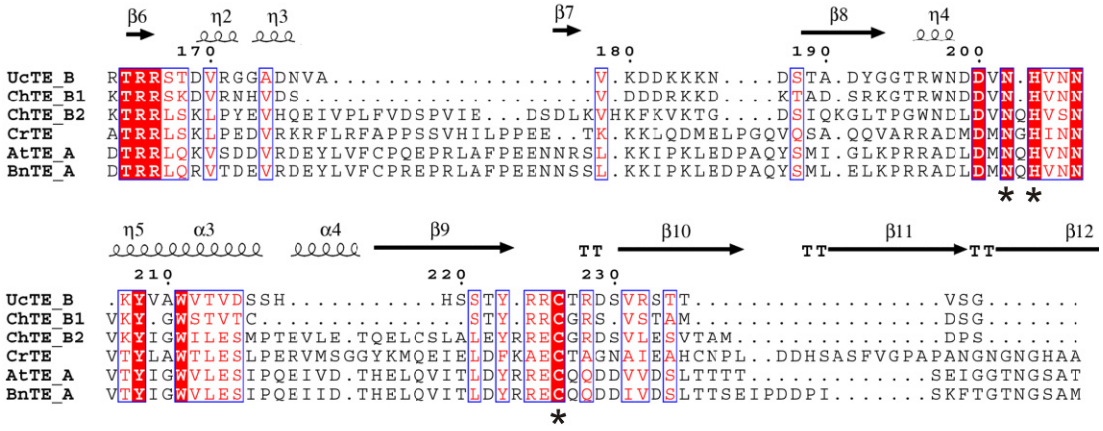


Figure 2-1 Thioesterase sequence alignments

Structure-based sequence alignment of FatA TEs from *Arabidopsis thaliana* (AtTE_A) and *Brassica napus* (BnTE_A), *Chlamydomonas reinhardtii* (CrTE), and FatB TEs from *Cuphea hookeriana* (ChTE_B1 and ChTE_B2) and *Umbellularia californica* (UcTE_B). The Cys-His-Asn catalytic triad is shown by an asterisk (*). Conserved residues are highlighted in red and similar residues in blue boxes.

An exhaustive search of medium chain-specific plant TEs for homology to the CrTE resulted in a panel of fourteen plant TEs (*SI*). Because there are currently no

crystal structures available for plant or algal TEs, homology models were constructed, and TEs were screened computationally for functional binding to CrACP.

In silico modeling of ACP-TE protein-protein interactions and TE-substrate interactions

While there are no plant or algal TE structures deposited in the PDB, two crystal structures (PDB: 2OWN and 2ESS) display homology and high sequence identity (37.9% and 41.4%, respectively) to CrTE around the active site residues Cys-His-Asn (8 Å). The modeled CrTE obtained from two different homology algorithms are nearly superimposable and in line with previous modeled TE structures including those from *Arabidopsis* [30] and *Jatropha* [29]. To investigate the binding mode of a fatty acid by CrTE and its substrate specificity, the modeled CrTE was subjected to non-biased docking of a stearate-pantetheine substrate (Figure 2-7). Consistently, the calculated top 10 models display the model with pantetheine substrates bound to CrTE in its binding tunnel (Figure 2-2) with the thioester exposed to the tentative active site. Experimentally observed fatty acyl substrates (14:0, 16:0, 16:4, 18:0, 18:1 and 18:3) were docked onto the model structure of CrTE and all fatty acids dock to the binding tunnel of CrTE (Figure 2-7). However, the 16:4 and 18:3 substrates do not bind in a favorable orientation, as expected since desaturations are typically derived from post-synthase desaturase enzymes.

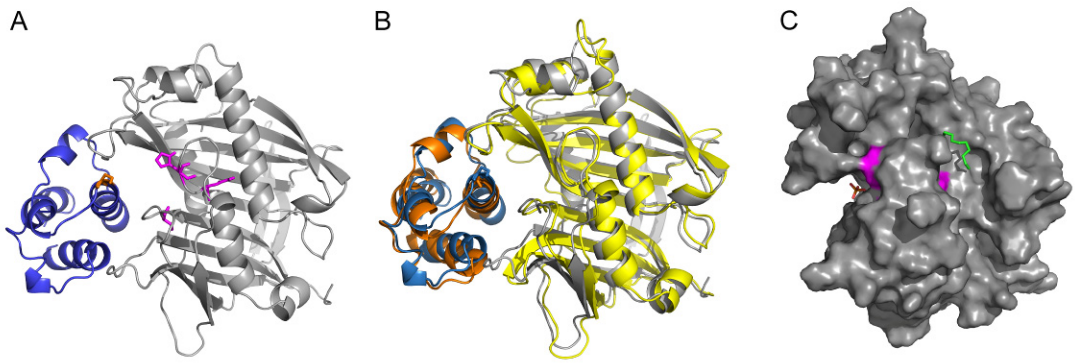


Figure 2-2 Thioesterase modeling, docking of ACP-TE protein-protein interactions, and blind substrate docking of fatty acid substrate to CrTE

A) Docking of CrTE (grey) with Cr-cACP (blue) showing a <10 Å distance between Cr-cACP Ser43 (orange) and the active site Cys306His270Asn268 triad (magenta) of CrTE. B) Docked complex of CrTE (grey) and ChTE (yellow) showing similar binding modes of Cr-cACP to both plant and algal thioesterases. C) Surface representation of blind docking of stearyl-4'-phosphopantetheine to CrTE showing the thioester bond of the substrate in close proximity to the TE active site and stearate in the binding tunnel of CrTE.

C. reinhardtii chloroplastic ACP (Cr-cACP) was modeled and docked onto TEs using the Cluspro server (Figure 2-2). We generated the homology model of Cr-cACP based on the template of the ACP from *Aquifex aeolicus* (PDB: 2EHTA), which shows a sequence identity of 52.6%, a q-meanscore of 0.7, and a residual RMSD of 0.08 between template and model. Models were compared with those obtained from the i-Tasser server [31] and no significant deviations between the two homology modeling algorithms were observed. The models generated from protein-protein docking were examined for binding of *apo*-ACP close to the TE active site and for juxtaposition of the ACP serine in relation to the TE active site cysteine. Docking of Cr-cACP to CrTE displayed large surface interactions between helix II of Cr-cACP

and the CrTE as well as a productive orientation of the ACP serine to TE cysteine (Figure 2-2), consistent with previous studies of protein-protein interactions between ACP and partner FAS enzymes [32] [33]. Examining electrostatic potential maps of Cr-cACP/CrTE models shows electrostatic complementarity, and docking predicts a 13.5 Å distance between Cr-cACP conserved serine and CrTE active site cysteine residue (Figure 2-10).

Virtually screening a panel of fourteen vascular plant TEs for interaction with Cr-cACP using a high-end computational cluster collecting data over several months [34] for each TE, *U. californica* TE (UcTE) and *C. hookeriana* TE (ChTE) emerged as promising candidates based on their ability to dock in proximity to the Cr-cACP Ser43 (Figure 2-2, Figure 2-8). Careful examination of Cr-cACP with UcTE and ChTE revealed that these ACP-TE models closely resemble the Cr-cACP/CrTE model, indicating a possible productive interaction between Cr-cACP and plant TEs. In contrast, modeling *C. reinhardtii* mitochondrial ACP (Cr-mACP) with these same plant TEs shows poor docking, suggesting that not all algal ACPs will successfully bind, and thus not catalytically interact, with plant TEs (Figure 2-9, Figure 2-10). Although the electrostatic potential map of Cr-mACP is similar to Cr-cACP, docking Cr-mACP to CrTE illustrates a different binding motif than that of Cr-cACP/CrTE, and as a result, the distance between Cr-mACP serine and CrTE cysteine is predicted to be 18.8 Å (Figure 2-10).

Characterization of fatty acid acyl carrier proteins of C. reinhardtii

To study ACP-TE interactions *in vitro*, we obtained both fatty acid ACPs identified in the *C. reinhardtii* genome, ACP1 and ACP2. Based upon alignment with plant mitochondrial and chloroplastic ACPs, homology, and BLAST analysis, ACP1 and ACP2 correspond to the mitochondrial and chloroplastic ACPs, respectively (Figure 2-11, *SI*). ACPs were synthesized in *E. coli* codon bias and cloned into pET28b expression vectors bearing an N-terminal-His_6 tag. Whereas Cr-mACP (ACP1) formed insoluble aggregates when expressed in *E. coli* and required refolding [35], Cr-cACP (ACP2) was expressed in *E. coli* in soluble form. Because we anticipated a stronger interaction between the ACP of the chloroplast and the CrTE, Cr-cACP was used as a positive control in crosslinking experiments (for details about Cr-mACP, see *SI*). After Ni-NTA purification, a mixture of *apo*-Cr-cACP and *holo*-Cr-cACP (Figure 2-3) was obtained. The conversion of *holo*-Cr-cACP to *apo*-Cr-cACP was catalyzed by ACP-hydrolase (ACPH) from *Pseudomonas aeruginosa* [36], which removed the 4'-phosphopantetheine moiety from Ser43 of Cr-cACP (Figure 2-3, Figure 2-12). The formation of *apo*-Cr-cACP was confirmed by mass spectrometry (Figure 2-13). A one-pot chemoenzymatic method was used to validate *in vitro* post-translational modification of *apo*-Cr-cACP [37]. As shown in Figure 2-3, incubation of *apo*-Cr-cACP with Sfp, a surfactin synthetase-activating PPTase from *B. subtilis* [38], Coenzyme A biosynthetic enzymes (CoA-A, CoA-D, and CoA-E), ATP, and fluorescent pantetheine analogue **1** resulted in the formation of fluorescently labeled *crypto*-Cr-cACP. This conversion validated proper folding and activity of Cr-cACP

and confirmed our ability to successfully load the activity-based probe onto ACP for further studies of protein-protein interactions [39].

Expression, purification, and activity of Cr, Uc, and Ch thioesterases

CrTE, UcTE, and ChTE genes were synthesized in *E. coli* codon bias and subcloned into pET21a expression vectors bearing a C-terminal FLAG tag. Expression in *E. coli* and FLAG affinity chromatography yielded pure TE protein. After reduction, plant and algal thioesterase activity was confirmed kinetically by monitoring hydrolysis of para-nitrophenylhexanoate (Figure 2-14) [40].

In vitro crosslinking assay to screen TEs for functional binding to Cr-cACP and design of TE substrate mimic

Recently, we developed a crosslinking assay based on enzymatic activity that interrogates specific protein complex formation in FAS, PKS, and NRPS modular synthases using pantetheine probes [26]. Work in our laboratory has demonstrated that activity-based crosslinking serves as a quantitative measure of the strength of functional binding interactions between ACP and modular FAS domains [41]. In this assay, selectivity arises from protein-protein interactions, rather than purely protein-substrate specificity, as shown by examining biochemically unrelated carrier proteins modified with the same alkylating functionality [26]. This method is far more selective than traditional techniques aimed at interpreting protein-protein interactions, thus preventing false positive readings from non-specific protein complex formation.

To investigate ACP-TE interactions, we sought to design a chemical probe that would selectively react with the active site of the TE. Chloroacrylic pantetheine probe

2 was used in previous studies to generate a site-specific crosslink between *E. coli* ACP and *E. coli* KS, triggered by protein-protein recognition [26]. Based on sequence alignment (Figure 2.1), the CrTE uses an active site cysteine as its catalytic residue during fatty acid biosynthesis, similar to the KS domain. However, the TE is fundamentally different from the KS, as it only acts upon the final acyl-ACP substrate, whereas the KS acts on intermediate chain lengths. When a mature fatty acid is assembled on the ACP, the TE first catalyzes acyl transfer to its active site cysteine residue (Figure 2-3). The product is then hydrolyzed as water attacks the pendant thioester to yield a free fatty acid. Modeling studies of CrTE illustrate a shallow hydrophobic tunnel at the surface of the enzyme, where the acyl chain presumably lies (Figure 2-2). Previously, a panel of pantetheine probes was used to examine the compatibility of ACP-KS partners, which established that substrate recognition also plays a critical role in catalysis [25]. Hypothesizing that the fatty acyl chain factors into enzymatic hydrolysis by the TE, we designed an activity-based substrate mimic containing a reactive bromide α to the site of nucleophilic attack by the TE and a long aliphatic tail (Figure 2-3). To test the substrate specificity and selectivity of the CrTE using activity-based probes, we designed a shorter α-bromohexyl pantetheine analogue composed of only six carbons (**6**). Chemical synthesis of α-bromopalmitic pantetheine probe **3** and its shorter analogue **6** (*SI*) enabled us to study ACP-TE interactions using three different activity-based probes (**2, 3,** and **6**).

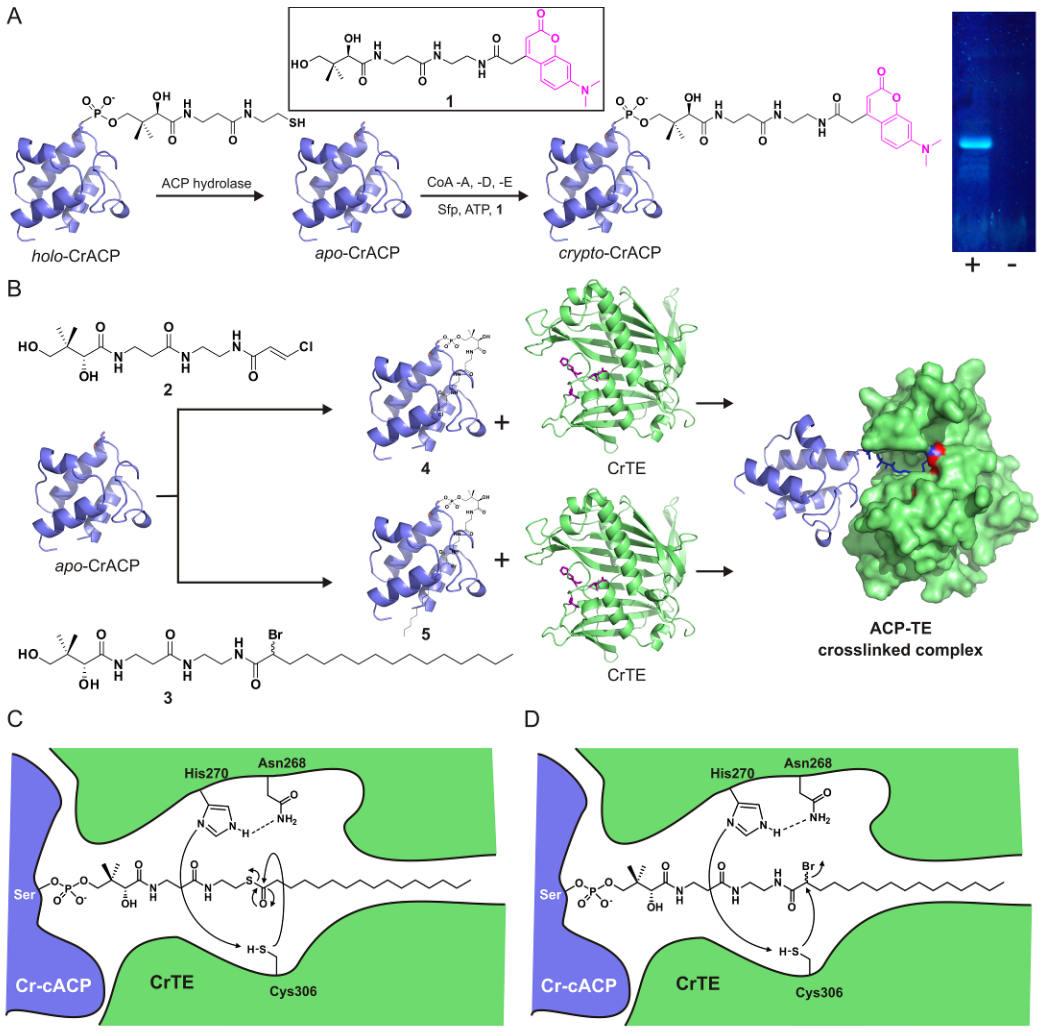


Figure 2-3 Schematic of activity-based crosslinking between CrACP and TEs

A) *Apo*-CrACP is formed by treating *holo*-CrACP with ACP hydrolase from *P. aeruginosa* [36], removing the pantetheine moiety from the conserved serine of CrACP. Presence of apo-CrACP is confirmed using a one-pot fluorescent labeling method [37], detected by visualization of a resulting SDS-PAGE gel at 365 nm. (+) formation of fluorescent crypto-CrACP; (-) control reaction in which fluorescent pantetheine analogue **1** was omitted. B) Activity-based crosslinking scheme. *Apo*-CrACP is incubated with **2** or **3**, Sfp, ATP, CoA-A, CoA-D, and CoA-E to form the corresponding *crypto*-CrACPs **4** and **5**. When *crypto*-CrACP is incubated with TE, protein-protein interactions trigger a site-specific covalent crosslinking reaction with the chloroacrylamide in **4** or the α-bromoamide moiety in **5**, forming an ACP-TE crosslinked complex. C) Enzymatic mechanism of the hydrolytic release of a fatty acid by CrTE using a Cys-Asn-His catalytic triad. D) Mechanism of irreversible crosslink between TE and crypto-CrACP containing a reactive bromide on the carbon α to the site of nucleophilic attack by the TE.

Manipulating the CoA biosynthetic pathway [39], functionalized CoA analogues were formed *in vitro* from chloroacrylamide **2** or α-bromopalmitamide **3** (Figure 2-3). Sfp [38] was used to attach substrate mimics to Ser43 of *apo*-Cr-cACP, generating reactive *crypto*-Cr-cACPs **4** and **5** (Figure 2-3) [37].

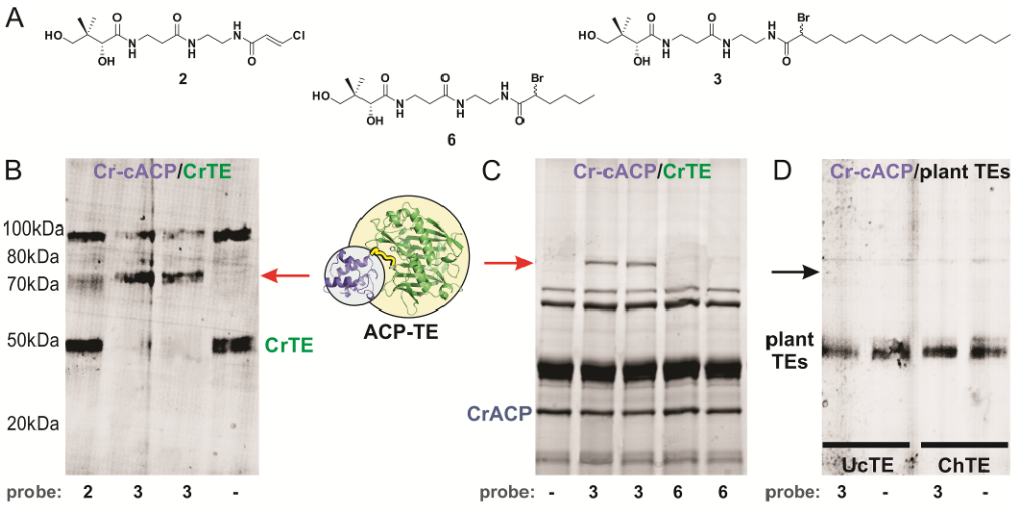


Figure 2-4 Activity-based crosslinking as a determinant of functional interaction with the *C. reinhardtii* cACP

A) Activity-based substrate mimics used in crosslinking assay; B) SDS-PAGE gel showing Cr-cACP/CrTE interaction. *Apo*-Cr-cACP was modified with pantetheine analogue **2** or **3** to generate the corresponding *crypto*-Cr-cACPs (**4** and **5**). *Crypto*-Cr-cACPs were incubated with CrTE and crosslinking was visualized by SDS-PAGE analysis following FLAG affinity purification of the CrACP/TE complex. The band observed at 50 kDa is the FLAG-tagged CrTE and the band detected at ~75 kDa is the ACP/TE complex. During overnight incubation at 37 °C, reduced CrTE shows spontaneous oxidation (bands at ~100 kDa). Pantetheine analogues used in crosslinking reactions are noted under the gel. A red arrow illustrates the ACP-TE complex in (B) and (C). C) *Apo*-Cr-cACP was reacted with either α-bromopalmitic pantetheine probe **3** or α-bromohexyl pantetheine probe **6** to generate *crypto*-Cr-cACPs with C16 and C6 acyl chains attached, respectively. Each *crypto*- Cr-cACP was incubated with CrTE and monitored for extent of crosslinking by SDS-PAGE analysis of Ni-NTA-purified reactions. D) *Apo*-Cr-cACP was modified with α-bromopalmitic pantetheine analogue **3** to form *crypto*-Cr-cACP, which was incubated with UcTE (left 2 lanes) or ChTE (right 2 lanes). Crosslinking was measured by SDS-PAGE analysis following FLAG affinity purification. The bands at ~50 kDa are plant TEs.

A crosslinked complex was observed between *crypto*-Cr-cACP and CrTE using both probes, indicating a positive interaction between the Cr-cACP and CrTE

(Figure 2-4). Importantly, incubation of α-bromopalmitic *crypto*-Cr-cACP **5** with CrTE resulted in full conversion of CrTE to a Cr-cACP/CrTE complex, indicating a strong functional interaction between the two proteins (Figure 2-4). In contrast, reacting chloroacrylic *crypto*-Cr-cACP **4** with CrTE did not result in a high yield of ACP/TE protein complex formation (Figure 2-4).

To study the effect of the length of the acyl chain attached to Cr-cACP on functional binding to CrTE, a shorter α-bromohexyl pantetheine probe **6** (*SI*) was used for activity-based crosslinking. In agreement with its predicted natural activity, a higher quantitative yield of crosslinked complex was detected between CrTE and longer chain α-bromopalmitic-ACP (Figure 2-4), indicating that chain length factors into enzymatic hydrolysis by the TE.

Based on the results of our initial computational docking screen of plant TEs, UcTE and ChTE were tested for their ability to crosslink with *crypto*-Cr-cACP. Contrary to modeling studies, no crosslink was detected between heterologous plant TEs and Cr-cACP due to an absence of protein-protein recognition (Figure 2-4). *Crypto*-Cr-cACPs formed from either chloroacrylic pantetheine probe **2**, α-bromopalmitic pantetheine probe **3**, or α-bromohexyl pantetheine probe **6** did not result in a crosslinked complex, indicating vascular plant TEs did not functionally interact with Cr-cACP *in vitro* (Figure 2-15).

Interestingly, *C. reinhardtii* mitochondrial ACP (Cr-mACP) did not orient favorably when docked onto CrTE, UcTE or ChTE (Figure 2-9, Figure 2-10). Indeed, a crosslinked complex was not detected between Cr-mACP and the CrTE of the chloroplast, again demonstrating selective recognition between ACP and TE domains

(Figure 2-16). These results demonstrate the specificity of protein-protein recognition between the algal chloroplastic ACP and TE.

Engineering TEs into Cr chloroplast and fatty acid analysis of transgenic strains

To determine whether plant TEs functionally interact with the Cr-cACP *in vivo* and to validate our *in vitro* activity-based crosslinking results, CrTE, UcTE, and ChTE were engineered into the *C. reinhardtii* chloroplast. A high level of protein expression was achieved by codon-optimizing the plant TE sequences for the *C. reinhardtii* chloroplast and using endogenous promoters and RNA elements to drive expression [21]. Microprojectile bombardment transformed *C. reinhardtii* strain 137c (mt+) with thioesterase-harboring plastid expression vectors [21]. PCR confirmed integration of the TE genes and homoplasmicity (Figure 2-17) and Western blot analysis validated expression of the TEs in the *C. reinhardtii* chloroplast (Figure 2-18) [42]. After purification from Cr lysate and reduction, plant and algal thioesterase activity was validated kinetically by monitoring hydrolysis of para-nitrophenylhexanoate (Figure 2-14) [40].

Transgenic algal strains were grown in a greenhouse to late log phase, harvested, and fatty acids converted into their corresponding methyl esters for GC/MS analysis (*SI*). Compared to wildtype, no significant change in fatty acid content or composition was detected in transgenic *C. reinhardtii* strains expressing the UcTE or ChTE (Figure 2-5). This was an indication that plant TEs did not functionally interact with Cr-cACP *in vivo* and thus were unable to hydrolyze fatty acids attached to Cr-cACP. Interestingly, the strain overexpressing native CrTE resulted in a short

circuiting of fatty acid chain elongation to increase myristic acid (14:0) content by 2.5 fold (p-value <0.05) (Figure 2-5). These results signify the first time protein-protein interactions have been correlated with product control in algae.

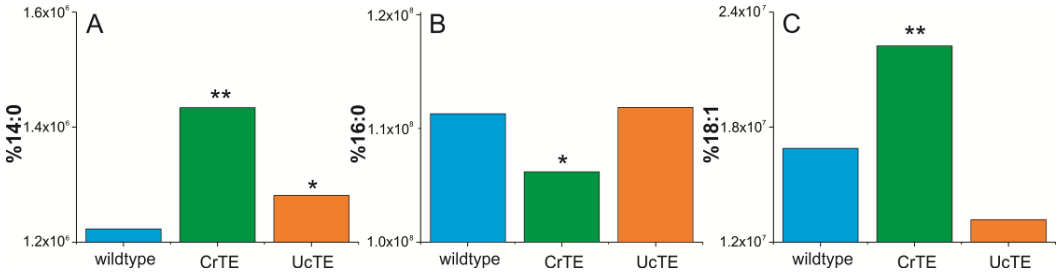


Figure 2-5 Fatty acid analysis of *C. reinhardtii* strains expressing thioesterases

Fatty acid composition of *C. reinhardtii* strains was determined by GC/MS analysis and comparison to authentic standards. Peak areas were integrated and compared to an external standard for quantification. Bar graphs denote abundances of (A) Myristic acid (14:0), (B) Palmitic acid (16:0), and (C) Oleic acid (18:1). Three separate cultures of each strain were analyzed for fatty acid content and composition, and data were recorded and averaged with a mean deviation of 7% in each experiment. Statistical analyses were performed using SPSS (v13.0), and for all data analysis, a p-value <0.5 was considered statistically significant.

Discussion

Fatty acid biosynthesis is orchestrated by a modular enzymatic complex within the algal plastid, similar to machinery present in prokaryotes [43]. The type II FAS iteratively condenses 2-carbon acetate units, each cycle resulting in a fully reduced extended carbon chain. All reactions occur with substrates bound as thioesters to a 4'-phosphopantetheine arm attached to ACP. In bacteria, interactions between the ACP and KS are required for proper metabolite assembly [26]. Further studies have established that substrate recognition is also important in KS-mediated condensation [25]. Here, we have shown that significant ACP-TE interactions as well as enzyme-

substrate interactions play a role in the hydrolysis of fatty acids within an algal chloroplast.

Docking of Cr-cACP to CrTE identifies a protein-protein recognition surface as well as proximity between the fatty-acyl thioester attached to the Cr-cACP and active site cysteine of CrTE (Figure 2-2), not dissimilar to ACP/TE docking in *Jatropha curcas* [29]. To trap transient ACP-TE interactions *in vitro*, we designed, synthesized, and implemented chemical probes inspired by the enzymatic activity of the TE (Figure 2-3). Activity-based probe **3** was attached to Cr-cACP to form *crypto*-Cr-cACP **4** (Figure 2-3) [35], and indeed, a positive functional interaction between CrTE and Cr-cACP triggered a site-specific crosslinking reaction, effectively trapping the ACP-TE complex (Figure 2-4). In fact, a crosslinked ACP-TE complex was observed between *crypto*-Cr-cACP and CrTE using both chloroacrylic pantetheine probe **2** and longer chain α-bromopalmitic pantetheine probe **3** (Figure 2-3), demonstrating a positive functional ACP-TE interaction (Figure 2-4). Incubation with longer chain α-bromopalmitic pantetheine probe **3** resulted in full conversion to an ACP-TE complex, whereas chloroacrylic pantetheine probe **2** left unreacted CrTE, demonstrating the specificity of the CrTE probe (Figure 2-4). We do not have any evidence that the racemic nature of the synthesized α-bromopalmitic pantetheine molecule (**3**) or its 6-carbon analogue (**6**) affects the crosslinking reaction, similar to previous studies [26], and there are ongoing efforts in our laboratory to study this further.

Although UcTE and ChTE appeared to orient in a similar conformation to Cr-cACP as wildtype CrTE in computational docking studies, no crosslinked complex

was detected using all three activity-based probes (**2**, **3**, and **6**), indicating that vascular plant TEs did not functionally interact with Cr-cACP *in vitro* (Figure 2-4, Figure 2-15, S10). To validate the results of our crosslinking assay *in vivo*, UcTE and ChTE were engineered into the *C. reinhardtii* chloroplast (Figure 2-17, Figure 2-18). Although expressions levels were high (Figure 2-18) and enzymatic activity was confirmed (Figure 2-14), no significant change in fatty acid content or composition was detected in transgenic Cr strains expressing vascular plant TEs, confirming that UcTE and ChTE were unable to hydrolyze fatty acids attached to Cr-cACP due to a lack of molecular recognition.

To clearly demonstrate the specificity of ACP-TE interactions, the Cr mitochondrial ACP (Cr-mACP) did not display favorable binding to CrTE *in silico* (Figure 2-9), as the model of docked Cr-mACP/CrTE predicted a greater distance between the ACP serine and TE cysteine than in the Cr-cACP/CrTE model (Figure 2-10). This result was validated *in vitro*, as Cr-mACP did not form a crosslinked complex with CrTE (Figure 2-16).

The longer chain α-bromo pantetheinyl-ACP served as a much better substrate mimic than chloroacrylic pantetheinyl-ACP (Figure 2-4). Similar selectivity has been demonstrated with polyketide KS domains that accept elongated hydrocarbon products, indicative of enhanced substrate binding [25]. Since the TE only acts upon a mature fatty-acyl-ACP, we propose that in addition to protein-protein interactions, the acyl chain attached to the ACP is important in guiding the thioester moiety towards the TE active site cysteine residue. Our modeling studies reveal a shallow hydrophobic binding cleft close to the CrTE active site (Figure 2-2). Perhaps there is a

synergistic effect at work, whereby the acyl chain helps orient the fatty acid substrate in the correct conformation for hydrolysis to occur.

To test whether the length of the acyl chain attached to the Cr-cACP factors into enzymatic hydrolysis by the CrTE, we synthesized a shorter chain derivative of **3** composed of only six carbons for activity-based crosslinking. Whereas C16 α-bromopalmitic pantetheine probe **3** resulted in full conversion to an ACP-TE species, its six-carbon analogue **6** yielded significantly less Cr-cACP/CrTE complex formation (Figure 2-4). These results indicate that in addition to ACP-TE interactions, chain length plays a role in the hydrolytic process.

To further characterize the CrTE *in vivo* and test its specificity, transgenic *C. reinhardtii* overexpressing native CrTE was analyzed for fatty acid content and composition. Wildtype *C. reinhardtii* produces mostly 16:0 fatty acids and only negligible amounts of 14:0 and 18:1 [44]. The shift in fatty acid profile to increased levels of 14:0 by overexpressing CrTE is notable in that CrTE does not display high homology to plant FatB TEs, which have varied affinities for saturated fatty acyl-ACPs [10]. In fact, the CrTE is more homologous to FatA TEs, which explains the slight increase in 18:1 (Figure 2-5). The availability of acyl-ACP substrates for TE hydrolysis determines the fatty acid content observed. For example, when *Arabidopsis thaliana* FatB TE was expressed in *E. coli,* an organism that produces mostly 16:0, there was a significant increase in 14:0 [45]. Therefore, if CrTE has specificity for saturated acyl-ACPs, its overexpression most likely causes a stoichiometric imbalance of CrTE to Cr-cACP in the plastid, which results in premature hydrolysis of available acyl-ACPs and accumulation of shorter chain fatty

acids. The observation that the pool of 16:0-ACPs was diminished in this strain compared to wildtype supports this hypothesis (Figure 2-5).

In plants, it is postulated that medium chain FatB TEs and FatA TEs are descendents of an ancient, ubiquitous FatB TE specific for 16:0-ACP [10]. A dedicated 16:0-acyl-ACP TE has not been detected in *C. reinhardtii*, but activity-based crosslinking studies with the 16:0-acyl-ACP probe indicate that CrTE can hydrolyze 16:0-ACP substrates. Crosslinking with the short-chain derivative (**6**) shows that CrTE is less selective for shorter chain acyl-ACPs, but it is still able to process them. *In vivo* characterization illustrates that CrTE can act upon 14:0-ACPs and 16:0-ACPs, as well as 18:1-ACPs, suggesting promiscuity in hydrolytic activity by CrTE. Docking fatty acid substrates of varied chain lengths and degrees of unsaturation provides further evidence that CrTE can hydrolyze both saturated and 18:1 fatty acids (Figure 2-8). We propose that CrTE is a unique TE with both FatA and FatB character.

Algal TEs remain largely uncharacterized, with the exception of a putative thioesterase (PtTE) from the diatom *Phaeodactylum tricornutum* [46]. PtTE is distinct from FatA and FatB TEs as well as prokaryotic TEs. However, overexpression of PtTE in *E. coli* increased the amount of mono-unsaturated fatty acids and total fatty acid content, similar to FatA TEs [46]. Upon sequencing of the *C. reinhardtii* genome [18], CrTE was initially annotated as a FatA TE based on homology to plant TEs. Yet, sequence alignment shows only dispersed regions of similarity with FatA TEs (Figure 2.1, Figure 2-6). Currently in the UniProt database, CrTE is designated a ‘Fat1’ TE [47]. Phylogenetic analysis reveals that CrTE exists in a separate clade from FatA

and FatB TEs along with other algal TEs (Figure 2-19). Perhaps the CrTE sits at the crossroads of thioesterase evolution, where it functioned as one of the first TEs to hydrolyze unsaturated fatty acids. This is the first evidence of a unique Fat1 TE with promiscuity and possibly high hydrolytic activity for saturated acyl-ACP substrates in algae.

Protein-protein interactions between the TE and ACP control the identity of the fatty acid end product, as illuminated by activity-based crosslinking. Here, we demonstrate for the first time that activity-based chemical probes can be used to trap FAS enzymes and study their interactions with endogenous biosynthetic machinery in algal chloroplasts. We have verified that *in vitro* activity-based crosslinking translates to expected phenotypes *in vivo*, as shown by engineering TEs into the Cr chloroplast and analyzing fatty acid content. Characterization of algal fatty acid biosynthetic enzymes has unveiled distinct features compared to established bacterial and plant type II FAS systems, exemplified by a unique TE detected in *C. reinhardtii*. Our results demonstrate that synergy between protein-protein interactions and substrate recognition mediates the hydrolysis of fatty acids by CrTE.

This work offers a new fundamental understanding of algal fatty acid synthase with which to guide future metabolic engineering. It is the first demonstration of protein-protein interactions used as a mechanism to control product identity. To accomplish this, we successfully designed and implemented activity-based probes as a means to understand the molecular recognition between the algal ACP and algal/plant TEs. In addition, we interrogated the specificity of the TE using three different

activity-based probes. This novel approach may be used to facilitate the engineering of fatty acid identity in algae and other organisms of interest for energy production.

Materials and methods

Protein modeling and docking

Proteins were modeled using Swissmodel [48] and docked using the Cluspro server [34]. The top 10 balanced models were manually examined using PyMol [49] and the distances between TE active site cysteine and ACP serine residues measured. Autodock was used to blindly dock fatty acid substrates onto the model of CrTE [50] (Figure 2-7, Figure 2-8, Figure 2-9).

Sequence alignment

Sequence-based alignments of ACPs and TEs were produced using TCoffee [51], and figures were created using ESPript [52]. Transit peptides were predicted using ChloroP1.1 [53] and TargetP1.1 [54]. Full TE and ACP sequence alignments are provided in Figure 2-6 and Figure 2-11.

Protein expression and purification

Cr-mACP, Cr-cACP, CrTE, UcTE, and ChTE proteins were expressed in *E. coli* and purified according to standard protocols and TEs reduced prior to use. ACP hydrolase (ACPH) was used to form *apo*-Cr-cACP (Figure 2-3 and Figure 2-12), confirmed by MS analysis (Figure 2-13). Thioesterase activity was confirmed kinetically by monitoring hydrolysis of para-nitrophenylhexanoate (Figure 2-14) [40].

Activity-based crosslinking as a determinant of functional binding

Crosslinking reactions contained phosphate buffer, pantetheine analogue (**2**, **3** or **6** in DMSO), $MgCl_2$, ATP, ACP, CoA-A, CoA-D, CoA-E and Sfp [26]. After 1h incubation at 37 °C, TE was added and the reaction mixture further incubated. As a negative control, enzymes were incubated without pantetheine analogue. Following overnight incubation, anti-FLAG resin or Ni-NTA resin was added directly to the reaction mixtures, incubated at room temperature, centrifuged, washed and proteins eluted with 1 M arginine (pH 3.5) or 100 mM imidazole.

Engineering TEs into C. reinhardtii chloroplast

Genes encoding UcTE, ChTE, and CrTE were codon-optimized for the Cr chloroplast and cloned into chloroplast expression vectors (16). Microprojectile bombardment was used to transform *C. reinhardtii* with TE constructs [55]. Integration of TE genes into the chloroplast genome was confirmed by PCR (Figure 2-17) and expression of Tes was validated by Western blot analysis (Figure 2-18) [42]. Plant and algal TE activity was measured kinetically by monitoring hydrolysis of para-nitrophenylhexanoate (Figure 2-14) [40].

Growth of transgenic C. reinhardtii strains and GC/MS analysis of fatty acid content

C. reinhardtii strains were grown on TAP media [56] agar plates supplemented with carbendazim, ampicillin and cefotaxime [57] under constant illumination. Transgenic strains were kept under kanamycin selection. Flasks containing TAP media were inoculated with single colonies and grown under constant shaking in a

greenhouse. After 3 days, cultures were centrifuged and the cell pellet resuspended in methanolic acid. The suspension was incubated for 30 min at 65 °C and fatty acid methyl esters extracted using hexanes. Fatty acid composition was determined by GC/MS analysis. Statistical analyses were performed using SPSS (v13.0), and for all data analysis, a p-value <0.5 was considered statistically significant.

.

Supplemental Information

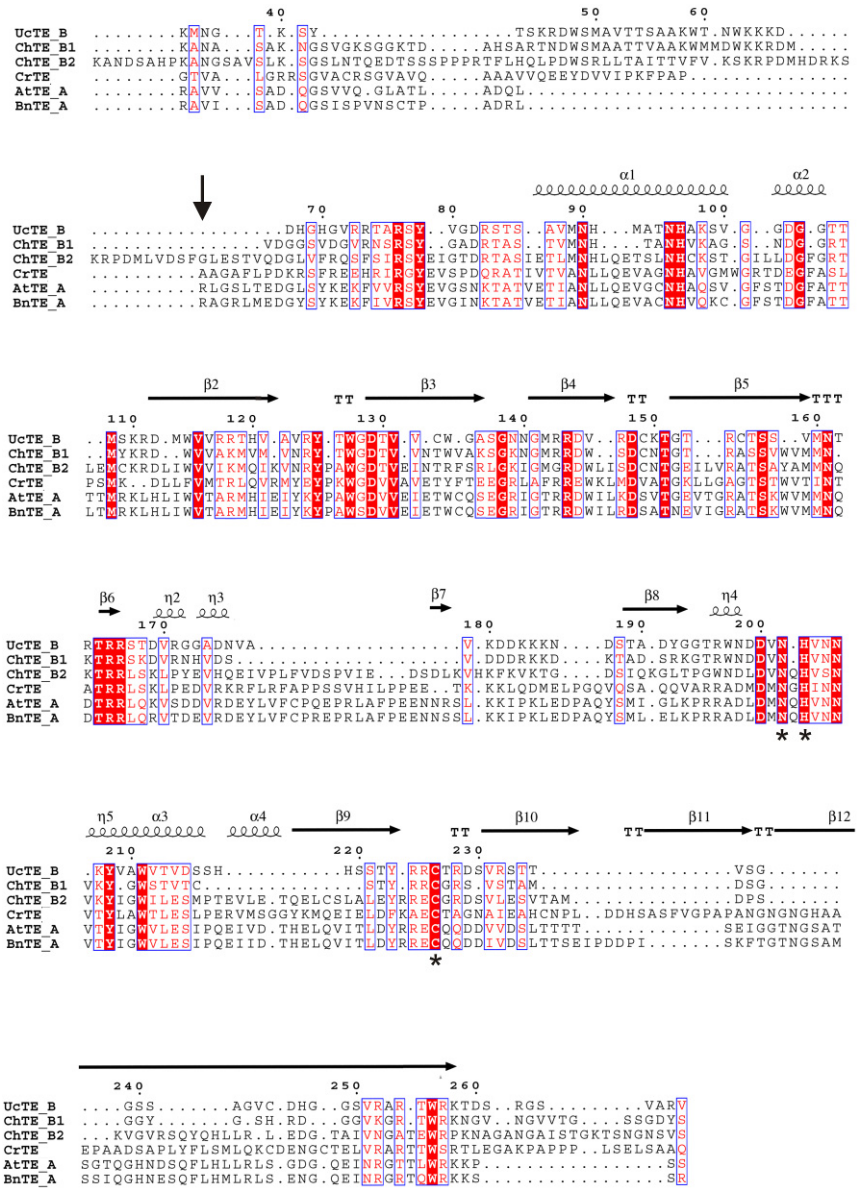


Figure 2-6 Sequence alignment of FatA and FatB TEs with CrTE

Structure-based sequence alignment of FatA TEs from *Arabidopsis thaliana* (AtTE_A), *Brassica napus* (BnTE_A), *Chlamydomonas reinhardtii* TE (CrTE) and FatB Tes from *Cuphea hookeriana* (ChTE_B1) and (ChTE_B2) and *Umbellularia californica* (UcTE_B). Conserved residues are highlighted in red and similar residues appear in blue boxes. A black arrow indicates the end of each transit peptide sequence. Cys-Asn-His catalytic triad is indicated by an asterix under each residue. Symbols are denoted as: α, α-helices; β, β-strands; η, 310 helices; TTT, strict α-turn; TT, strict β-turn. Sequence-based alignments were produced using Tcoffee [51]. The figure was created using ESPript [52].

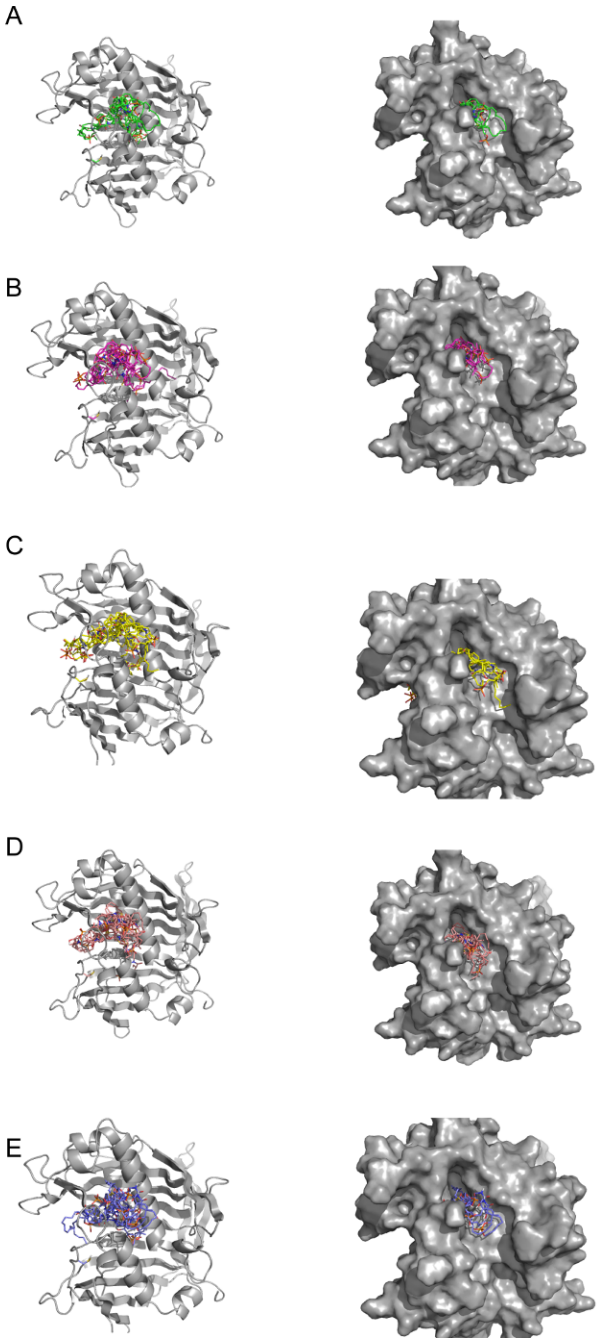


Figure 2-7 Docking of fatty acid PPTs to CrTE

Autodock4 was used to blindly dock various fatty acid-PPTs to CrTE. A) C16:0-PPT, B) C16:4-PPT, C) C18:0-PPT, D) C18:1-PPT and E) C18:3-PPT. Left: cartoon representation of CrTE (grey) and top 10 models of substrate as sticks. Right: surface representation of CrTE (grey) and top 10 models of substrate as sticks [PPT: phosphopantetheine]

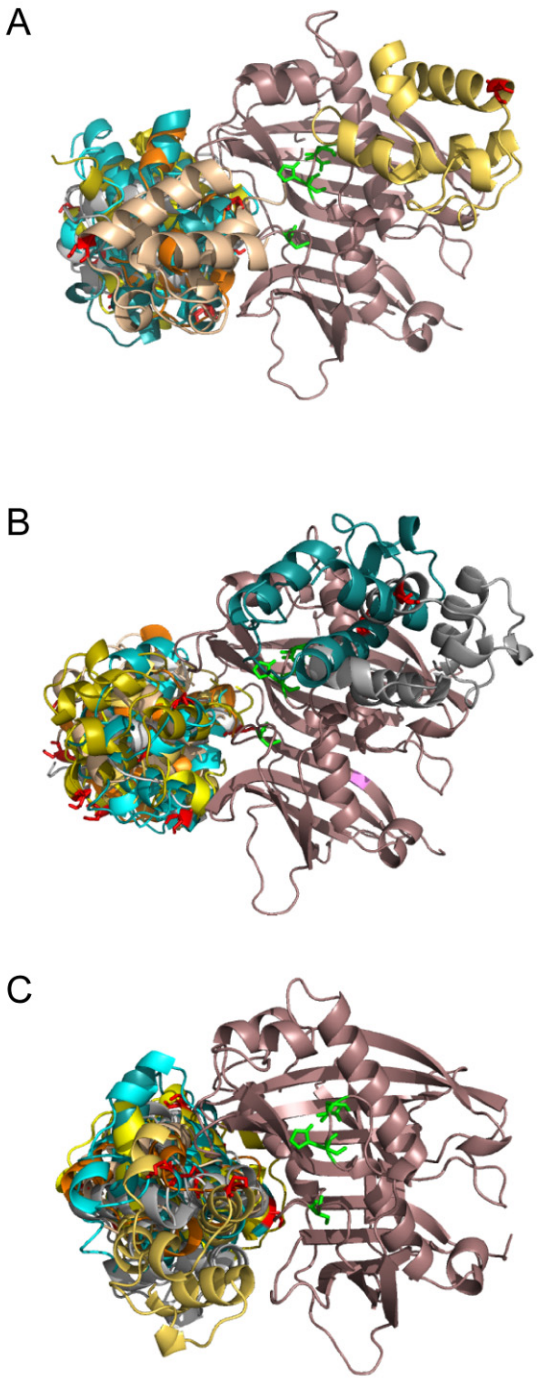


Figure 2-8 Ensemble representation of protein-protein docking of Cr-cACP with TEs

A) UcTE, B) ChTE and C) CrTE

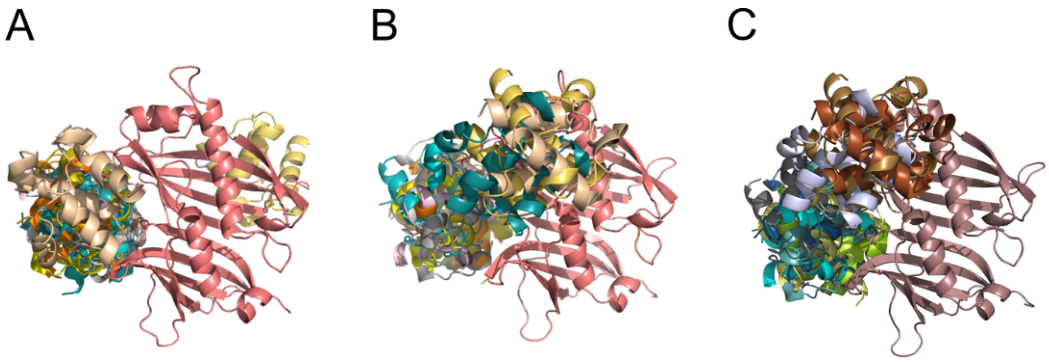


Figure 2-9 Docking of Cr-mACP to CrTE, UcTE, and ChTE

Docking of *C. reinhardtii* mitochondrial ACP (Cr-mACP) to thioesterases showing less favorable docking characteristics (< 50% of the ACP models docks to the tentative active site of the TEs). A) UcTE, B) ChTE, C) CrTE

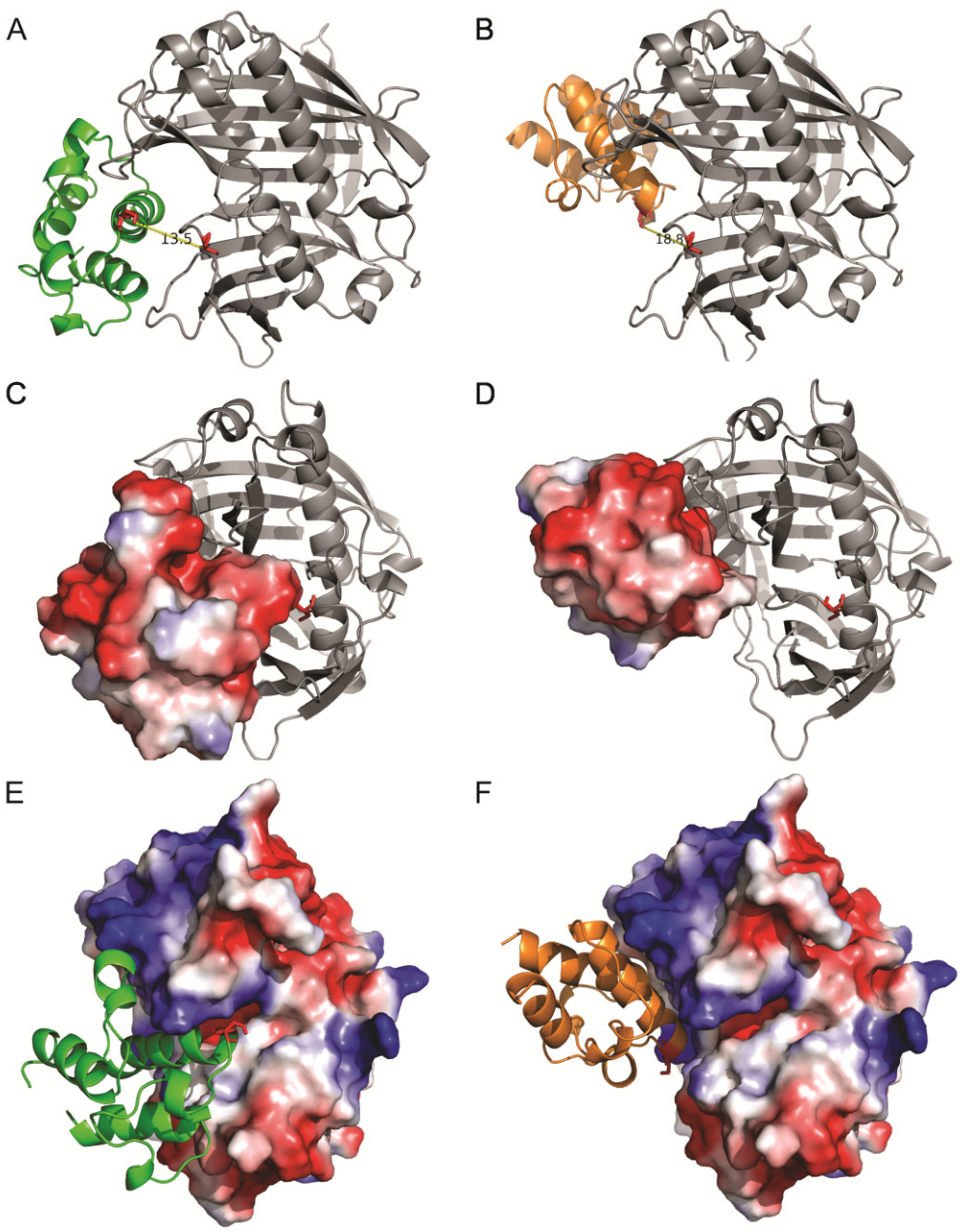


Figure 2-10 Electrostatic potential maps of modeled Cr-cACP/CrTE and Cr-mACP/CrTE complexes

Cr-cACP (green) docked to CrTE (grey) showing a 13.5 Å distance between conserved serine on Cr-cACP and active site cysteine of CrTE; B) Cr-mACP (orange) docked to CrTE (grey) showing a 18.8 Å distance between conserved serine on Cr-mACP and active site cysteine of CrTE; C) Cr-cACP (electrostatic surface display) docked to CrTE (ribbon, grey); D) Cr-mACP (electrostatic surface display) docked to CrTE (ribbon, grey); E) Cr-cACP (ribbon, green) docked to CrTE (electrostatic surface display); F) Cr-mACP (ribbon, orange) docked to CrTE (electrostatic surface display).

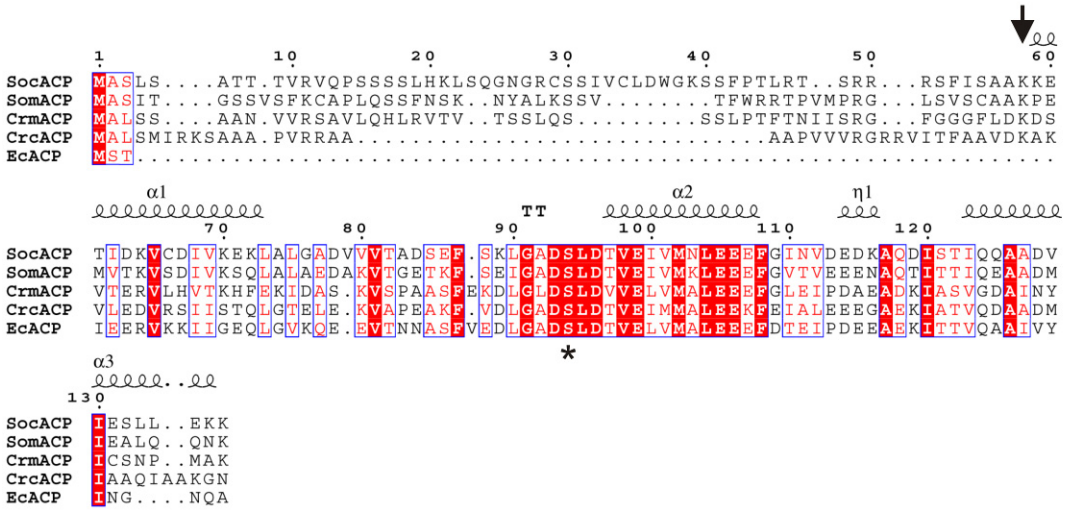


Figure 2-11 Sequence alignment of ACPs

Structure-based sequence alignment of ACPs from *Escherichia coli* (Ec), *Spinacia oleracea* mitochondrial ACP (So-mACP), *Spinacia oleracea* chloroplastic ACP (So-cACP), *Chlamydomonas reinhardtii* mitochondrial ACP (Cr-mACP), and *Chlamydomonas reinhardtii* chloroplastic ACP (Cr-cACP). Conserved serine is indicated by an asterix. A black arrow indicates the ends of transit peptide sequences. Symbols are denoted as: α, α-helices; β, β-strands; η, 310 helices; TTT, strict α-turn; TT, strict β-turn. Sequence-based alignments were produced using Tcoffee [51]. The figure was created using ESPript [52].

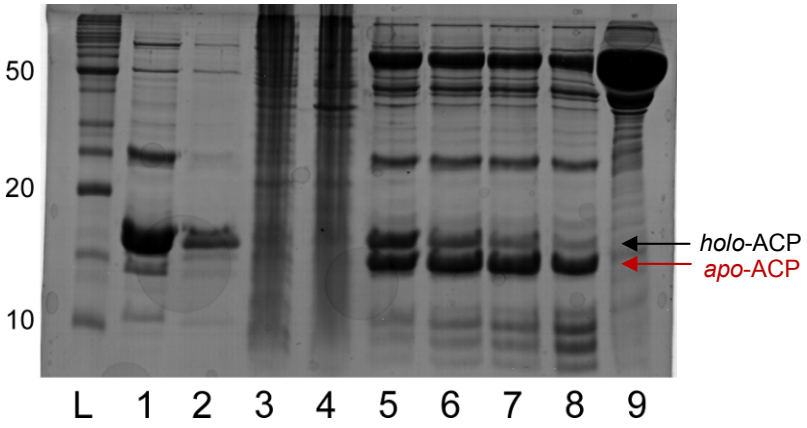


Figure 2-12 Apofication activity of ACP hydrolase on Cr-cACP

L: Benchmark ladder. 1) Ni-NTA purified Cr-cACP (elution 1), 2) Ni-NTA purified Cr-cACP (elution 2), 3 and 4 are crude cell lysates of Cr-cACP expression, 5) apofication of Cr-cACP at t = 1h, 6) t = 4h, 7) t = 8h, and 8) t = 24h, showing the faster running *apo*-Cr-cACP being formed upon cleavage of the phosphopantetheine arm by ACP hydrolase. This was confirmed by mass spectrometry (Figure 2-13) and testing the ability of *apo*-Cr-cACP to get loaded using fluorescent pantetheine analogue **1** (Figure 2-3). The bands at 24 kDa are presumably the disulfide bonded *holo*-Cr-cACP dimer, which can be reduced by the addition of DTT. Proteolysis (bands < 10 kDa) can be prevented by addition of 0.5% sodium azide and 1 mM PMSF to the reaction mixture (data not shown). 9) Purified ACP hydrolase.

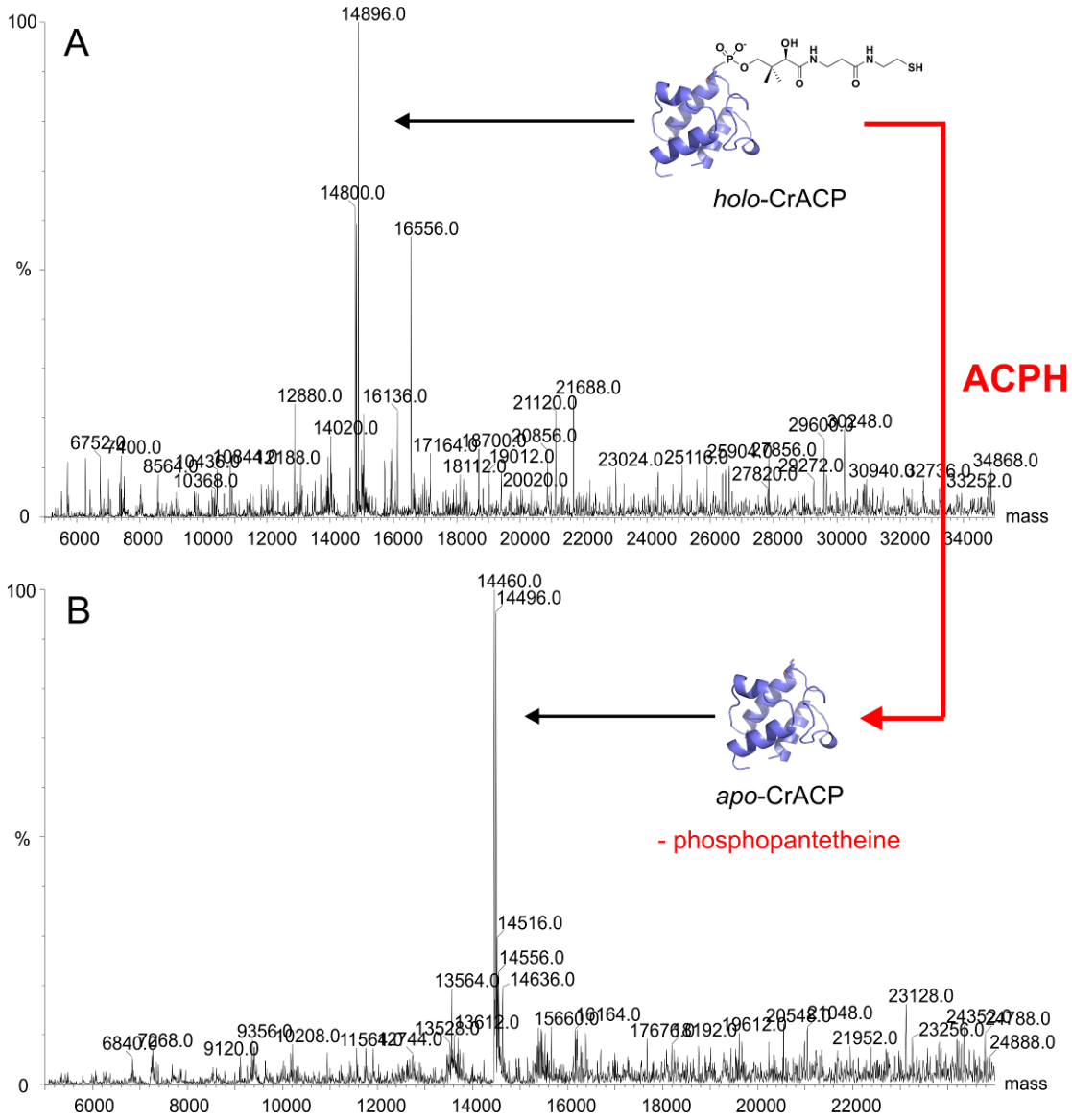


Figure 2-13 Mass spectra of *holo*-CrACP and *apo*-CrACP

A) *holo*-CrACP and B) *apo*-CrACP. ESI-MS were recorded in positive mode. ACP-hydrolase (ACPH) from *Pseudomonas aeruginosa* [20] was used to convert *holo*-CrACP to *apo*-CrACP *in vitro*, which was confirmed by MS to verify the loss of the pantetheine moiety from ACP. We validated the activity of *apo*-CrACP using a one-pot chemoenzymatic method [3], in which incubation of *apo*-CrACP with Sfp, CoA A, D, and E, ATP and fluorescent pantetheine analogue **1**, as shown in Figure 2-3.

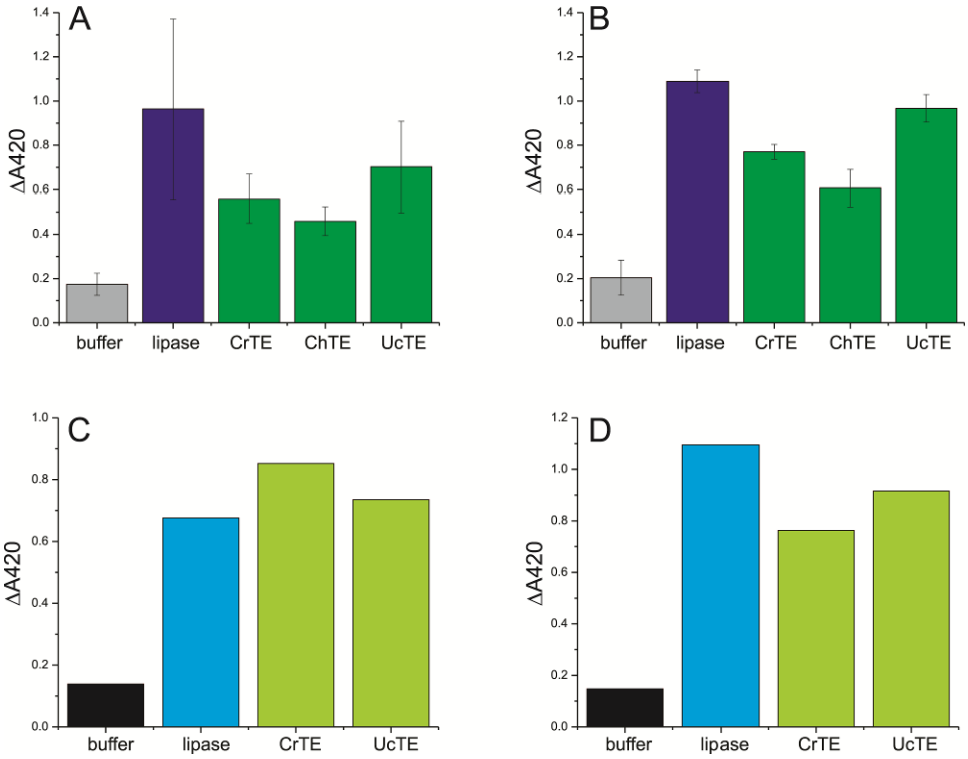


Figure 2-14 Thioesterase activity

Activity of thioesterases and porcine pancreas type II lipase were determined by monitoring the hydrolysis of para-nitrophenylhexanoate for 16 hours at 30 °C. A) pH 7, TEs expressed in *E. coli*; B) pH 8, TEs expressed in *E. coli*; C) pH 7, TEs expressed in *C. reinhardtii*; and D) pH 8, TEs expressed in *C. reinhardtii*.

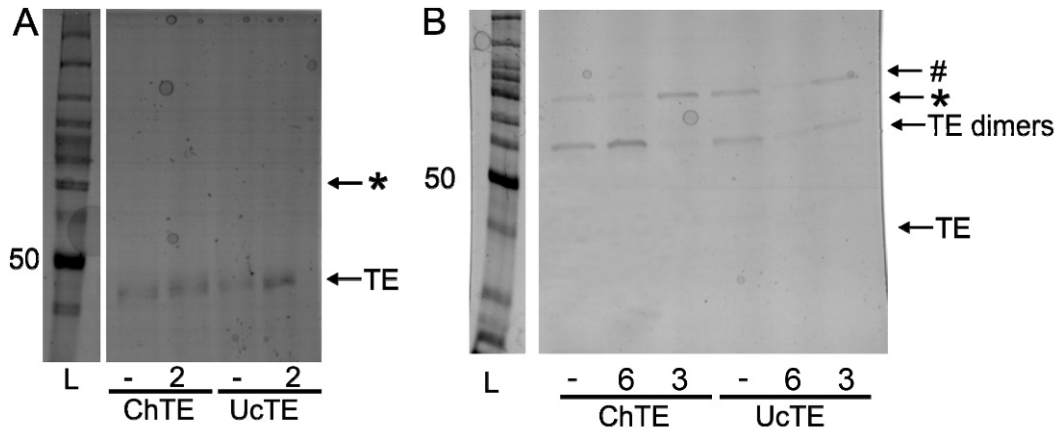


Figure 2-15 Activity-based crosslinking of *C. reinhardtii* cACP and plant TEs

A) *crypto*-Cr-cACP was formed using CoA –A, -D, -E, Sfp, ATP, and chloroacrylic pantetheine analogue 2, and tested for its ability to functionally interact with ChTE and UcTE. Reactions were purified by anti-FLAG resin and eluted with 1 M arginine (pH 3.5) and loaded onto an 8% SDS-PAGE gel. B) *crypto*-Cr-mACP was formed using chloroacrylic pantetheine analogue 6 and 3 and tested for its ability to functionally interact with ChTE and UcTE. Reactions were purified by anti-FLAG resin and eluted with 1 M arginine (pH 3.5) and loaded onto an 8% SDS-PAGE gel. No crosslinked complex was observed between the Cr-cACP and either ChTE or UcTE using all three crosslinking probes, indicating a lack of functional interaction between the two domains. During 24h incubation at 37° C, reduced plant TEs show spontaneous oxidation (bands at ~100 kDa). L = Benchmark protein ladder; Cr = *C. reinhardtii*; Ch = *C. hookeriana*; Uc = *U. californica*. (+) denotes addition of pantetheine analogue 2 (A), 6 or 3 (B); (-) indicates negative control reactions in which pantetheine analogues were omitted. (*) indicates the position a crosslinked complex would be observed. (#) is a co-purifying contamination.

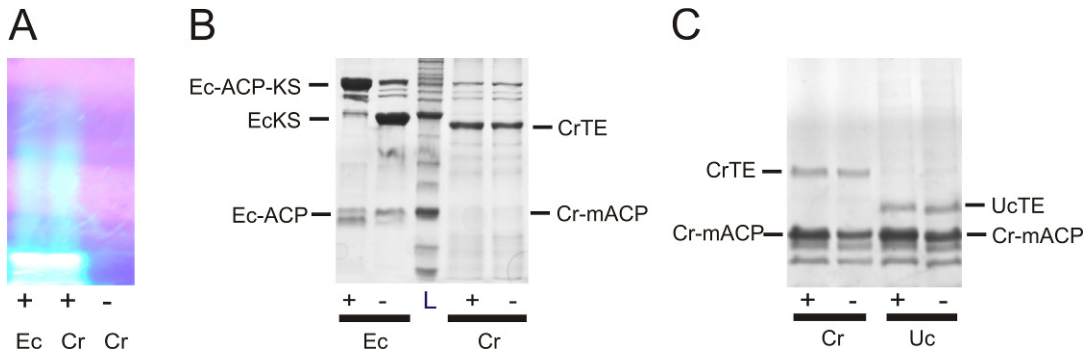


Figure 2-16 Fluorescent loading of *C. reinhardtii* mACP and crosslinking with TEs

A) *crypto*-Cr-mACP was formed using CoA –A, -D, -E, Sfp, ATP, and fluorescent pantetheine analogue 1 [37], validating post translational modification of Cr-mACP. *E. coli* ACP was loaded with 1 as a positive control. Reactions were loaded on an 8% SDS-PAGE gel, visualized at 365 nm. B) *crypto*-Cr-mACP was formed using chloroacrylic pantetheine analogue 2 and tested for its ability to functionally interact with CrTE; no crosslinked complex was observed between the Cr-mACP and CrTE (right 2 lanes), indicating a lack of functional interaction between the two proteins. The *E. coli* ACP and KSII serve as a positive experimental control; a crosslinked complex is observed between the *E. coli* ACP and KSII (left 2 lanes). Reactions were visualized on a 12% SDS-PAGE gel stained with Coomassie; C.) The Cr-mACP fails to generate a crosslinked complex with CrTE (left 2 lanes) and UcTE (right 2 lanes), demonstrating a lack of protein-protein interactions between either the CrTE or UcTE and Cr-mACP, visualized on a Coomassie-stained 12% SDS-PAGE gel. L = Benchmark protein ladder; Ec = *E. coli*; Cr = *C. reinhardtii*; Uc = *U. californica*. (+) addition of pantetheine analogue 1 (A) or 2 (B and C); (-) negative control reactions in which pantetheine analogues were omitted.

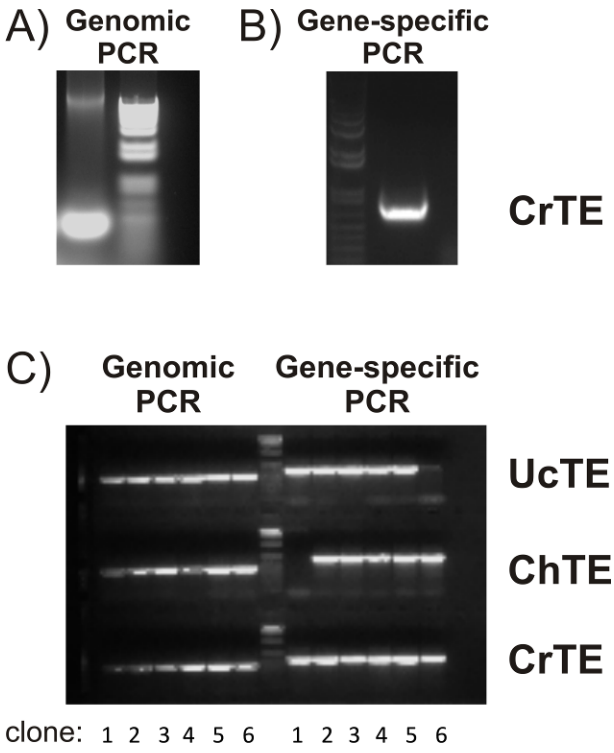


Figure 2-17 PCR screen to identify *C. reinhardtii* transformants harboring TEs

Initial screening of primary Cr transformants was carried out using PCR to determine which clones have proper insertion of exogenous genes. Clones showing proper integration into the chloroplast genome were rescreened to ensure genetic stability. A) PCR analysis of genomic chloroplast DNA to determine degree of homoplasmy in Cr transformed with CrTE. B) PCR on inserted nucleotide sequence of Cr transformed with CrTE. C) Left: PCR on genomic chloroplast DNA to determine degree of homoplasmy in Cr transformed with UcTE, ChTE and CrTE. Right: PCR on inserted nucleotide sequence of Cr transformed with UcTE, ChTE and CrTE.

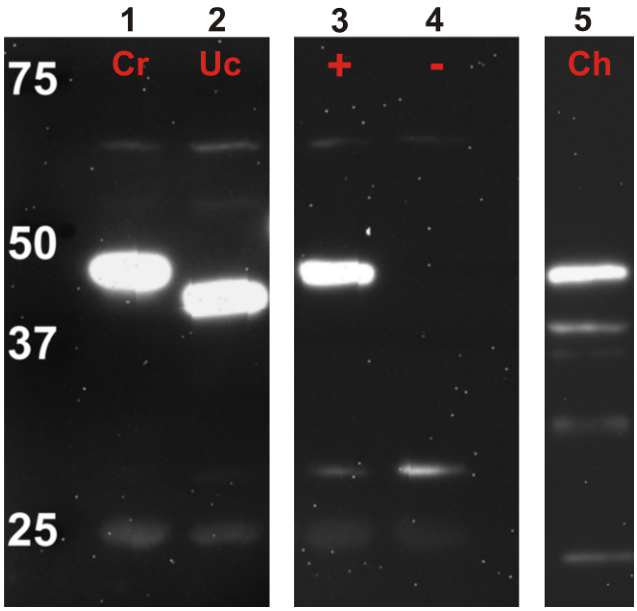


Figure 2-18 Western blot analysis confirming expression of TEs in *C. reinhardtii* chloroplast

To confirm heterologous TE protein expression in the Cr chloroplast, Western blot analysis was conducted on transgenic Cr strains. Protein was purified via affinity purification using M2 FLAG resin and the Western blot was carried out using anti-FLAG antibody. Gel images show Western blot results for TEs overexpressed in *C. reinhardtii*. Each lane detects 97hioesterases97 of a different TE in transgenic Cr chloroplasts: 1) CrTE, 2) UcTE, 3) CrTE [positive control] 4) wildtype strain 137c (mt+) [negative control], 5) ChTE. White numbers (left of gel) indicate molecular weight in kDa. Red letters above each lane denote TE detected: Cr = *C. reinhardtii*; Uc = *U. californica*; Ch = *C. hookeriana*; (+) 97hioesterases97 of native CrTE; (-) wildtype Cr strain 137c.

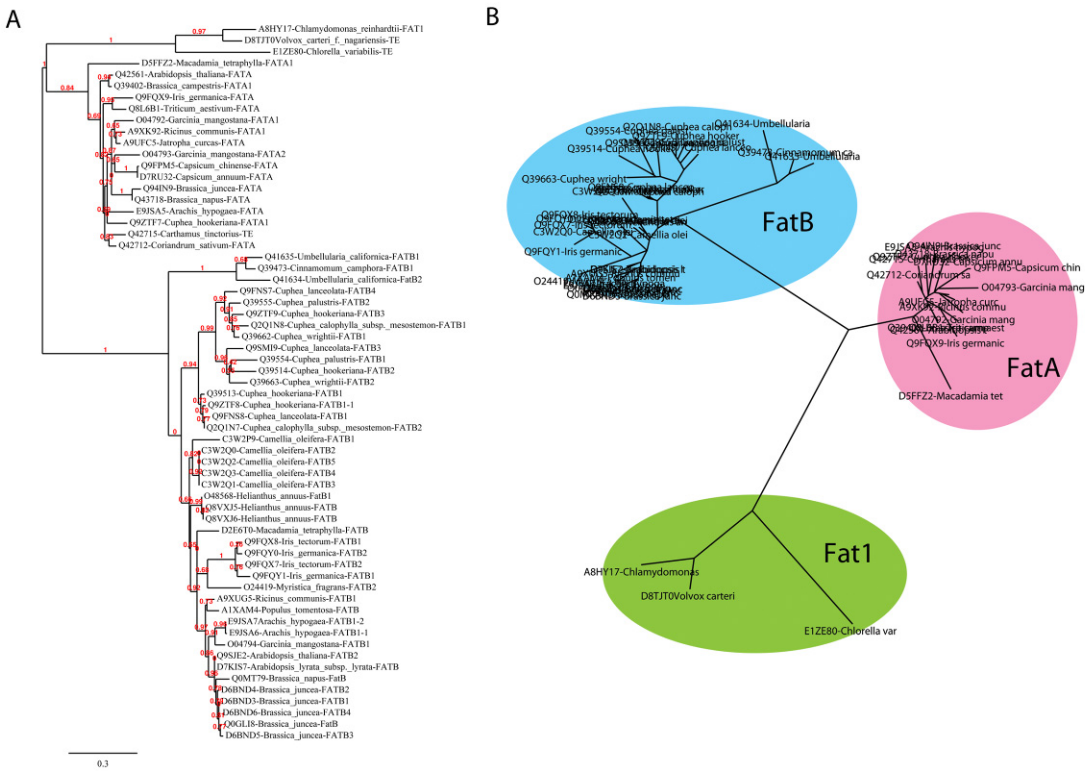


Figure 2-19 Thioesterase phylogeny

The phylogenetic relationship between FatA, FatB, and Fat1 TE protein sequences was analyzed using a web-based phylogeny analysis tool [58]; A) Phylogram of the relationship between FatA, FatB, and Fat1 TEs, generated by Phylogeny.fr [58]; B) Radial representation of the phylogram shown in A.

Materials and Methods

Materials

All chemicals were purchased from Sigma-Aldrich, Fisher Science and VWR in biochemical grade. CoA –A, -D, -E and Sfp were expressed and purified using previously published methods [37]. Protein concentrations were determined using Bradford reagent. Buffers and media were prepared with ultrapure water. Anti-FLAG M2 affinity gel and anti-FLAG antibody were obtained from Sigma-Aldrich. Genomic DNA purification kit was purchased from Promega.

Protein docking

Protein sequences of Cr-cACP (ACP2, Q6UKY5_CHLRE), Cr-mACP (ACP1, Q6UKY4_CHLRE), CrTE (FAT1, A8HY17_CHLRE), ChTE (FATB2, Q39514_CUPHO) and UcTE (FATB1, Q41635) were obtained from the UniProt database. No structures of these proteins are known, thus models were prepared using the automated Swissmodel server [48], which predicts protein structure based on homology modeling. The four steps used by Swissmodel are template superposition, target-template alignment, model building and energy minimization. In the latter, deviations in the protein structure, which have been introduced by the modeling algorithm when joining stiff fragments, are regularized by using the steepest descent energy minimization of the GROMOS96 force field. Models were compared with the models obtained from the i-Tasser server [31] and no significant deviations between the two homology modeling algorithms were observed. Cr-cACP (ACP2, 117 aa) was modeled after ACP from *Aquifex aeolicus* (PDB: 2EHTA, aa 40-114), which shows a

sequence identity of 52.6%, a q-meanscore of 0.7 and a residual RMSD of 0.08. Cr-mACP (ACP1, 128 aa) models after malonyl-ACP from *Bacillus subtilis* (PDB: 2X2BA, aa 50-124) with a sequence identity of 52% and a q-meanscore of 0.3 and a residual RMSD of 0.09. While there are no algal or plant TE crystal structures available, there are two TE structures that show high homology and sequence identity to CrTE in the active site fold. CrTE (395 aa) was modeled after the TE from *Lactobacillus plantarum* (PDB: 2OWNA, aa 84-378), which shows a 17.6% overall sequence identity, 0.4 q-meanscore, and residual RMSD 0.47 Å, and the TE from *Bacteroides thetaiotaomicron* vpi-5482 (PDB: 2ESSA, aa 83-375, 17.5% overall sequence identity). PyMol structures from these two Tes were aligned with the CrTE homology model and residues in the catalytic fold were manually identified and aligned. The *L. plantarum* TE shows a 37.9% identity and the *B. thetaiotaomicron* TE shows a 41.4% sequence identity to CrTE around its active site Cys-His-Asn (8 Å).

With CrACP and CrTE homology models in hand, a panel of fourteen vascular plant Tes were identified that were homologous to CrTE and specific for short chain fatty acids. Tes included those from *Helianthus annuus* (FatB TE, 16:0 TE), *Umbellularia californica* (FatB1 TE), *Cinnamomum camphora* (FatB TE), *Cuphea hookeriana* (16:0 TE, FatB1 TE, FatB2 TE), *Jatropha curcas* (FatB1 TE, FatB2 TE, FatA TE), *Oryza sativa* (FatB TE), *Cuphea lanceolata* (FatB TE), *Cuphea wrightii* (FatB TE, 16:0 TE). Both UcTE (382 aa) and ChTE (415 aa) were also modeled after the TE from *Lactobacillus plantarum*, using the sequences of 99-367aa and 133-398aa, which have overall sequence identities of 18.2% and 22.6% and q-meanscore of 0.5 and 0.4, respectively.

Proteins were docked using the Cluspro server [34]. The top 10 balanced models were manually examined using PyMol [49], and the distances between 101hioesterases active site cysteine residues and acyl carrier protein serine residues measured.

Autodock was used to blindly dock fatty acid-phosphopantetheine (PPT) substrates onto the model of CrTE [50]. The protein structure was prepared by adding hydrogens using H++ [59] and removing water. The ligand was prepared by extracting the stearate-phosphopantetheine structure from the structure of spinach ACP (PDB: 2FVA) and the 18:0-phosphopantetheine was energy minimized using the PRODRG2 server [60]. Other fatty acid-phosphopantetheine substrates were energy minimized similarly. Autodock tools [61] were used to define torsional freedom. The cysteine active site residue was defined as a flexible residue, whereas the rest of the protein was considered rigid. A grid was deployed over the whole protein structure enabling the ligand to be docked in any location. A Lamarckian genetic algorithm and 2.5 Mi energy evaluations were used to dock the ligand onto the protein, using Autodock4. The obtained models were manually evaluated for relevance.

Characterization of mitochondrial and chloroplastic acyl carrier proteins of *C. reinhardtii*

Two sequences for *C. reinhardtii* fatty acid acyl carrier proteins exist in the NCBI and UniProt databases, ACP1 and ACP2. Based upon alignment with plant mitochondrial and chloroplastic ACPs, homology, and BLAST analysis, it is likely that ACP1 and ACP2 correspond to the mitochondrial and chloroplastic ACPs,

respectively. Cr-mACP (ACP1), a 128 amino acid protein with a molecular weight of 13.7 kDa and predicted pI of 4.80. Cr-cACP (ACP2), a 117 amino acid protein with a molecular weight of 12.4 kDa and predicted pI of 4.91. Sequence alignment of the Cr-cACP and Cr-mACP with ACPs from spinach (*Spinacia oleceria*) and bacteria (*Escherichia coli*) reveals a similar four-helix bundle, as shown in Figure 2-11. There is ambiguity when predicting transit peptides of the Cr-cACP and Cr-mACP, suggesting possible dual targeting of CrACPs to mitochondria and chloroplasts.

Cr-cACP contains a 54 amino acid transit peptide according to TargetP1.1 [54]. Interestingly, this algorithm predicts mitochondrial localization for Cr-cACP. ChloroP1.1 predicts a 15 residue plastid targeting sequence for Cr-cACP [53], whereas Mitoprot II v1.01 predicts a 38 amino acid cleavage site following a mitochondrial targeting sequence [62]. Predotar v1.03 also predicts mitochondrial localization for the Cr-cACP with slight possibility of localization to the chloroplast [63].

Cr-mACP contains a 64 amino acid mitochondrial transit peptide according to TargetP1.1 and mitochondrial localization [54]. A 23 residue chloroplast targeting peptide (cTP) is predicted for Cr-mACP by ChloroP1.1 [53], whereas Mitoprot II v1.01 predicts a 50 residue cleavage site following a mitochondrial targeting sequence [62]. Predotar v1.03 predicts chloroplastic localization for the Cr-mACP with slight possibility of localization to the mitochondria [63].

The ambiguity of these predictions suggests dual targeting of CrACPs. Typically, there are two ways a single gene can transport a product to both mitochondria and plastids. One scenario is that a ‘twin’ targeting sequence (where a cTP and mTP exist in tandem at the N-terminus) is used and thus the protein can be

transported to either organelle. Or, the protein can have an 'ambiguous' targeting presequence that can be recognized by both organelles. In this case, either there is a novel signal (not currently known) that a shared import mechanism recognizes (both in mitochondria and chloroplasts) or there is a shared signal that both paths recognize [64]. Although we used a variety of algorithms to predict localization of each CrACP, there was certainly ambiguity for each protein. It is known that these current programs are not well adapted to dual targeting sequences. For instance, networks used in Predotar have difficulty due to interference between the overlapping signals in the ambiguous targeting sequences. Therefore, more experimental evidence is necessary in assigning either of the CrACPs a specific localization. It is clear from crosslinking studies that Cr-cACP and CrTE interact, giving confidence that Cr-cACP is indeed the ACP involved in plastid fatty acid biosynthesis. BLAST analysis with known plant ACPs depicts CrACP1 as the Cr-mACP and CrACP2 as the Cr-cACP.

Table 2-1 Strains, plasmids, and restriction sites

Enzyme	Vector	Organism	Restriction Sites
Cr-cACP	pET-28b	*C. reinhardtii*	Nde1, Xho1
Cr-mACP	pET-28b	*C. reinhardtii*	Nde1, Xho1
CrTE	pET-21a	*C. reinhardtii*	Nde1, Xba1
UcTE	pET-21a	*U. californica*	Nde1, Xba1
ChTE	pET-21a	*C. hookeriana*	Nde1, Xba1
ACPH	pET-24b	*P. aerigunosa*	Nde1, Xho1

Cr-cACP, Cr-mACP, CrTE, ChTE and UcTE gene design

Entelechon (Regensburg, Germany; www.entelechon.com) was used to backtranslate Cr-cACP and Cr-mACP into *E. coli* codon bias and the genes were synthesized by GenScript (Piscataway, NJ). The CrTE, UcTE and ChTE were codon

optimized for *E. coli* and synthesized into a pET21a vector (DNA 2.0, Menlo Park, CA) in which a FLAG tag was added to the carboxy terminus.

Cloning of Cr-cACP, Cr-mACP, CrTE, ChTE and UcTE

Synthesized genes were subcloned into the multiple cloning site of pET28b (Cr-cACP and Cr-mACP) and pET21a (UcTE, CrTE, ChTE) vectors using Nde1 and Xho1 restriction sites and Nde1 and Xba1 restriction sites, respectively.

Expression of Cr-cACP

The Cr-cACP/pET28b plasmid was transformed into chemically competent BL21 (DE3) *E. coli* cells. 5 ml cultures in LB media, containing 50 μg/ml kanamycin, were grown from a single colony at 37 °C overnight. 1 liter cultures of LB media containing 50 mg/L kanamycin were inoculated with these 5 ml overnight cultures and grown for 4h at 37 °C until they reached an OD_{600} of 0.8, at which point they were induced with 500 μM IPTG (final concentration) and grown overnight at 25 °C. Cells were harvested by centrifugation at 4 °C for 30 minutes at 2000 rpm, and the obtained pellets were frozen at -80 °C.

Expression of Cr-mACP

The Cr-mACP/pET28b plasmid was transformed into chemically competent BL21 (DE3) *E. coli* competent cells. 5 ml LB cultures containing 50 μg/ml kanamycin were grown from a single colony at 37 °C overnight. 1 liter cultures of LB containing 50 mg/L kanamycin were inoculated with these 5 ml overnight cultures and grown for 4h at 37 °C. When they reached an OD_{600} of 0.8, cultures were induced with 1 mM

IPTG (final concentration) and grown for 4h at 37 °C. Cells were harvested by centrifugation at 4 °C for 30 minutes at 2000 rpm, and the resultant pellets were frozen at -80 °C.

Expression of TEs

TE plasmids (pET21a/FLAG/CrTE, pET21a/FLAG/UcTE FatB1 and pET21a/FLAG/ChTE FatB2) were transformed into chemically competent BL21 (DE3) *E. coli* cells. 5 ml LB cultures containing 100 μg/ml ampicillin were grown from a single colony at 37 °C overnight. 1 liter cultures of LB containing 100 mg/L ampicillin were inoculated with these 5 ml overnight cultures and grown for 4h at 37 °C (OD_{600} of 0.8), induced with 500 μM IPTG (final concentration), and grown overnight at 25 °C. Cells were harvested by centrifugation at 4 °C for 30 minutes at 2000 rpm. Cell pellets were frozen at -80 °C.

Expression of ACP hydrolase

The *Pseudomonas aerigunosa* ACP hydrolase (ACPH) [36] in either pET24b (MBP-LVPRGSH-ACPH construct) or pMALc2x (MBP-ENLYFQ-HHHHHH-ACPH construct) plasmid was transformed into chemically competent BL21(DE3) *E. coli* cells and 5 ml LB cultures containing 50 μg/ml kanamycin or 100 μg/ml ampicillin were grown from a single colony at 37 °C overnight. 1 liter cultures of LB, in the presence of 50 mg/l kanamycin or 100 mg/l ampicillin, were inoculated with these 5 ml overnight cultures and grown for 4h at 37 °C. When they reached an OD_{600} of 0.8, cultures were induced with 500 μM IPTG (final concentration) and grown overnight at

25 °C. Cells were harvested by centrifugation at 4 °C for 30 minutes at 2000 rpm, and the obtained pellets frozen at -80 °C.

Purification of *holo*-CrACP

The pellet from a 1 liter culture was resuspended in 20 ml of lysis buffer (50 mM potassium phosphate pH 7.4, 500 mM sodium acetate, 0.1 mM EDTA, 20% glycerol) and lysed using a French pressure cell at 4 °C. Cell debris was spun down at 10,000 rpm at 4 °C for 1h. The His_6-tagged protein was purified by Ni-NTA affinity chromatography. ACP was identified as *holo*-ACP by high resolution MS.

Apofication of CrACP and purification of *apo*-CrACP

The purified *holo*-CrACP (0.5 mg/ml) was transferred to dialysis tubing (3000 kDa MWCO) and ACPH, Tris-HCl (pH 8), NaCl, $MgCl_2$, and $MnCl_2$ added to final concentrations of 1 mg/ml and 50, 100, 15 and 1 mM, respectively. This mixture was dialyzed overnight at room temperature against 1 L of 50 mM Tris-HCl (pH 8), 100 mM NaCl, 15 mM $MgCl_2$, and 1 mM $MnCl_2$. The mixture containing the apofied CrACP was purified by Ni-NTA affinity chromatography, concentrated, and further purified by size-exclusion chromatography (FPLC). ACP was identified as *apo*-ACP by high resolution MS.

Purification of Cr-mACP

The pellet from a 1 liter culture was suspended in 20 ml of lysis buffer (50 mM potassium phosphate pH 7.4, 500 mM sodium acetate, 0.1 mM EDTA, 20% glycerol) and lysed using a French pressure cell at 4 °C. Cell debris and inclusion bodies were

spun down at 10,000 rpm at 4 °C for 1h. The pellet was dissolved in 6 M guanidine hydrochloride and the His_6-tagged protein was purified by Ni-NTA affinity chromatography under denaturing conditions. The unfolded protein was folded by sequential overnight dialysis in 6, 3, 2, 1, 0.5 M guanidine hydrochloride at 4 °C.

Purification of TEs

The pellet from 1 liter of culture was suspended in 20 ml of lysis buffer (50 mM potassium phosphate pH 7.4, 500 mM sodium acetate, 0.1 mM EDTA, 20% glycerol) and lysed using a French pressure cell at 4 °C. Cell debris was spun down at 10,000 rpm at 4 °C for 1h. The FLAG-tagged proteins were purified by FLAG-antibody affinity chromatography and eluted off the resin with 1 M arginine pH 3.5. Alternatively, the crude lysate was concentrated and purified using size-exclusion chromatography (FPLC).

Reduction of TEs

TEs were reduced with 10 mM DTT in TBS buffer (pH 7.4) and dialyzed into PBS buffer (pH 6.5) prior to use.

Synthesis of para-nitrophenylhexanoate

Hexanoic acid (1g), N-hydroxysuccinimide (1 eq.) and Et_3N (1.1 eq.) were dissolved in CH_2Cl_2, cooled to 0 °C and 1.1 eq. EDC slowly added while stirring. The mixture was allowed to warm to room temperature and stirred overnight. The activated ester was added dropwise to an ice-cold solution of 4-nitrophenol in CH_2Cl_2. After 3h at 0 °C and 5h at room temperature, ice was added and the mixture carefully acidified,

extracted with ethyl acetate and the aqueous layer basified and extracted. The combined organic layers were dried and evaporated. The product was recrystallized form hot ethanol. (90%, LCMS-ESI calcd. (M+H) 237.2 found 237.2, ^{1}H NMR (400 MHz, $CDCl_3$) δ 0.90 (t, 3H), 1.3 (m, 2H), 1.5 (m, 2H), 2.3 (t, 2H), 7.55 (d, 2H), 8.23 (d, 2H)).

Thioesterase activity assay

Chlamydomonas reinhardtii thioesterase (CrTE), *Cuphea hookeriana* thioesterase (ChTE) and *Umbellularia californica* thioesterase (UcTE) were expressed in *E. coli* strain BL21 and purified by anti-FLAG resin, as described above. Separately, dense 100 ml TAP media cultures supplemented with the appropriate antibiotic of *C. reinhardtii* strains (chloroplastic) transformed with UCTE, ChTE and CrTE, were grown for 5 days, centrifuged, lysed via French pressure cell and the proteins purified by anti-FLAG resin, following the same procedure as for the enzymes expressed in *E. coli*. Protein concentrations were determined by Bradford assay and the amounts used in the assay varied accordingly. Lipase and thioesterases were freshly reduced with dithiothreitol (pH 8) for 1 hour at room temperature and carefully buffer exchanged against 50mM Tris-HCl pH 6.5 containing 5% glycerol. Thioesterase activity was determined by its ability to catalyze hydrolysis of 4-nitrophenylhexanoate. In a total volume of 150 µl, each reaction contained 10 mM Tris buffer (pH 7 or 8), 1.3% acetonitrile, 0.02% triton-X and 0.1 mg/ml enzyme. The mixtures of buffer, triton-X and enzymes were pre-incubated for 10 minutes at 37 °C and subsequently pre-warmed substrate added. Hydrolysis of para-

nitrophenylhexanoate (1 mg/ml, dissolved in acetonitrile) at 30 °C was followed for 16 hours, in duplicate, on a Perkin Elmer HTS 7000 Bioassay reader, at 340 nanometer.

Thioesterase phylogeny

A phylogenetic tree comprising various plant FatA TEs, FatB TEs, and the CrTE was created using web-based tools. All complete FatA, FatB and Fat1 protein sequences were obtained from the UniProt protein database and submitted for analysis using the one-click mode with default settings of Phylogeny.fr [58]. The CrTE, a 'Fat1' TE, exists in a separate clade from FatA and FatB TEs phylogenetically (Figure 2-19). CrTE, an algal thioesterase, likely sits somewhere between bacterial and plant TEs evolutionarily. Plants have evolved two different fatty acid termination pathways. In the prokaryotic pathway, an acyl transferase (AT) catalyzes transfer of palmitic acid directly from 16:0-ACP to glycerol-3-phosphate. Alternately, a FatA TE can hydrolyze oleoyl-ACP releasing free oleic acid for export from the plastid and incorporation into membrane lipids [65].

It has been postulated that the FAS of ancient chloroplasts produced only saturated fatty acid products, indicated by the fact that cyanobacteria produce predominantly 16:0-ACPs [66]. Phylogenetic analyses depict the ubiquitous FatA 18:1-ACP TE as derived from an ancient 16:0-ACP TE [10], which presumably evolved by competing with AT domains for 16:0-ACP substrates. It has also been hypothesized that medium chain specific FatB TEs evolved from the ancient 16:0-specific TE [10]. Crosslinking data suggests that the CrTE is able to siphon 16:0-

ACPs. Fatty acid analysis of *C. reinhardtii* overexpressing CrTE shows an increase in 18:1 fatty acids (Fig. 5). Thus, all data and literature taken together, it is possible that the CrTE ('Fat1') can function as a FatA/B hybrid, able to hydrolyze both saturated and unsaturated fatty acids in the *C. reinhardtii* chloroplast.

Purification of ACPH

The pellet from a 1 liter culture was suspended in 20 ml of lysis buffer (50 mM potassium phosphate pH 7.4, 500 mM sodium acetate, 0.1 mM EDTA, 20% glycerol) and lysed using a French pressure cell at 4 °C. Cell debris was spun down at 10,000 rpm at 4 °C for 1h. The MBP-tagged protein was purified by amylose affinity chromatography and eluted from the resin with 100 mM maltose.

One-pot loading of coumarin-pantetheine (1) onto ACP

Typical one-pot loading reactions contain 50 mM phosphate buffer (pH 8), 2 mM coumarin-pantetheine (**1**) (in DMSO), 12.5 mM $MgCl_2$, 8 mM ATP, 0.4 μg/ml ACP, 0.1 μg/ml CoA-A, 0.1 μg/ml CoA-D, 0.1 μg/ml CoA-E and 0.1 μg/ml Sfp [37]. This mixture (50 μl) was incubated for 1h at 37 °C and proteins separated on a 15% SDS-PAGE gel.

Crosslinking experiments

Crosslinking reactions contained 50 mM phosphate buffer (pH 8), 2 mM crosslinker (**2**, **3** or **6** in DMSO), 12.5 mM $MgCl_2$, 8 mM ATP, 0.4 μg/ml ACP, 0.1 μg/ml CoA-A, 0.1 μg/ml CoA-D, 0.1 μg/ml CoA-E and 0.1 μg/ml Sfp. After 1h incubation at 37 °C, thioesterase (final concentration 0.1 μg/ml) was added and the

reaction mixture further incubated. After overnight incubation, a 10 µl suspension of anti-FLAG resin (or Ni-NTA resin) was directly added to the mixtures, incubated for 30 minutes at room temperature, centrifuged, washed with sterile pH 7.4 TBS buffer, and the proteins eluted with 1 M arginine (pH 3.5) or 100 mM imidazole prior to being loaded onto 8% SDS-PAGE gels. Crosslinked complexes were confirmed by high resolution MS.

Mass analysis of proteins by in gel digestion

Crosslinked bands were excised from the gel, washed with H_2O, 50/50 H_2O/ACN, ACN and 100 mM ammonium bicarbonate prior to drying under vacuum (speed vacuum). Trypsin in a buffer containing $CaCl_2$ and ammonium bicarbonate was added to the dried gel piece and the mixture incubated for 5h at 37 °C. The supernatant was dried under vacuum and 20 µl 5% formic acid added prior to MS analysis.

Chlamydomonas reinhardtii chloroplast engineering

Table 2-2 *C. reinhardtii* strains

Name	Type	Enzyme	Organism	Sequence
Cr_wt	Strain	Cr 137c (mt+)	*C. reinhardtii*	Wildtype strain [47]
Cr_CrTE	Strain	Cr 137c (mt+) CrTE	*C. reinhardtii*	Strain transformed with CrTE [42]
Cr_UcTE	Strain	Cr 137c (mt+) UcTE	*C. reinhardtii*	Strain transformed with UcTE [42]
Cr_ChTE	Strain	Cr 137c (mt+) ChTE	*C. reinhardtii*	Strain transformed with ChTE [42]

Table 2-3 *C. reinhardtii* primers

Number	Type	Enzyme	Organism	Sequence
Primer 1	Primer	3' rev for CrTE	Cr	ATTCCACCATCTGCTAATGCTT
Primer 2	Primer	3' rev for UcTE	Uc	CATTCCAACGAGGAGTTAAAC
Primer 3	Primer	3' rev for ChTE	Ch	ACTCCCGCGGTACTCAATGGTG
Primer 4	Primer	3HB fwd wt	Cr	CGCCACTGTCATCCTTTAAGT
Primer 5	Primer	3HB fwd	Cr	TGTTTGTTAAGGCTAGCTGC
Primer 6	Primer	psbA 5' UTR fwd	Cr	GTGCTAGGTAACTAACGTTTGATTTTT
Primer 7	Primer	Control fwd homoplasty	Cr	CCGAACTGAGGTTGGGTTTA
Primer 8	Primer	Control rev homoplasty	Cr	GGGGGAGCGAATAGGATTAG
Primer 9	Primer	psbA 5' UTR fwd	Cr	GGAAGGGGACGTAGGTACATAAA
Primer 10	Primer	psbA 3' rev	Cr	TTAGAACGTGTTTTGTTCCCAAT
Primer 11	Primer	psbC 5' UTR fwd	Cr	TGGTACAAGAGGATTTTTGTTGTT
Primer 12	Primer	psbD 5' UTR fwd	Cr	TGGTACAAGAGGATTTTTGTTGTT
Primer 13	Primer	atpA 5' UTR fwd	Cr	CCCCTTACGGGCAAGTAAAC

Thioesterases plasmid construction for transformation into *C. reinhardtii*

All DNA and RNA manipulations were carried out using standard methods previously described [21]. Codon-optimized TE genes were cloned into a chloroplast

expression vector, in which the insert is under the control of the 5'UTR and promoter sequence for the *psbD* gene and 3'UTR for the *psbA* gene. The kanamycin resistance gene (*aphA6*) is regulated by the 5'UTR and promoter sequence for *atpA* and 3'UTR for the *rbcL* gene. The transgene cassette is flanked by 5' and 3' homology regions, which target the entire cassette to the 3HB locus of *Chlamydomonas reinhardtii.*

Transformation of thioesterase plasmids into *C. reinhardtii 137c (mt+)*

C. reinhardtii cells were grown to late log phase (7 days) in TAP media in the presence of 0.5 mM 5-fluorodeoxyuridine at 23 °C, under constant illumination of 450 lux on a rotary shaker at 100 rpm. Cells (50 ml) were harvested by centrifugation (4000 x g, 23 °C, 5 minutes) and the supernatant was decanted and the cells resuspended in 4 ml TAP media prior to transformation by particle bombardment [55]. All transformations were performed under kanamycin (100 μg/ml) selection.

PCR was used to identify the transformed strains (Figure 2-17). For PCR analysis, 10^6 cells from plate or liquid culture are suspended in 10 mM EDTA and heated to 95 °C for 10 minutes and then cooled to 23 °C. This mixture was added to a typical PCR cocktail, containing $MgCl_2$, dNTPs, PCR primer pairs, DNA polymerase and water. $MgCl_2$ and EDTA concentrations were carefully controlled. Additionally, genomic DNA (extracted with a Promega Genomic DNA purification kit) was used as a template if crude algal lysates yielded unclear results.

To identify strains containing the thioesterase gene, a primer pair was used in which one anneals to a site in the 5'UTR and the other to the thioesterase coding segment. Desired clones are those that yield the desired size PCR fragment. To

determine the degree to which the endogenous gene locus has been displaced (heteroplasmic versus homoplasmic), a PCR reaction containing two sets of primers was used. The first pair amplifies the endogenous locus targeted by the expression vector. The second pair amplifies a constant, or control, region that is not targeted by the expression vector. The number of cycles was increased to >30 to enhance sensitivity. The most desired clones are those that yield a product from the constant region but not from the endogenous region.

To confirm expression of each thioesterase protein in transgenic *C. reinhardtii* strains, Western blot analysis was performed (Figure 2-18). $5x10^8$ cells were collected by centrifugation from TAP media by centrifugation. Cells were resuspended in cold lysis buffer (25 mM Tris, 137 mM NaCl, 2.7 mM KCl, pH 7.4) supplemented with an EDTA-free protease inhibitor cocktail (Roche). Cells were lysed by sonication, followed by centrifugation. Proteins were purified by FLAG affinity chromatography at 4 °C and the protein eluted off the resin using 0.1 M glycine pH 3.0. Proteins are separated by SDS-PAGE and transferred to PVDF membrane. The membrane is blocked with blocking solution for 30 minutes at room temperature, followed by incubation with HRP-anti-FLAG antibody for 12h at 4 °C, washed, and visualized by HRP-substrate and chemiluminescent detection.

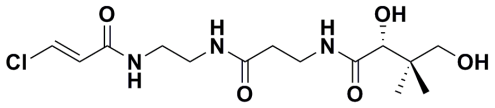
***Chlamydomonas reinhardtii* growth and fatty acid analysis**

Chlamydomonas reinhardtii strains were maintained on TAP media [56], supplemented with 40 μg/ml carbendazim, 500 μg/ml ampicillin and 100 μg/ml cefotaxime (TAP+C+A+C) [57] (and in the case of the plant TEs, an additional 100

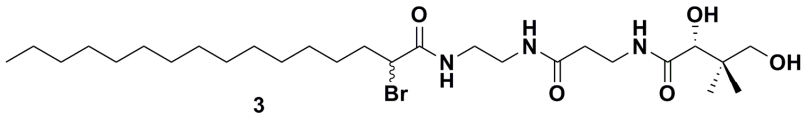
μg/ml kanamycin (+K)), agar plates under constant illumination. Sterile 150 ml flasks containing 50 ml TAP+C+A+C (+K) media were inoculated with single colonies and grown under constant shaking in a greenhouse. Cell density of the cultures was measured. After 3 days, cultures were centrifuged at 4000 rpm at 4 °C, and the cell pellet resuspended in 2 ml of a 1 M methanolic acid solution. The suspension was incubated for 30 minutes at 65 °C and the fatty acid methyl esters extracted using 2 ml of hexanes. The fatty acid composition was determined by GC/MS analysis on an Agilent 7890A GC system connected to a 5975C VL MSD quadrupole MS (EI). Helium was used as a carrier gas. A method employing a gradient of 110 °C to 200 °C at 15 °C /min followed by 20 minutes at 200 °C on a 60 meter DB23 column was used to accurately measure fatty acid methyl ester content and composition. Statistical analyses were performed using SPSS (v13.0), and for all data analysis, a p-value <0.5 was considered statistically significant.

Activity-based pantetheine probes used in crosslinking studies

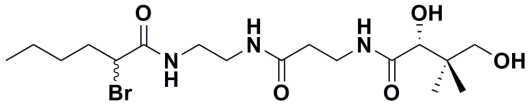
Pantetheine analogue **2** was used in previous studies to elucidate ACP-KS interactions in prokaryotic systems (26). Here, C16-α-bromoamide probe **3** was designed as a site-specific inhibitor of thioesterase activity. As such, a reactive bromide was placed α to the site of nucleophilic attack by the TE. We hypothesized that the longer aliphatic tail of **3** could possibly help orient the amide towards the TE active site. To test TE specificity, a six carbon analogue (**6**) of **3** was also envisioned. Activity-based probes **3** and **6** were synthesized by adapting methods previously described [25].



2



3



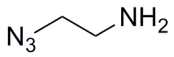
6

Summary of synthesis of C16-α-bromoamide probe (3)

α-Bromopalmitamide **3** was synthesized using a modified procedure of our prior work [26]. In short, 2-azidoethanamine was coupled to PMB-protected pantothenate, followed by reduction of the terminal azide to the amine. PyBOP-

catalyzed peptide bond formation between commercially available 2-bromopalmitate and the free amine followed by deprotection afforded compound **3**.

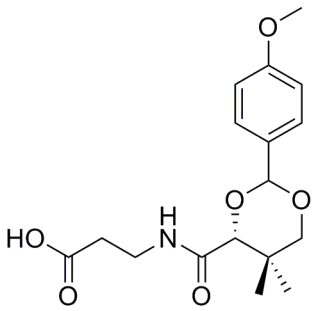
Synthesis of 2-azidoethanamine (S2)



S2

2-Bromoethylamine hydrobromide (2.0 g, 9.76 mmol) was added to a solution of sodium azide (1.9 g, 29.3 mmol) in H_2O (10 mL). The stirred solution was heated to 75 °C for 24 h before it was cooled to 0 °C. Et_2O (10 mL) was added followed by solid KOH (3 g). The organic phase was separated and the aqueous layer extracted with Et_2O (3x50 mL). The combined organic layers were dried with Na_2SO_4 and the solvent carefully removed by air flow. The colorless oil was dried overnight in a desiccator (90%) to afford azide **S2**. ^{1}H NMR (400 MHz, $CDCl_3$): δ = 1.68 (s), 2.87 (t), 3.36 (t) ppm. ^{13}C NMR (100 MHz, $CDCl_3$): δ = 41.45, 54.73 ppm. ESI-MS $[M+H]^+$ calcd. 86.06, found 86.10.

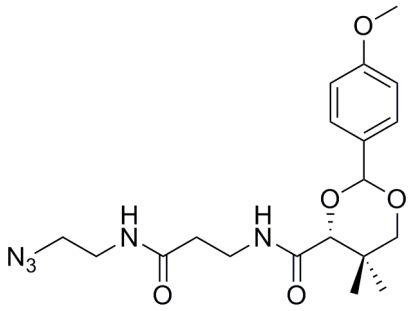
Synthesis of 3-((4*R*)-2-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-4-carboxamido)propanoic acid (S3)



S3

D-Pantothenic acid hemicalcium salt (10 g, 42.0 mmol) was dissolved in anhydrous DMF (50 mL). Concentrated H_2SO_4 (1.3 mL, 42 mmol) was added dropwise and stirred for 30 minutes. p-Anisaldehyde dimethyl acetal (7.2 mL, 42 mmol) and CSA (1.0 g, 4.6 mmol) were added and the reaction was stirred for 16 h. The reaction mixture was partitioned between EtOAc (250 mL) and H_2O (250 mL). The organic layer was washed with saturated $NaHCO_3$ (5x) and brine (5x). The organic layer was dried (Na_2SO_4) and evaporated. The resulting white solid was washed with cold dichloromethane to remove any remaining p-anisaldehyde dimethyl acetal to yield the desired product as a white crystalline solid (75%). ^{1}H-NMR ($CDCl_3$, 400 MHz): δ= 9.87 (s), 7.41 (d), 6.91 (d), 5.45 (s), 4.10 (s), 3.79 (s), 3.69-3.49 (m), 2.59 (t), 1.09 (s), 1.08 (s). ^{13}C-NMR ($CDCl_3$, 100 MHz): δ= 177.03, 169.75, 160.41, 130.40, 127.64, 113.96, 101.42, 83.90, 78.68, 55.52, 34.31, 34.03, 33.33, 22.00, 19.31. ESI-MS $[M+H]^+$ calcd. 337.15, found 337.20.

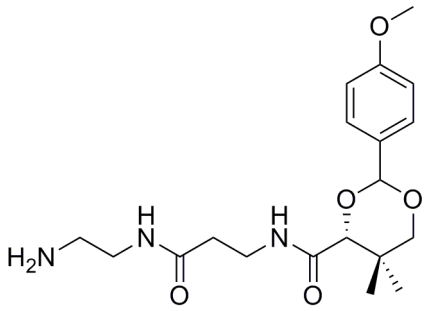
Synthesis of (4*R*)-N-(3-(2-azidoethylamino)-3-oxopropyl)-2-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-4-carboxamide (S4)



S4

Pantothenate **S3** (3.3 g, 9.76 mmol), 2-azidoethanamine (840 mg, 9.76 mmol), EDC (3.7 g, 19.5 mmol), and HOBt (3.0 g, 19.5 mmol) were combined and dissolved in dry DMF (10 mL). DIPEA (3.4 mL, 19.5 mmol) was added, and the reaction was allowed to stir overnight. The reaction mixture was partitioned between EtOAc and H_2O and the organic phase washed with saturated $NaHCO_3$ (3 x 50 mL) and brine (3 x 50 mL). Elution through a plug of Si-gel with EtOAc gave the pure product **S4** as a yellow oil (75%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.36 (d), 7.06 (bt), 6.93 (bt), 6.85 (d), 5.39 (s), 3.99 (s), 3.73 (s), 3.62 (q), 3.46 (m), 3.26 (m), 2.37 (t), 1.04 (s), 1.01 (s). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 171.62, 169.74, 160.40, 130.34, 127.74, 113.89, 101.49, 84.00, 78.57, 55.50, 39.06, 35.88, 35.08, 33.22, 22.02, 19.31. ESI-MS $[M+H]^+$ calcd. 406.20, found 406.20.

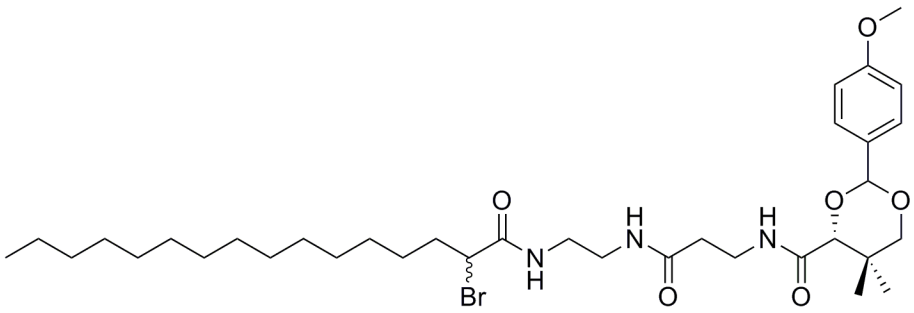
Synthesis of (4*R*)-N-(3-(2-aminoethylamino)-3-oxopropyl)-2-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-4-carboxamide (S5)



S5

To a solution of azide **S4** (2.9 g, 7.2 mmol) in THF (10 mL), PPh_3 (2.1 g, 8.0 mmol), and water (1 ml) were added and stirred at room temperature for 24 h. Evaporation of the solvent under reduced pressure gave oil from which amine **S5** was isolated by flash chromatography (CH_2Cl_2 to 15% MeOH/CH_2Cl_2/1.5% Et_3N) as a golden oil (70%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.41 (d), 7.09 (bt), 6.92 (d), 6.54 (bt), 5.45 (s), 4.07 (s), 3.81 (s), 3.68 (q), 3.55 (m), 3.25 (m), 2.77 (t), 2.63 (q), 2.43 (t), 1.89 (bs), 1.10 (s), 1.08 (s). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 172.53, 170.10, 160.34, 130.28, 127.74, 113.87, 101.44, 83.96, 78.48, 55.46, 40.63, 35.80, 35.50, 33.15, 21.89, 19.22. ESI-MS $[M+H]^+$ calcd. 380.21, found 380.30.

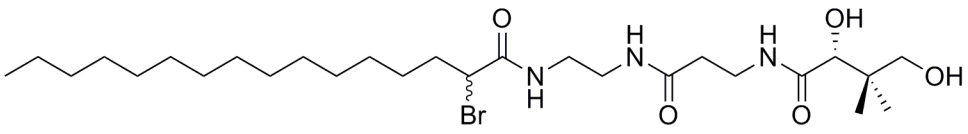
Synthesis of (4*R*)-N-(3-(2-(2-bromohexadecanamido)ethylamino)-3-oxopropyl)-2-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-4-carboxamide (S6)



S6

Amine **S5** (500 mg, 1.32 mmol), PyBOP (754 mg, 1.45 mmol), and α-bromopalmitate (442 mg, 1.32 mmol) were combined and dissolved in dry CH_2Cl_2 (25 mL). DIPEA (460 μL, 2.64 mmol) was added at 0 °C and stirred overnight. The solvent was removed under reduced pressure, and the resultant oil was purified by column chromatography (CH_2Cl_2 to 5% CH_2Cl_2/MeOH) to yield compound **S6** as a white solid (90%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.41 (d), 7.09 (bs), 6.89 (d), 5.43 (s), 4.19 (m), 4.04 (s), 3.77 (s), 3.64 (q), 3.50 (m), 3.30 (m), 2.88 (bs), 2.39 (t), 2.01 (bs), 1.22 (bs), 1.06 (s), 1.03 (s), 0.85 (t). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 172.53, 170.65, 170.06, 160.44, 130.27, 127.80, 113.94, 101.55, 83.98, 78.57, 55.54, 50.69, 40.16, 39.31, 36.32, 35.67, 35.31, 33.27, 32.15, 29.94-29.60, 29.17, 27.62, 22.93, 22.08, 19.36, 14.38. ESI-MS $[M+H]^+$ calcd. 696.35, found 696.50.

Synthesis of (*R*)-2-bromo-N-(2-(3-(2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl)hexadecanamide (3)



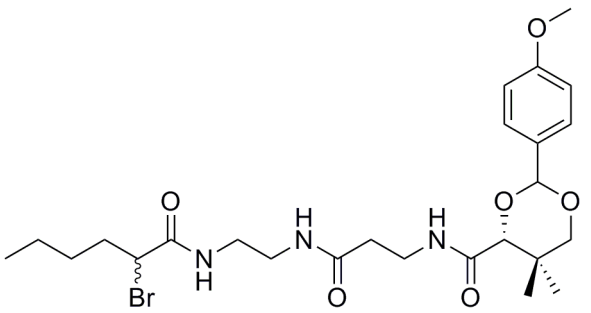
3

PMB-protected bromopalmitoyl pantetheine **S6** (50 mg, 0.07 mmol) was dissolved in 10 mL dioxane and cooled. An equivalent volume of 1 M HCl (aq) was added dropwise. The reaction was allowed to slowly attain room temperature and stirred for 2 h. Once the reaction was complete, solid $NaHCO_3$ was added to the reaction mixture until the solution was neutral. The mixture was evaporated and extracted with CH_2Cl_2, the solvent evaporated and the crude product purified by column chromatography (CH_2Cl_2 to 20% MeOH/CH_2Cl_2) to yield the white solid product as a mixture of diastereomers (80%). ^{1}H-NMR (400 MHz, $(CD_3)_2SO$) δ 8.27 (bt), 7.90 (bt), 7.69 (bt), 5.35 (d), 4.45 (t), 4.27 (t), 3.68 (d), 3.29-3.05 (m), 2.48 (bs), 2.23 (t), 1.21 (bs), 0.85 (t), 0.78 (s), 0.75 (s). ^{13}C-NMR (100 MHz, $(CD_3)_2SO$) δ 173.51, 171.37, 169.23, 86.94, 75.70, 68.72, 50.32, 39.24, 38.62, 35.88, 35.46, 35.28, 31.98, 29.73-29.40, 28.94, 27.46, 22.78, 21.59, 20.98, 14.63. ESI-MS $[M+H]^+$ calcd. 578.31, found 578.40.

Summary of synthesis of C6-α-bromoamide probe (6)

The synthesis of α-bromopalmitamide **6** was achieved by PyBOP-catalyzed peptide bond formation between the free amine **S5** and commercially available 2-bromohexanoic acid to afford **S7**, followed by deprotection.

Synthesis of (4*R*)-N-(3-(2-(2-bromohexanamido)ethylamino)-3-oxopropyl)-2-(4-methoxyphenyl)-5,5-dimethyl-1,3-dioxane-4-carboxamide (S7)

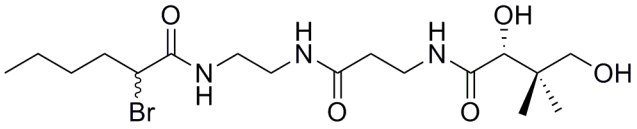


S7

Amine **S5** (250 mg, 0.66 mmol), PyBOP (343 mg, 0.66 mmol), and 2-bromohexanoic acid (94 µl, 0.66 mmol) were combined and dissolved in dry CH_2Cl_2 (15 mL). DIPEA (230 µL, 1.32 mmol) was added at 0 °C and stirred overnight. The reaction mixture was extracted with saturated aqueous citric acid, the solvent removed under reduced pressure, and the resultant oil was purified by column chromatography (CH_2Cl_2 to 10% MeOH/CH_2Cl_2) to yield compound **S7** as a golden clear oil (90%). ^{1}H-NMR (400 MHz, $CDCl_3$) δ 7.42 (d), 7.22 (bs), 7.01 (bt), 6.90 (d), 6.81 (bs), 5.44 (s), 4.20 (m), 4.06 (s), 3.80 (s), 3.67 (q), 3.52 (bq), 3.34 (bs), 3.26 (bm), 2.41 (t), 2.04 (bm), 1.96 (bm), 1.32 (bm), 1.09 (s), 1.06 (s), 0.88 (t). ^{13}C-NMR (100 MHz, $CDCl_3$) δ 172.20, 170.26, 169.87, 160.47, 130.36, 127.76, 113.96, 101.56, 84.07, 78.66, 55.56,

50.99, 40.56, 39.50, 36.40, 35.52, 35.22, 33.30, 29.62, 22.19, 22.10, 19.38, 14.07. ESI-MS $[M+H]^+$ calcd. 556.19, found 556.30.

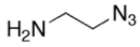
Synthesis of (*R*)-2-bromo-N-(2-(3-(2,4-dihydroxy-3,3-dimethylbutanamido)propanamido)ethyl)hexanamide (6)

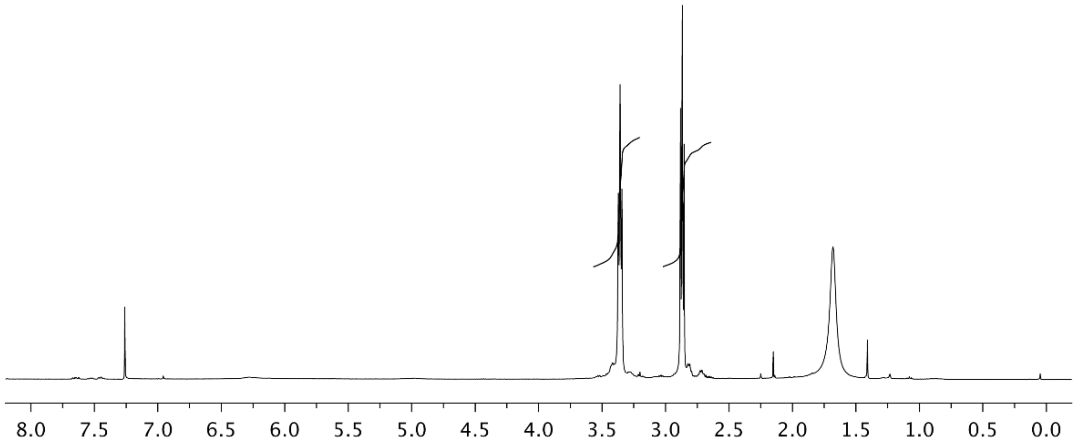


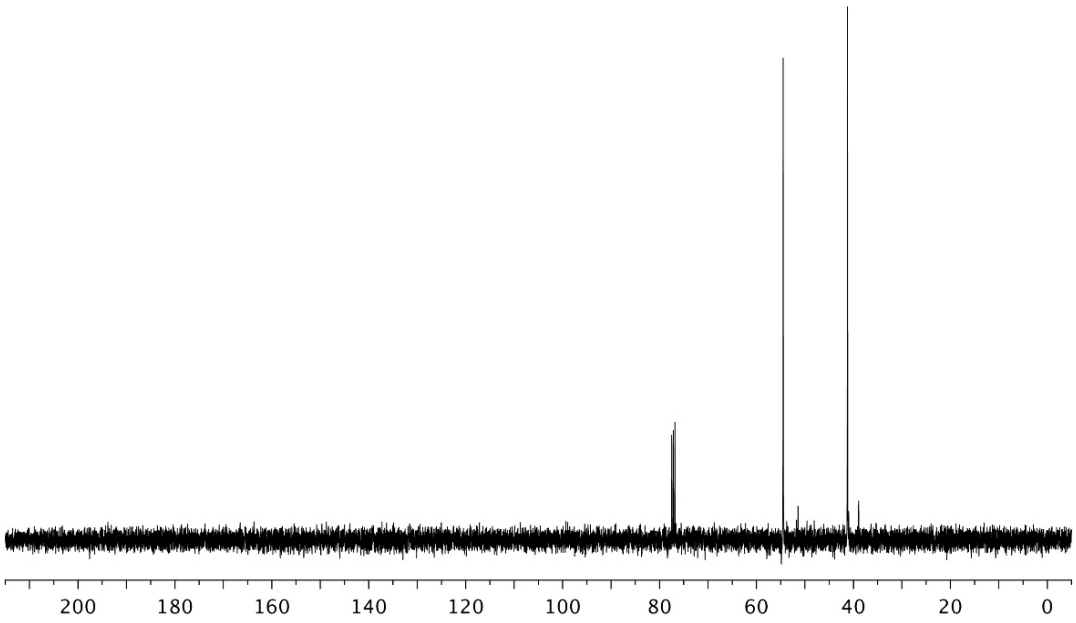
6

PMB-protected bromohexanoyl pantetheine **S7** (50 mg, 0.09 mmol) was dissolved in 10 mL dioxane and cooled. An equivalent volume of 1 M HCl (aq) was added dropwise. The reaction was allowed to slowly attain room temperature and stirred for 4 h. Once the reaction was complete, solid $NaHCO_3$ was added to the reaction mixture until the solution was neutral. The mixture was evaporated and the crude product purified by column chromatography (CH_2Cl_2 to 20% MeOH/CH_2Cl_2) to yield the product as a mixture of diastereomers (80%). ^{1}H-NMR (400 MHz, $(CD_3)_2SO$) δ 8.28 (bt), 7.89 (bt), 7.69 (bt), 5.35 (d), 4.45 (t), 4.27 (t), 3.68 (d), 3.29-3.05 (m), 2.48 (bs), 2.23 (t), 1.90 (m), 1.83 (m), 1.27 (m), 0.85 (t), 0.78 (s), 0.75 (s). ^{13}C-NMR (100 MHz, $(CD_3)_2SO$) δ 173.51, 171.39, 169.24, 75.72, 68.72, 50.30, 39.24, 38.62, 35.90, 35.47, 35.03, 29.65, 22.14, 21.60, 20.99, 14.43. ESI-MS $[M+H]^+$ calcd. 438.15, found 438.20.

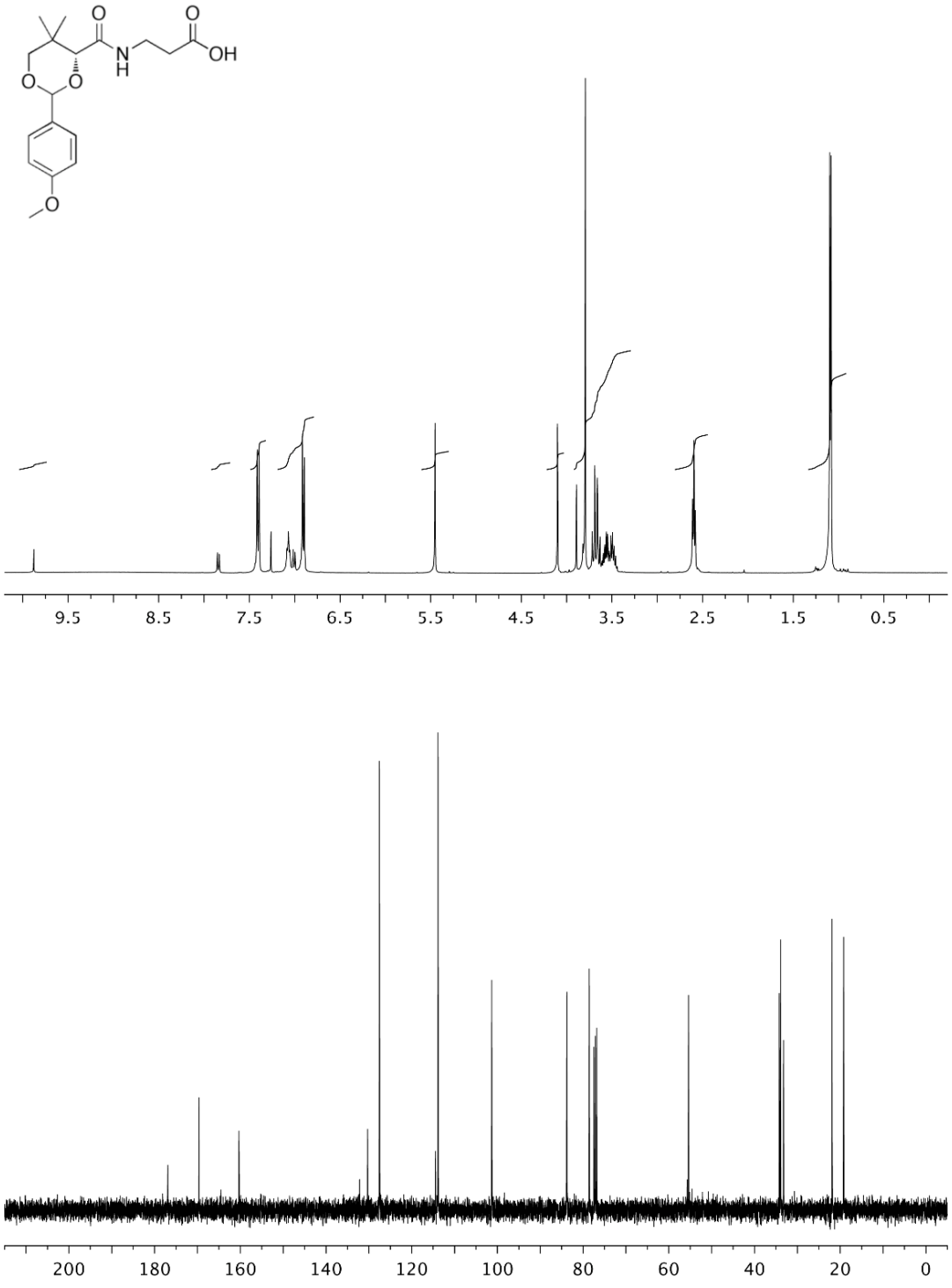
^{1}H NMR (400 Mhz) and ^{13}C NMR (400 Mhz) of amine **S2** in $CDCl_3$



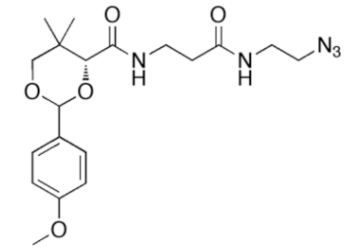


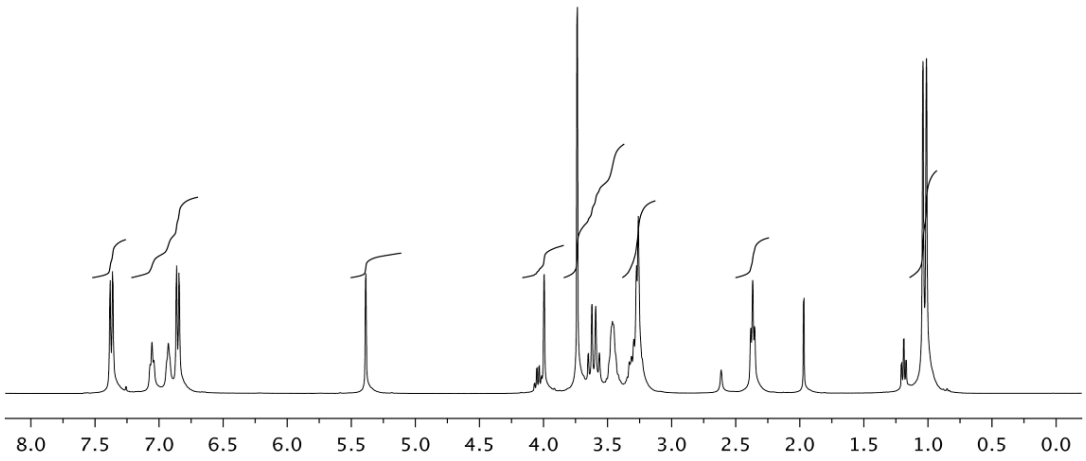


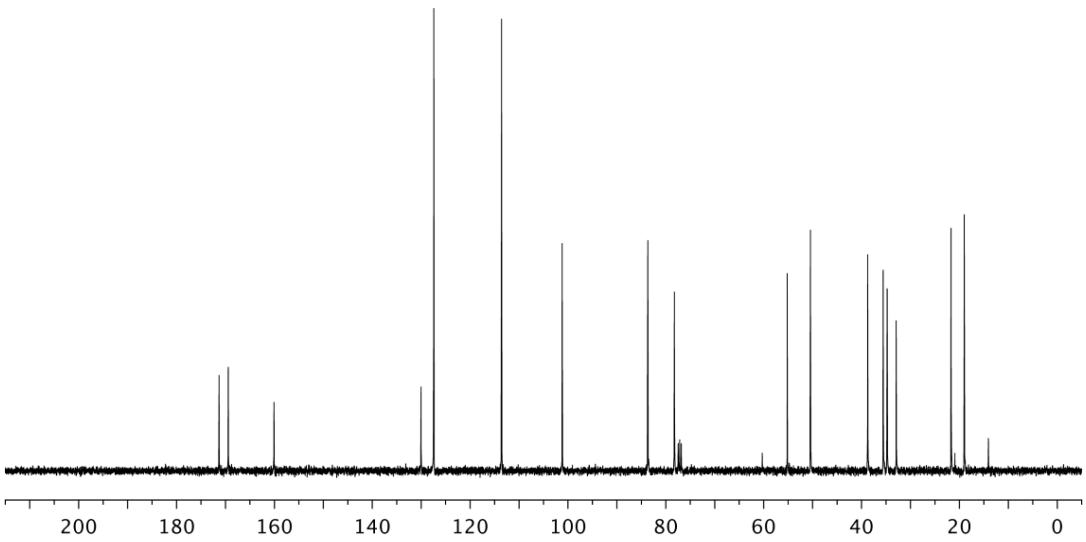
^{1}H NMR (400 Mhz) and ^{13}C NMR (400 Mhz) of acid **S3** in $CDCl_3$



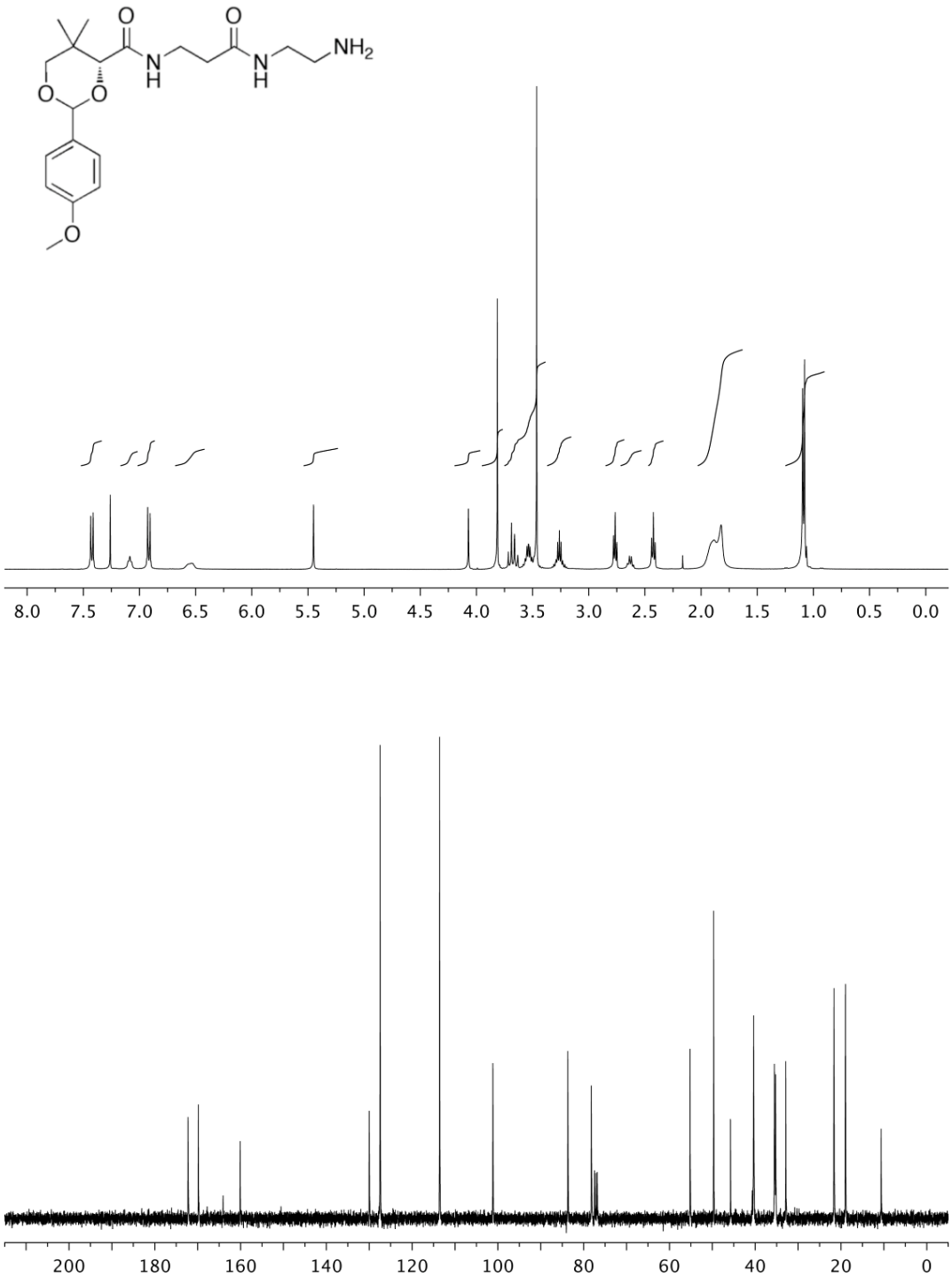
^{1}H NMR (400 Mhz) and ^{13}C NMR (400 Mhz) of azide **S4** in $CDCl_3$



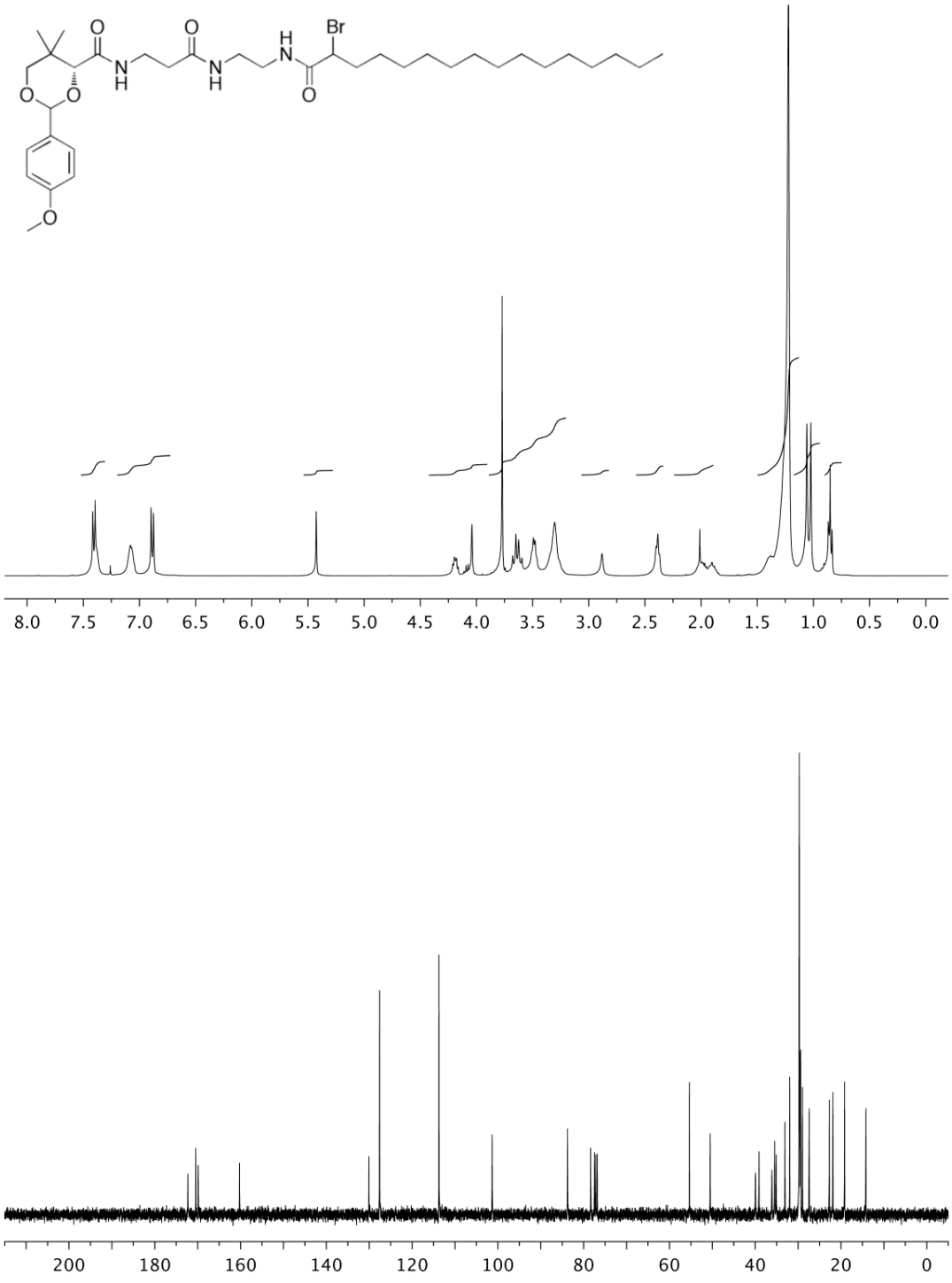




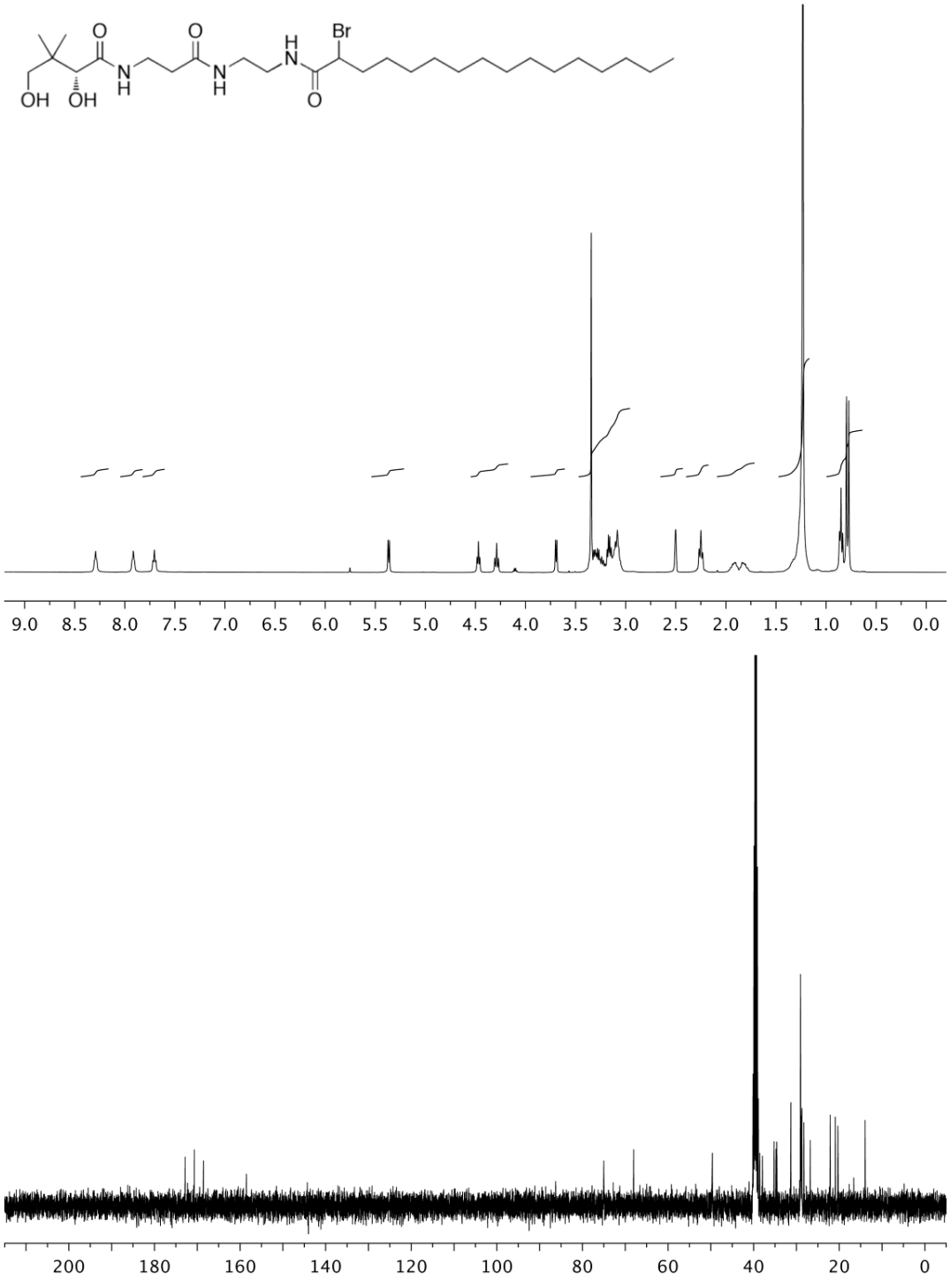
^{1}H NMR (400 Mhz) and ^{13}C NMR (400 Mhz) of amine **S5** in $CDCl_3$



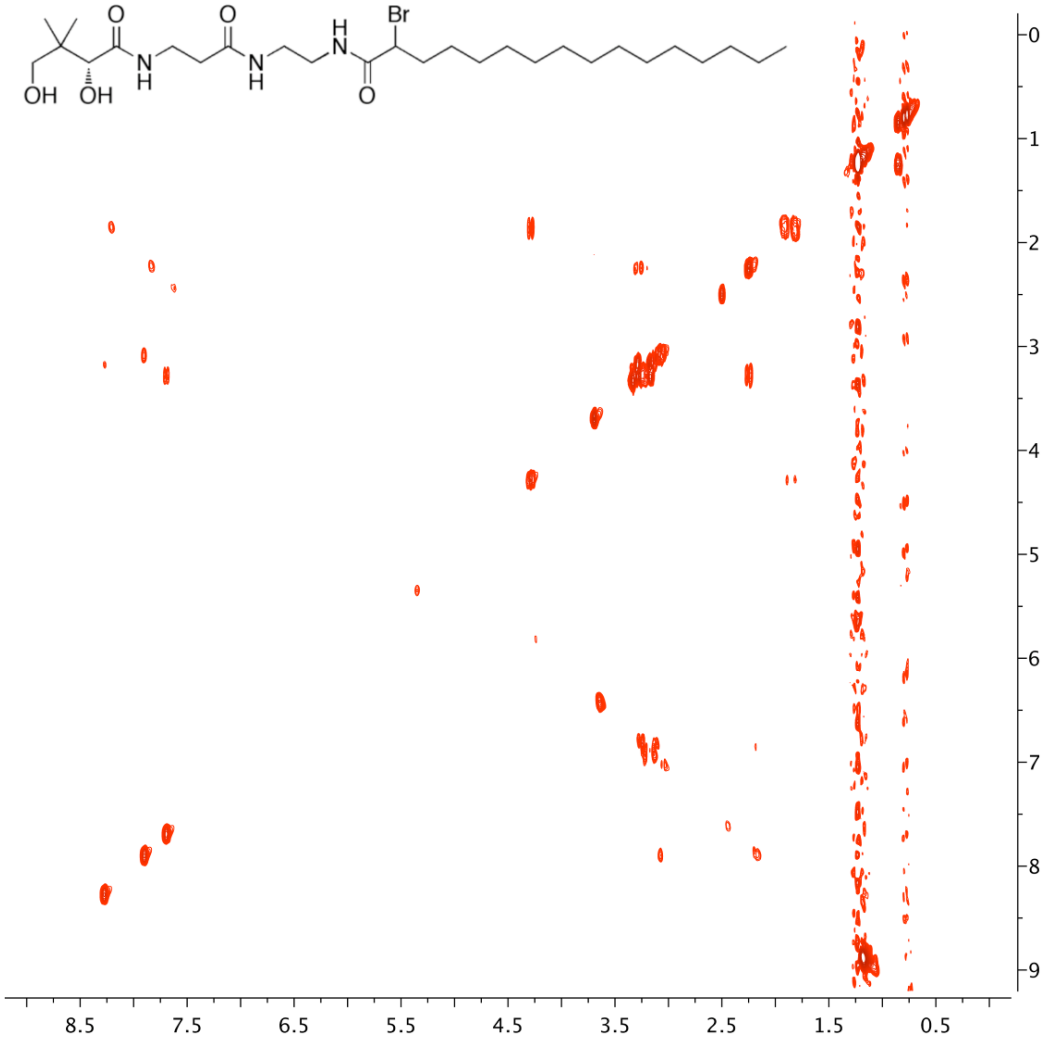
^{1}H NMR (400 Mhz) and ^{13}C NMR (400 Mhz) of amide **S6** in $CDCl_3$



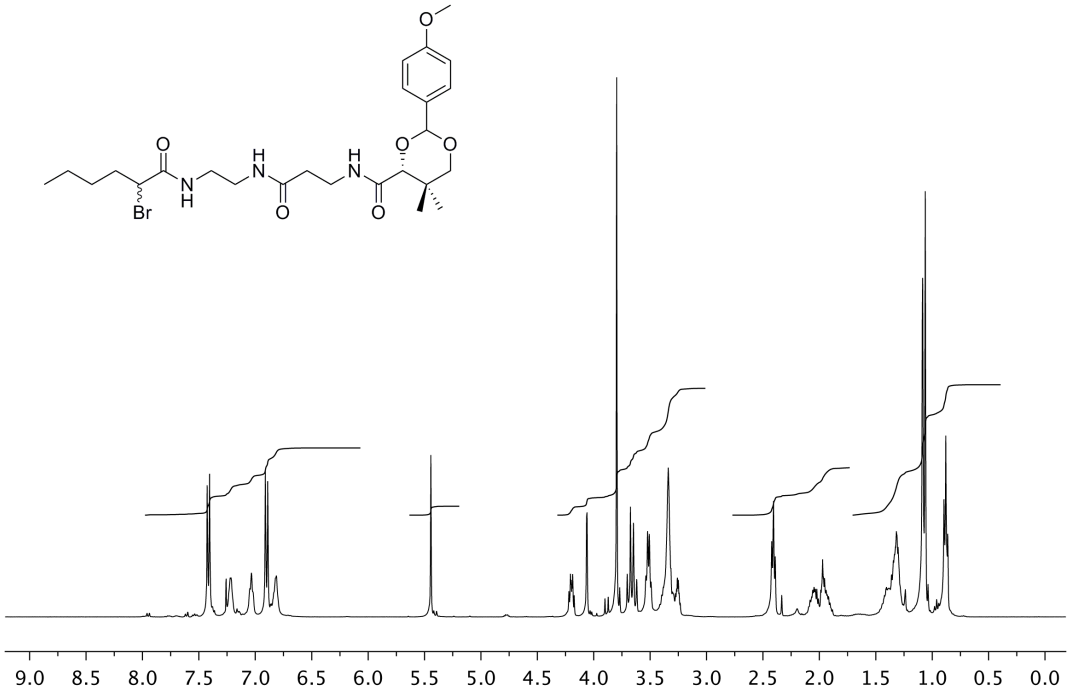
^{1}H NMR (400 Mhz) and ^{13}C NMR (400 Mhz) of probe **3** in d_6–DMSO

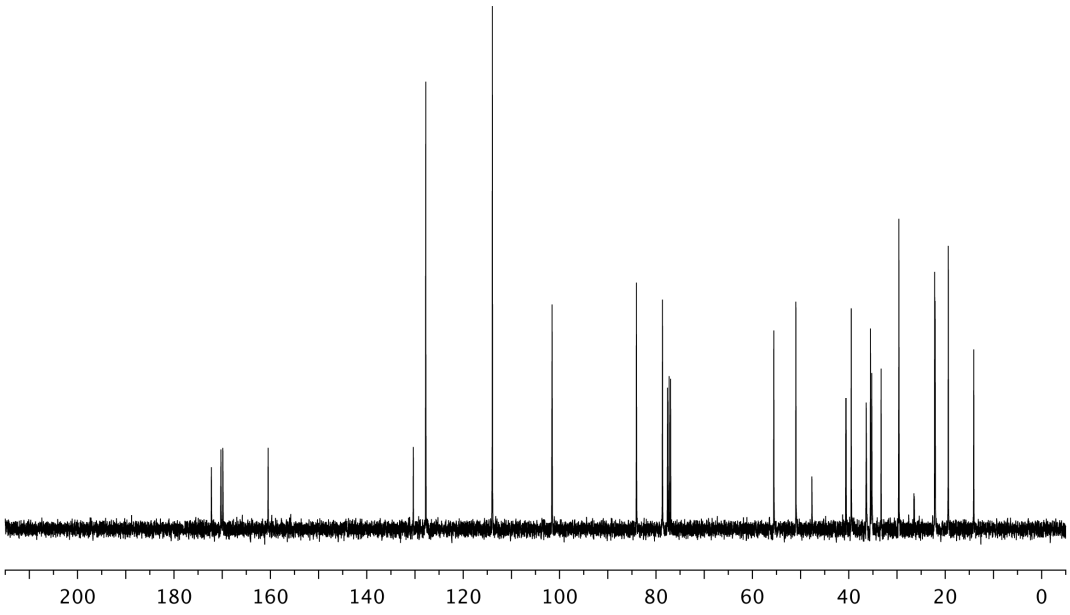


gCOSY (500 Mhz) of probe **3** in d_6–DMSO

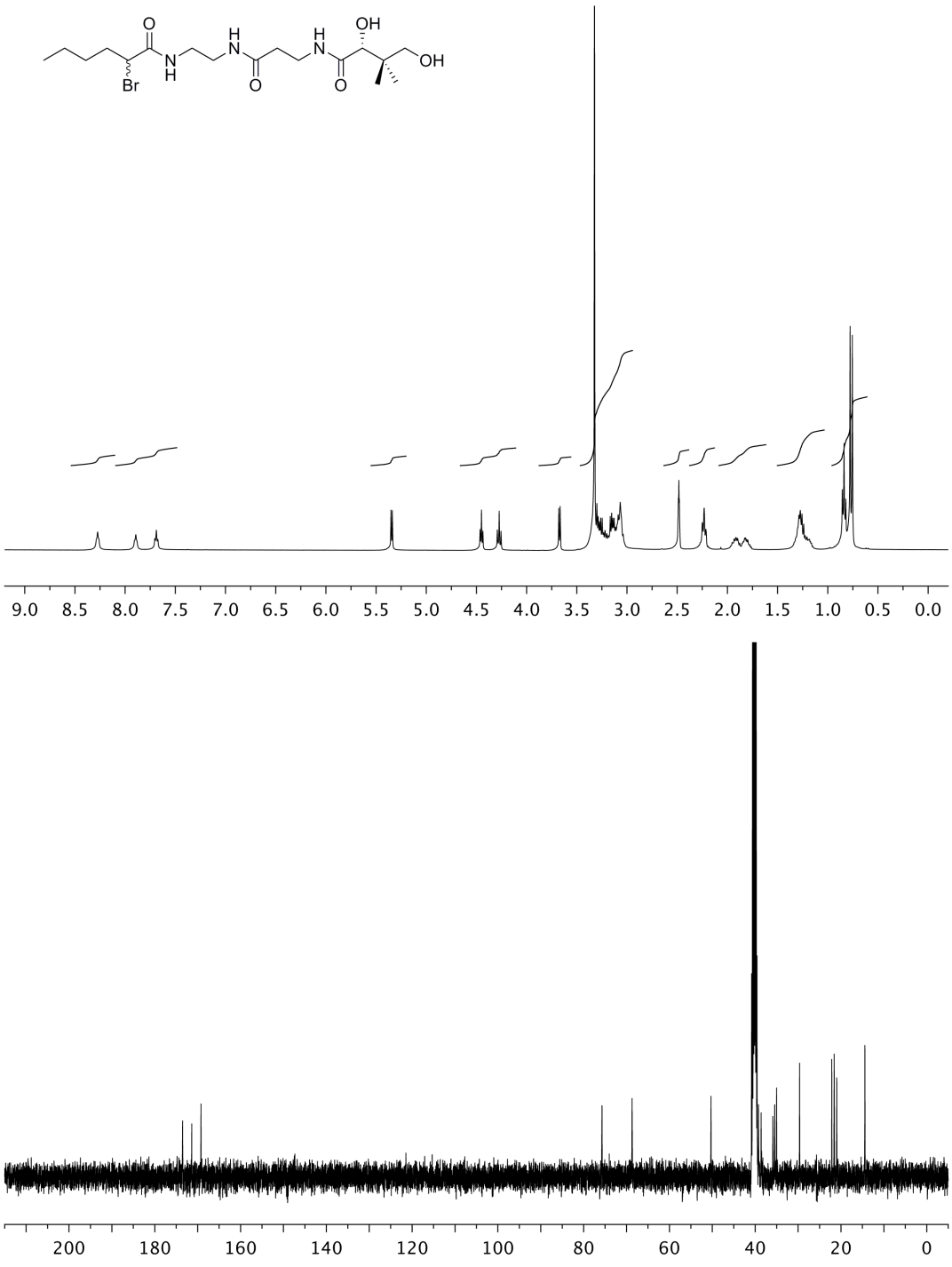


^{1}H NMR (400 Mhz) and ^{13}C NMR (100 Mhz) of probe **S7** in $CDCl_3$





^{1}H NMR (400 Mhz) and ^{13}C NMR (100 Mhz) of analog **6** in DMSO-d_6



References

1. Chisti Y (2007) Biodiesel from microalgae. Biotechnology Advances 25: 294-306.

2. Scott SA, Davey MP, Dennis JS, Horst I, Howe CJ, Lea-Smith DJ, Smith AG (2010) Biodiesel from algae: challenges and prospects. Current Opinion in Biotechnology 21: 277-286.

3. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. Plant Journal 54: 621-639.

4. Durrett TP, Benning C, Ohlrogge J (2008) Plant triacylglycerols as feedstocks for the production of biofuels. Plant Journal 54: 593-607.

5. Mata TM, Martins AA, Caetano NS (2010) Microalgae for biodiesel production and other applications: A review. Renewable and Sustainable Energy Reviews 14: 217-232.

6. Yu WL, Ansari W, Schoepp NG, Hannon MJ, Mayfield SP, Burkart MD (2011) Modifications of the metabolic pathways of lipid and triacylglycerol production in microalgae. Microbial Cell Factories 10: 1-11.

7. Thelen JJ, Ohlrogge JB (2002) Metabolic engineering of fatty acid biosynthesis in plants. Metabolic Engineering 4: 12-21.

8. Voelker TA, Worrell AC, Anderson L, Bleibaum J, Fan C, Hawkins DJ, Radke SE, Davies HM (1992) Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. Science 257: 72-74.

9. Salas JJ, Ohlrogge JB (2002) Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. Archives of Biochemistry and Biophysics 403: 25-34.

10. Jones A, Davies HM, Voelker TA (1995) Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. Plant Cell 7: 359-371.

11. Davies HM (1993) Medium chain acyl-ACP hydrolysis activities of developing oilseeds. Phytochemistry 33: 1353-1356.

12. Radakovits R, Eduafo PM, Posewitz MC (2011) Genetic engineering of fatty acid chain length in *Phaeodactylum tricornutum*. Metabolic Engineering 13: 89-95.

13. Liu X, Sheng J, Curtiss III R (2011) Fatty acid production in genetically modified cyanobacteria. Proceedings of the National Academy of Sciences, USA 108: 6899-6904.

14. Ryall K, Harper JT, Keeling PJ (2003) Plastid-derived type II fatty acid biosynthetic enzymes in chromists. Gene 313: 139-148.

15. Zornetzer GA, Tanem J, Fox BG, Markley JL (2010) The length of the bound fatty acid influences the dynamics of the acyl carrier protein and the stability of the thioester bond. Biochemistry 49: 470-477.

16. Weissman KJ, Hong H, Popovic B, Meersman F (2006) Evidence for a protein-protein interaction motif on an acyl carrier protein domain from a modular polyketide synthase. Chemistry & Biology 13: 625-636.

17. Meier JL, Burkart MD (2009) The chemical biology of modular biosynthetic enzymes. Chemical Society Reviews 38: 2012-2045.

18. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, et al. (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318: 245-251.

19. Walker TL, Collet C, Purton S (2005) Algal transgenics in the genomic ERA. Journal of Phycology 41: 1077-1093.

20. Specht E, Miyake-Stoner S, Mayfield S (2010) Micro-algae come of age as a platform for recombinant protein production. Biotechnology Letters 32: 1373-1383.

21. Mayfield SP, Franklin SE, Lerner RA (2003) Expression and assembly of a fully active antibody in algae. Proceedings of the National Academy of Sciences, USA 100: 438-442.

22. Weissman KJ, Mueller R (2008) Protein-protein interactions in multienzyme megasynthases. ChemBioChem 9: 826-848.

23. Jones S, Thornton JM (1996) Principles of protein-protein interactions. Proceedings of the National Academy of Sciences, USA 93: 13-20.

24. Miernyk JA, Thelen JJ (2008) Biochemical approaches for discovering protein-protein interactions. Plant Journal 53: 597-609.

25. Worthington AS, Hur GH, Meier JL, Cheng Q, Moore BS, Burkart MD (2008) Probing the compatibility of type II ketosynthase-carrier protein partners. ChemBioChem 9: 2096-2103.

26. Worthington AS, Rivera H, Jr., Torpey JW, Alexander MD, Burkart MD (2006) Mechanism-based protein crosslinking probes to investigate carrier protein-mediated biosynthesis. ACS Chemical Biology 1: 687-691.

27. Cantu DC, Chen Y, Reilly PJ (2010) Thioesterases: A new perspective based on their primary and tertiary structures. Protein Science 19: 1281-1295.

28. Facciotti MT, Yuan L (1998) Molecular dissection of the plant acyl-acyl carrier protein thioesterases. Fett - Lipid 100: 167-172.

29. Dani KGS, Hatti KS, Ravikumar P, Kush A (2011) Structural and functional analyses of a saturated acyl ACP thioesterase, type B from immature seed tissue of *Jatropha curcas*. Plant Biology 13: 453-461.

30. Mayer KM, Shanklin J (2005) A structural model of the plant acyl-acyl carrier protein thioesterase FatB comprises two helix/4-stranded sheet domains, the N-terminal domain containing residues that affect specificity and the C-terminal domain containing catalytic residues. Journal of Biological Chemistry 280: 3621-3627.

31. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function prediction. Nature Protocols 5: 725-738.

32. Chan DI, Vogel HJ (2010) Current understanding of fatty acid biosynthesis and the acyl carrier protein. Biochemical Journal 430: 1-19.

33. Zhang YM, Rao MS, Heath RJ, Price AC, Olson AJ, et al. (2001) Identification and Analysis of the Acyl Carrier Protein (ACP) Docking Site on β-Ketoacyl-ACP Synthase III. Journal of Biological Chemistry 276: 8231-8238.

34. Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: An automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 20: 45-50.

35. Schaeffer ML, Agnihotri G, Kallender H, Brennan PJ, Lonsdale JT (2001) Expression, purification, and characterization of the Mycobacterium tuberculosis acyl carrier protein, AcpM. Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids 1532: 67-78.

36. Murugan E, Kong R, Sun H, Rao F, Liang Z-X (2010) Expression, purification and characterization of the acyl carrier protein phosphodiesterase from *Pseudomonas aeruginosa*. Protein Expression and Purification 71: 132-138.

37. Worthington AS, Burkart MD (2006) One-pot chemo-enzymatic synthesis of reporter-modified proteins. Organic & Biomolecular Chemistry 4: 44-46.

38. Quadri LEN, Weinreb PH, Lei M, Nakano MM, Zuber P, Walsh CT (1998) Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. Biochemistry 37: 1585-1595.

39. La Clair JJ, Foley TL, Schegg TR, Regan CM, Burkart MD (2004) Manipulation of carrier proteins in antibiotic biosynthesis. Chemistry & Biology 11: 195-201.

40. Meier JL, Barrows-Yano T, Foley TL, Wike CL, Burkart MD (2008) The unusual macrocycle forming thioesterase of mycolactone. Molecular BioSystems 4: 663-671.

41. Worthington AS, Porter DF, Burkart MD (2010) Mechanism-based crosslinking as a gauge for functional interaction of modular synthases. Organic & Biomolecular Chemistry 8: 1769-1772.

42. Mendez M, O'Neill B, Burkart MD, Behnke C, Lieberman S, Bielinski V, Poon Y (2010) Production of fatty acids by genetically modified photosynthetic organisms. US Patent No. WO 2010/019813.

43. Fujita Y, Matsuoka H, Hirooka K (2007) Regulation of fatty acid metabolism in bacteria. Molecular Microbiology 66: 829-839.

44. James GO, Hocart CH, Hillier W, Chen H, Kordbacheh F, Price GD, Djordjevic MA (2011) Fatty acid profiling of *Chlamydomonas reinhardtii* under nitrogen deprivation. Bioresource Technology 102: 3343-3351.

45. Dormann P, Voelker TA, Ohlrogge JB (1995) Cloning and expression in *Escherichia coli* of a novel thioesterase from *Arabidopsis thaliana* specific for long-chain acyl-acyl-carrier proteins. Archives of Biochemistry and Biophysics 316: 612-618.

46. Gong Y, Guo X, Wan X, Liang Z, Jiang M (2011) Characterization of a novel thioesterase (PtTE) from *Phaeodactylum tricornutum*. Journal of Basic Microbiology 51: 666-672.

47. Harris EH (1989) The Chlamydomonas sourcebook: A comprehensive guide to biology and laboratory use. Academic Press: San Diego, CA.

48. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22: 195-201.

49. PyMol (v1.4) Schrödinger, LLC.

50. Morris GM, Goodsell DS, Huey R, Olson AJ (1996) Distributed automated docking of flexible ligands to proteins: Parallel applications of AutoDock 2.4. Journal of Computer-Aided Molecular Design 10: 293-304.

51. Armougom F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, Kedua V, Notredame C (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-coffee. Nucleic Acids Research 34: W604-W608.

52. Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15: 305-308.

53. Emanuelsson O, Nielsen H, Von Heijne G (1999) ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. Protein Science 8: 978-984.

54. Nielsen H, Engelbrecht J, Brunak S, vonHeijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Engineering 10: 1-6.

55. Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM, Jones AR, Randolph-Anderson BL, Robertson D, Klein TM, Shark KB, et al. (1988) Chloroplast transformation in Chlamydomonas with high velocity microprojectiles. Science 240: 1534-1538.

56. Gorman DS, Levine RP (1965) Cytochrome F and plastocyanin - their sequence in photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. Proceedings of the National Academy of Sciences, USA 54: 1665-1669.

57. Kan Y, Pan J (2010) A one-shot solution to bacterial and fungal contamination in the green alga *Chlamydomonas reinhardtii* culture by using an antibiotic cocktail. Journal of Phycology 46: 1356-1358.

58. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research 36: W465-W469.

59. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A (2005) H++: a server for estimating pK(a)s and adding missing hydrogens to macromolecules. Nucleic Acids Research 33: W368-W371.

60. Schuttelkopf AW, van Aalten DMF (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. Acta Crystallographica Section D 60: 1355-1363.

61. Sanner MF (1999) Python: A programming language for software integration and development. Journal of Molecular Graphics & Modelling 17: 57-61.

62. Claros MG, Vincens P (1996) Computational method to predict mitochondrially imported proteins and their targeting sequences. European Journal of Biochemistry 241: 779-786.

63. Small I, Peeters N, Legeai F, Lurin C (2004) Predotar: A tool for rapidly screening proteomes for N-terminal targeting sequences. Proteomics 4: 1581-1590.

64. Peeters N, Small I (2001) Dual targeting to mitochondria and chloroplasts. Biochimica et Biophysica Acta 1541: 54-63.

65. Dormann P, Frentzen M, Ohlrogge JB (1994) Specificities of the acyl-acyl-carrier protein (ACP) thioesterase and glycerol-3-phosphate acyl transferase for octadecenoyl-ACP isomers - identification of a petroselinoyl-ACP thioesterase in Umbelliferae. Plant Physiology 104: 839-844.

66. Araki S, Sakurai T, Kawaguchi A, Murata N (1987) Positional distribution of fatty acids in glycerolipids of the marine red alga *Porphyra yezoensis*. Plant and Cell Physiology 28: 761-766.

Chapter 2 is adapted from a recently submitted article to Plos One, on which I was the primary author. Joris Beld assisted with the *in silico* modeling and *in vitro* crosslinking experiments, including protein expression and purification and synthesis of chemical probes. Craig Behnke and Mike Mendez assisted with engineering of TEs into the *C. reinhardtii* chloroplast. Yongzuan Su (UCSD Mass Spectrometry Facility) helped with mass spectrometric analsysis.

Chapter 3 : Evolution of fatty acid synthase enzymes dictates interchangeability

Abstract

Microalgae naturally accumulate large quantities of oil and other energy-dense metabolites, and as such, they have emerged as a feedstock for renewable biofuels and coproducts. Engineering algal fatty acid biosynthesis promises to create strains capable of producing fungible fuels on scale in a sustainable fashion. However, recent engineering attempts have highlighted both a lack of fundamental understanding of algal fatty acid biosynthesis and the interactions between catalytic domains. As the acyl carrier protein (ACP) must catalytically interact with all fatty acid synthase (FAS) enzymes in the biosynthesis of a fatty acid, for successful engineering of FAS, one must consider protein-protein interactions between the target FAS enzyme and the ACP of the host organism.

In recent work, we demonstrated that protein-protein interactions between the ACP and acyl-ACP thioesterase mediate the hydrolysis of fatty acids in green algae, and we have also shown that ACP-KS interactions are required in bacterial fatty acid biosynthesis. Here, we investigate the compatibility between FAS enzymes and ACPs across species using phylogenetic analyses, *in silico* protein-protein docking, *in vitro* activity-based crosslinking between FAS enzymes and ACPs, and *in vivo* engineering of FAS enzymes. We propose that ACP-KS interactions are permissive across species and that ACP-TE interactions are selective. This hypothesis is based on our analysis of the evolutionary origin of each distinct FAS enzyme. Our findings will help predict successful protein-protein interactions between FAS enzymes and ACPs to enable the metabolic engineering of FAS in any species.

Introduction

Petroleum products are ancient deposits of photosynthetic algae transformed over time by pressure and temperature derived mostly from algal fatty acids [1]. As fatty acids represent the most abundant source of reduced carbon in Nature, they are currently used in the synthesis of biodiesel, a renewable alternative to petroleum [2].

Microalgae were first explored as a potential biodiesel feedstock during the Aquatic Species Program (ASP), funded from 1978-1996 in the United States to investigate oil-accumulating microorganisms as a source of renewable fuel [3]. However, low yields and crop protection remained major hurdles in the field [3]. Oleaginous microalgae are now reemerging as a biodiesel feedstock because of their innate ability to efficiently capture solar radiation and store it in energy-dense lipid molecules [4], while growing at fast rates in a broad range of environments [5]. However, although oleaginous species accumulate lipids, the fatty acid biosynthetic pathway in wildtype microalgae is not optimized for fatty acid synthesis, as fatty acids only represent a fraction of necessary biomolecules within a healthy algae cell (i.e. proteins, carbohydrates and nucleic acids). Additionally, fatty acids present in algal lipids are not of ideal length or saturation for biodiesel composition [6]. Metabolic engineering could facilitate the creation of algal strains that produce high levels of lipids containing fatty acids with select characteristics for biodiesel fuel [7]. Fatty acid biosynthetic enzymes have been engineered into heterologous microalgae with the goal of altering the fatty acid chain length [8] [9] and increasing oil content [10], and the limited success of these efforts highlights the lack of understanding of algal fatty acid synthase (FAS) and compatibility of FAS domains across species.

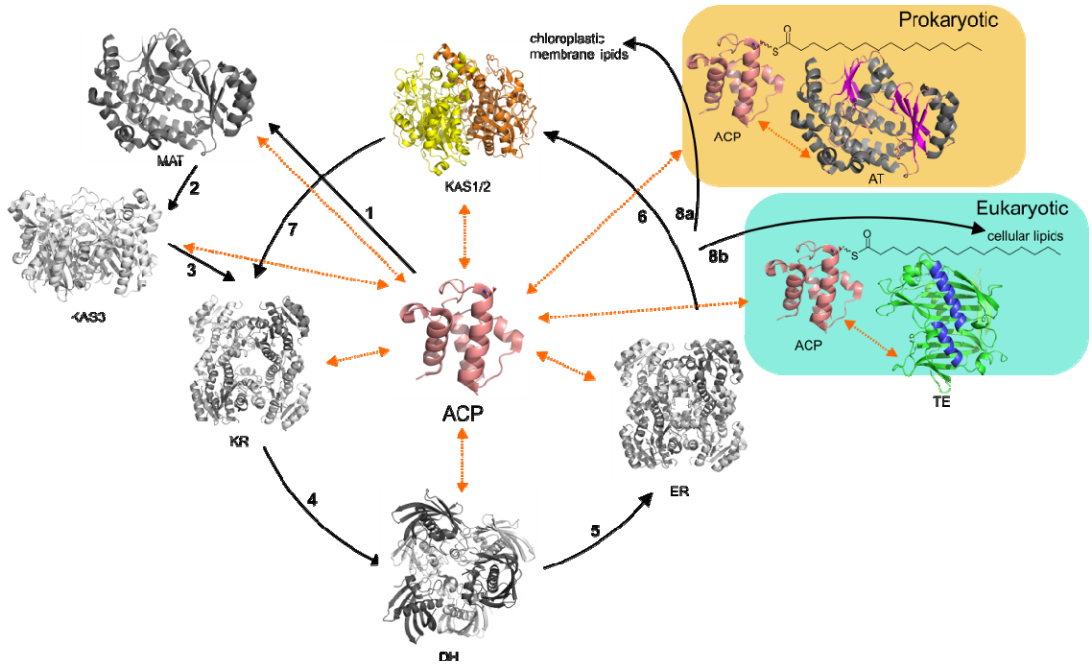


Figure 3-1 Type II fatty acid biosynthesis in prokaryotes and eukaryotes

(1) Fatty acid biosynthesis is initiated by the formation of malonyl-ACP from malonyl-CoA and *holo*-ACP by malonyl-acyltransferase (MAT, FabD). (2) β-ketoacyl ACP synthase III (KSIII, KAS3 or FabH) is responsible for the initial condensation of malonyl-ACP with acetyl-CoA. (3) β-ketoacyl ACP ketoreductase (KR, FabG) reduces the keto group and (4) β-ketoacyl ACP dehydratase (DH, FabA/FabZ) eliminates water. (5) β-ketoacyl ACP enoylreductase (ER, FabI) reduces the double bond using NADH. Iterative activity of KS, KR, DH and ER domains extends the chain by two carbons each cycle. β-ketoacyl ACP synthase I (KSI, KAS1, FabB) extends the chain from C4 to C16 and β-ketoacyl ACP synthase II (KS2, KAS2, FabF) forms C18 from C16 (6 and 7). In the prokaryotic termination pathway, fatty acids are transferred from acyl-ACP to glycerol-3-phosphate for phospholipid formation using an acyl transferase (AT) (8a), and in the eukaryotic pathway, mature fatty acids are hydrolyzed from acyl-ACP by an acyl-ACP thioesterase (TE) (8b).

Amphiphilic molecules (such as fatty acids) gave rise to the phenomenon of compartmentalization early in evolution, their structure facilitating micelle formation to effectively shield the inner cellular environment from water [11] [12] [13] [14] [15]. In primitive organisms, such as archaea and bacteria, as well as in eukaryotes such as algae and plants, fatty acids are biosynthesized by a group of dissociated catalytic

enzymes called a type II fatty acid synthase (FAS). This is in contrast to the type I FAS of higher eukaryotes, where all catalytic domains are fused to a single protein. Essential to FAS is an acyl carrier protein (ACP), which tethers the growing metabolite to a flexible 4'-phosphopantetheiene arm as it functionally interacts with each of the seven catalytic domains to elongate the fatty acid in cycles of two carbon units (Figure 3-1). Importantly, ACP buries the fatty acid in its hydrophobic core to protect it from hydrolysis throughout the biosynthetic process [16]. Protein-protein recognition with catalytic FAS domains triggers a conformational change in ACP that elicits the acyl chain from the pocket for iterative cycles of elongation and reduction [17]. Thus, protein-protein interactions, in addition to enzyme-substrate interactions, are essential in fatty acid biosynthesis.

A key distinction between prokaryotic and eukaryotic type II FAS is the evolution of an alternate fatty acid biosynthetic termination pathway in eukaryotes (Figure 3-1) [18] [19]. In addition to the prokaryotic termination pathway utilizing acyl transferase (AT) domains, which transfer fatty acids directly from acyl-ACP to glycerol-3-phosphate for membrane formation, early eukaryotes, such as microalgae, contain a thioesterase (TE) domain that catalyzes the hydrolysis of fatty acids from acyl-ACP, effectively releasing free fatty acids and allowing their export from the plastid for incorporation into cellular lipids [19]. Primitive TEs, such as those present in microalgae, are rather promiscuous in the acyl-ACP substrates they act upon [9]. For example, in recent work, we were able to show that the sole fatty acid TE (CrTE, a 'Fat1' TE) detected in the genome of model green alga *Chlamydomonas reinhardtii* [20] hydrolyzes a range of saturated and unsaturated substrates [9]. Additionally, a

single TE was found in the genome of *Phaeodactylum tricornutum* (Pt), also distinct from known plant and bacterial TEs, and its overexpression increased amounts of 16:0, 16:1, 18:0 and 18:1 fatty acids [21], indicating promiscuity. Vascular plants, on the other hand, have evolved much more specific TEs with high substrate specificity. Plant FatA TEs hydrolyze 18:1-ACPs and plant FatB TEs act upon a range of saturated substrates ranging from 8-18 carbons [22]. Additionally, the seeds of certain plant species, such as *Cuphea hookeriana* [23], contain a redundant fatty acid biosynthetic pathway for the production of medium chain (C8-C14) fatty acids and contain TEs with specificities corresponding to the medium chain phenotype [24].

As the eukaryotic fatty acid biosynthetic pathway is terminated by TE-mediated hydrolysis of a mature fatty acid and release from the synthase, the TE functionally determines the identity of the fatty acid end product [25]. In efforts to manipulate the fatty acid content of oil crops to create 'designer oils', highly specific plant TEs have been engineered into several heterologous plant species effectively altering their fatty acid composition [26]. Due to the relatively high homology between plants and algae, it seems plausible that plant TEs could also alter the fatty acid profile when engineered into algae. However, in recent attempts aimed at altering the fatty acid chain length in algae, specific plant TEs were introduced into green algae (*Chlamydomonas reinhardtii*) and diatoms (*Phaeodactylum tricornutum*), and these efforts were only marginally successful [9] [8]. Work in our laboratory has demonstrated that protein-protein interactions between the TE and ACP are required in fatty acid hydrolysis [9]. Thus, when engineered into the algal chloroplast, it is likely

that plant TEs do not functionally interact with the algal ACP and are unable to siphon fatty acids.

Whereas plant TEs did not significantly alter the algal fatty acid composition, when engineered into bacteria (*Escherichia coli*) and cyanobacteria (*Synechoccus elongatus*), an altered fatty acid content was observed [27] [28] [29] [30] [31] [32] [33] [34]. In contrast to algae and plants, aerobic bacteria and cyanobacteria do not have dedicated fatty acid TEs [35]. Possible scenarios for the observed phenomenon that plant TEs can function to alter the bacterial fatty acid profile *in vivo* are that positive bacterial ACP-plant TE interactions occur [36] [37] [17] [9], the plant TEs recognize the acyl chain of the acyl-ACP substrate [38], or the plant TEs cause a disruption to bacterial FAS regulation [39] [40]. The evolutionary origin of the algal/plant TE is unknown. The algal/plant plastid, where fatty acid biosynthesis occurs, evolved through a endosymbiotic event whereby a unicellular protist (host) engulfed a free-living cyanobacterium [41]. The type II FAS present in algal/plant plastids is derived from ancient cyanobacterial FAS [42], and as cyanobacteria produce mostly saturated fatty acids, it has been postulated that plant TEs are descendents of an ancient plastidial TE of the FatB type [43]. However, in this work, we propose that the algal/plant TE evolved from the lineage of the eukaryotic host.

The ketoacyl synthases (KSs) mediate the Claisen condensation reaction that elongates a growing fatty acid by two carbon units upon each catalytic cycle (Figure 3-1). KS domains catalyze carbon-carbon bond formation and are therefore essential in the biosynthesis of fatty acids. KSIII initiates fatty acid biosynthesis by condensing malonyl-ACP with acetyl-CoA. KSI elongates a growing fatty acid from 6-16

carbons, and KSII forms stearoyl-ACP from palmitoyl-ACP. In contrast to TE domains, there is clear phylogenetic evidence that reveals a common ancestor of all KSs, an archaeal thiolase [44] (Figure 3-2). Over time, elongation enzymes (KSI and II) and initiation enzymes (KSIII, KCS, and CHS) developed from this predecessor, an example of divergent evolution [44]. Chalcone synthases (CHS), elongases and thiolases utilize a very similar mechanism and have similar folds [45]. In this superfamily of enzymes, the nucleophilic α-anion of an acylthioester (which form a C-C bond by attacking the electrophilic carbonyl of another molecule) is either generated in a non-decarboxylative or decarboxylative fashion – the latter a more recent optimization of this enzyme class.

Thiolases, from which KSs are derived [44] (Figure 3-2), catalyze both forward and reverse reactions. Thiolase I (3-ketoacyl-CoA thiolase) catalyzes the removal of an acetyl group from acyl-CoA in fatty acid beta-oxidation, and thiolase II (also known as acetyl-CoA transferase) catalyzes the reverse reaction, condensing two molecules of acetyl-CoA together. Thiolases are non-decarboxylative and contain an C(N/H)HC active site, whereas KS enzymes are decarboxylative and use CHH (I and II) or CHN (III) for catalysis. Some algal KS domains have been annotated in genomes of sequenced microalgae and there appears to be a high degree of homology of algal KSs to known bacterial and plant KSs [46].

Engineering fatty acid biosynthetic enzymes into heterologous organisms has illuminated an apparent directionality in the ability of FAS enzymes to interact across species. We pose that the evolution of FAS enzymes dictates their interchangeability. In particular, we hypothesize that ACP-KS interactions are permissive across

bacterial, algal and plant species due to common ancestry, and that the ACP-TE interaction is more selective. Here, we use phylogenetic analyses, *in silico* modeling of protein-protein interactions, *in vitro* activity-based crosslinking as a screen for protein-protein interactions, and *in vivo* engineering of FAS enzymes to illustrate compatibility between partner FAS enzymes and ACP. We focus on the ACP, KS, and TE domains of bacterial, algal, and plant type II fatty acid synthase.

Results

Sequence alignment and comparison of bacterial, algal, and plant ACPs

Structure-based sequence analysis of bacterial, algal, and plant ACPs shows that the α-helical secondary structure, as well as the serine residue which serves as the site of post-translational modification, is completely conserved across species (Figure 3-17).

Sequence alignment of bacterial, algal, and plant KSs

Sequence alignment of bacterial, algal, and plant KSs (KSI, II, III) illustrates a high degree of sequence conservation between KS types across species (Figure 3-8, Figure 3-9, Figure 3-10) [46]. Due to the high primary sequence variability between thiolases and KSs, the active site residues do not align, but they all contain a CHN, CHH, or CHC catalytic triad [44].

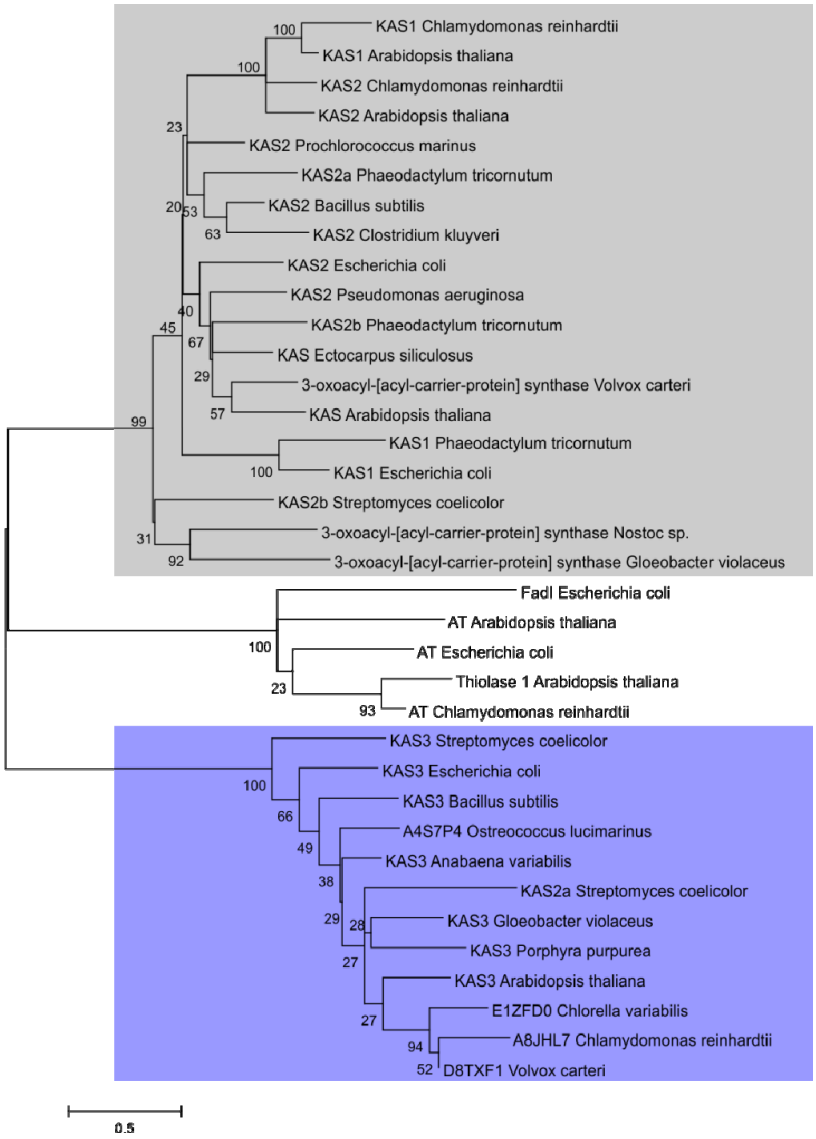


Figure 3-2 Ketosynthase phylogeny

Detailed phylogenetic analysis of the thiolase superfamily was conducted with sequences obtained from the UniprotKB database using Bayesian inference [75]. Sequences were aligned with MUSCLE (3.8.31) [72] and Bayesian trees constructed using MrBayes (3.2.0) [75].

KS Phylogeny

Phylogenetic analysis reveals clustering of KSIs, KSIIs, and KSIIIs in all organisms interrogated, indicating KS domains are similar across species (Figure 3-2, Figure 3-11). We began by assembling the primary sequences of ketoacyl synthases and thiolases from 10 algae, 10 plants, 10 bacteria, 10 cyanobacteria and 10 archaea (*SI*). Although KSIII is not annotated in any algal genome, BLAST analysis of known plant KSIIIs against sequenced algal genomes identifies putative KSIIIs in algae (*SI*). Constructing a phylogenetic tree of KS domains from bacteria, plants and algae clearly illustrates a division between thiolases, KSI/II, and KSIII enzymes in all species (Figure 3-2, Figure 3-11). Our data is consistent with the rooted phylogenetic tree constructed by Jiang et al. [44], and the putative algal KSs we identified herein fit very well within this thiolase ancestry (Figure 3-2).

Sequence alignment of bacterial, algal, and plant TEs

To offer insight into the origin of the eukaryotic TE, we first conducted a BLAST analysis of all known TEs in algae, plants, and bacteria (*SI*). We readily identified three examples of TEs present in anaerobic bacteria (*Desulfovibrio africanus*, *Desulfovibrio desulfuricans*, and *Desulfovibrio aespoeensis*), but were unable to locate TE sequences in any cyanobacterial genome (*SI*). Amino acid sequence alignment of TEs from the three identified anaerobic bacteria, along with TEs from sequenced algal species and characterized plant TEs, illustrates that the TEs from anaerobic bacteria contain many regions of sequence conservation with algal TEs and highly-specific plant TEs (Figure 3-12, Figure 3-3). The active site residues of all

representative TEs are identical (His and Asn), with the exception that the TEs from anaerobic bacteria do not seem to contain a catalytic Cys residue (Figure 3-12).

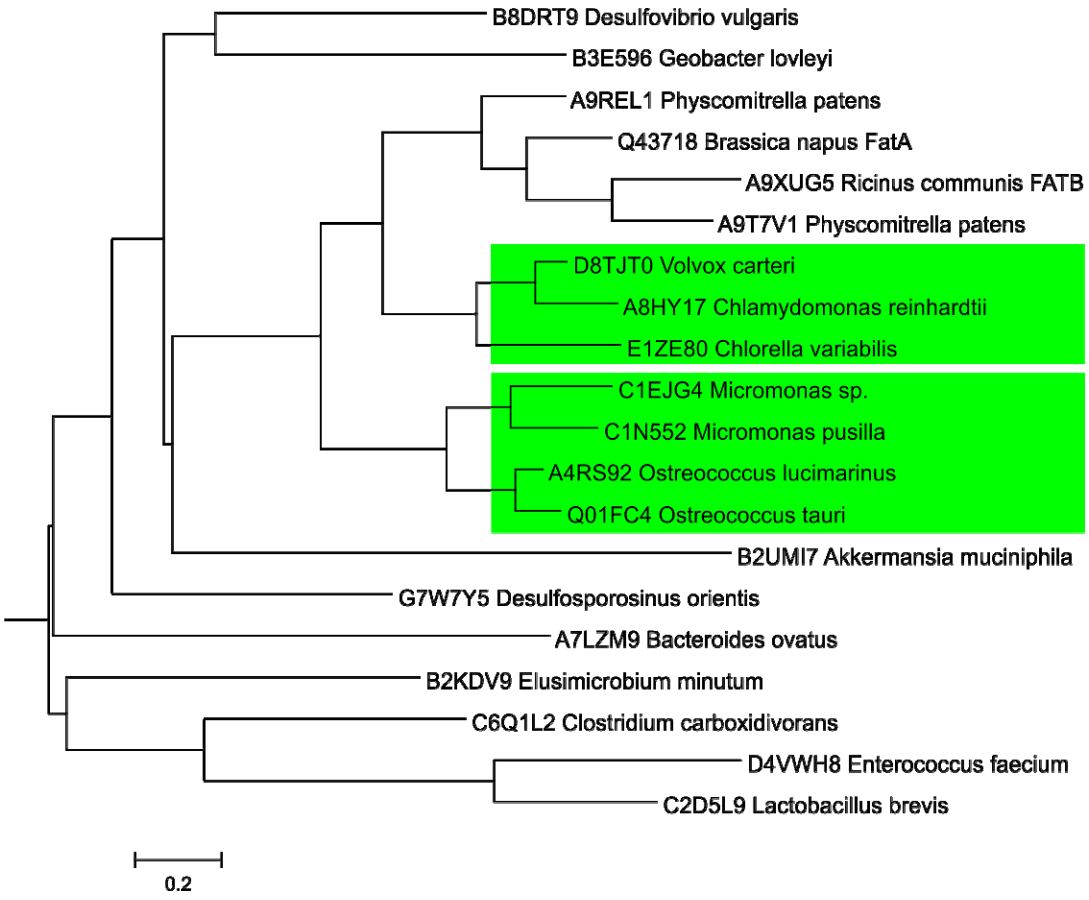


Figure 3-3 Thioesterase phylogeny

Detailed phylogenetic analysis of the acyl-ACP thioesterase family was conducted with sequences obtained from the UniprotKB database using Bayesian inference [75]. Sequences were aligned with MUSCLE (3.8.31) [72] and Bayesian trees constructed using MrBayes (3.2.0) [75].

TE Phylogeny

To validate our hypothesis that primitive TEs from anaerobic bacteria are the ancestors of algal and plant TEs, we constructed a rooted phylogenetic tree comprised of TE sequences from anaerobic bacteria, TEs from all sequenced algae, and plant

FatA and FatB TEs (Figure 3-3, Figure 3-13). As shown in Figure 3-3, there is a clear divergence over time from anaerobic bacterial TEs to algal Fat1 TEs to plant FatA and FatB TEs. These results suggest that TEs originated in anaerobic bacteria, and not the cyanobacterial symbiont as previously thought [43]. This argument is further strengthened by the absence of fatty acid TE sequences in bacterial and cyanobacterial genomes.

Our constructed phylogenetic tree comprised of TEs from various species depicts clustering of different groups of TEs (Figure 3-3, Figure 3-13). Plant FatA TEs, plant FatB TEs, algal Fat1 TEs, and prokaryotic TEs all exist as separate clades (Figure 3-3). Interestingly, there are two separate clades of algal Fat1 TEs (Figure 3-3), which appear to correlate with distinct algal fatty acid profiles (Figure 3-14). Additionally, we observe a eukaryotic TE between the algal Fat1 TEs and the plant FatA and FatB TEs in the moss *Physcomitrella patens*, also in a distinct clade (Figure 3-3, Figure 3-13).

Following a thorough evolutionary analyses of KS and TE domains, we next sought to investigate interspecies ACP-TE and ACP-KS interactions *in silico*, *in vitro*, and *in vivo*.

In silico modeling of ACP-KS and ACP-TE protein-protein interactions

As an initial screen for protein-protein interactions, we generated homology model structures of ACPs, KSs, and TEs and conducted a virtual docking experiment between ACPs and FAS domains [9]. A bacterial ACP (*Escherichia coli*) was docked to representative KSIIs from a bacterium (*Escherichia coli*), an alga (*Chlamydomonas reinhardtii*), and a plant (*Ricinus communis*). The algal KSII and plant KSII docked

extremely well with the bacterial ACP (Figure 3-4, Figure 3-18), similar to the native bacterial ACP-KS interaction, suggesting a productive interaction between the two proteins. Extracting the EcACP from its crystal structure in which it is bound to a partner enzyme (PDB ID: 3EJB) prior to docking the ACP structure to various KS domains shows an even more favorable *in silico* ACP-KS interaction across species (Figure 3-19).

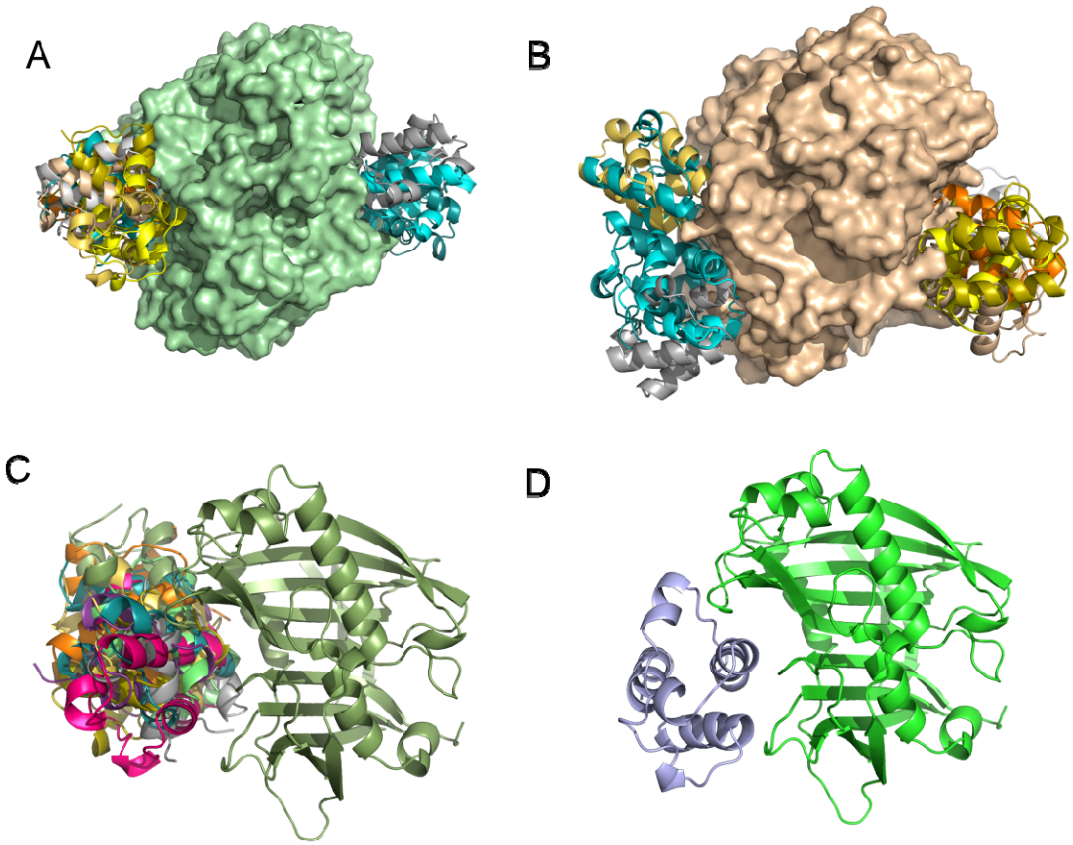


Figure 3-4 In silico docking of ACP-KS and ACP-TE interactions across species

A) EcACP-CrKSII, B) CrACP-EcKSII, C) EcACP-CrTE, D) CrACP-CrTE

Docking the algal ACP (Cr) to the bacterial KSII (Ec) (Figure 3-4, Figure 3-20) shows favorable *in silico* binding. Additionally, the plant ACP (Rc) docks well to the

algal KSII (Cr) (Figure 3-21). In the reverse evolutionary direction, the plant KSII (Rc) binds favorably with the algal ACP (Cr) model (Figure 3-7, Figure 3-20). Taken together, this virtual screen suggests positive protein-protein interactivity between ACPs and KSs across species. In contrast to the KS domain, where we propose that ACP-KS interactions are permissive across species as there is high sequence conservation between evolutionarily diverse species, we hypothesize that the ACP-TE interaction is more selective and unidirectional. We have previously shown that plant TEs (*Umbellularia californica* and *Cuphea hookeriana*) do not functionally interact with algal ACPs (*C. reinhardtii*) *in vitro* and *in vivo* [9]. However, plant TEs are functional *in vivo* in TE-less bacteria in cyanobacteria [33]. Docking EcACP to CrTE (Figure 3-4) shows a favorable binding orientation between the two domains, comparable to, but not as strong as the native CrACP-CrTE *in silico* interaction (Figure 3-22).

In vitro crosslinking assay between ACPs and FAS enzymes

To screen for compatibility between ACPs and FAS domains across several prokaryotic and eukaryotic species, we used an *in vitro* crosslinking assay developed in our laboratory [47] (Figure 3-5). Site-specific pantetheine analogue probes are designed to trap the transient interactions occurring between the FAS domain of interrogation and ACP [47]. This assay is a highly selective *in vitro* screen for functional protein-protein interactions with ACP, and kinetic and thermodynamic studies validate that activity-based crosslinking serves as a quantitative analysis of the strength of functional binding between partner FAS enzymes and ACP [48].

Based on sequence alignment, phylogenetic analyses of KS domains, and ACP-KS virtual docking, it seems likely that ACPs and KS domains are interchangeable across species. To test this hypothesis, we first examined whether the algal ACP could functionally interact with a bacterial KS. We previously synthesized and used chloroacrylamide pantetheine analogue **1** to study ACP-KS interactions in bacterial FAS [47]. Thus, the chloroplastic ACP from model green alga *Chlamydomonas reinhardtii* was loaded with activity-based KS probe **1** and further incubated with the KSII from *Escherichia coli* (Figure 3-5). Indeed, functional compatibility between the algal ACP and bacterial KSII was observed (Figure 3-5). Importantly, the high yield of this ACP-KS formation was the same as observed for *E. coli* ACP with *E. coli* KSII (Figure 3-5). This demonstrates that the algal ACP is able to interact with the bacterial KSII in a similar fashion as the bacterial ACP *in vitro*.

We next sought to examine whether the algal/plant TEs interact with the *E. coli* ACP, or whether the observed *in vivo* activity in bacteria and cyanobacteria is simply the case of substrate recognition. To test whether eukaryotic TEs functionally interact with bacterial ACPs, the *E. coli* ACP was loaded with activity-based TE probes **2** and **3** (Figure 3-5), and *crypto*-EcACPs were incubated with the Fat1 algal TE from *C. reinhardtii*. Indeed, protein-protein recognition between the algal TE and bacterial ACP triggers a crosslinking reaction, albeit not as high yielding as the ACP-KS interaction (Figure 3-5). Because the TE also is able to recognize the acyl chain attached to ACP, both a short chain (C6, **2**) and a long chain (C16, **3**) TE probe were used for crosslinking. In contrast to ACP-TE formation in alga *C. reinhardtii*, where a higher yield was observed with C16 analogue **3**, both probes resulted in the same yield

of crosslinked product in the EcACP-CrTE experiment (Figure 3-5), indicative of both protein-protein interactions and substrate recognition. The plant TEs did not crosslink to the EcACP using either **2** or **3** (Figure 3-29), suggesting acyl-chain recognition as the mechanism by which they alter the fatty acid profile in *E. coli.*

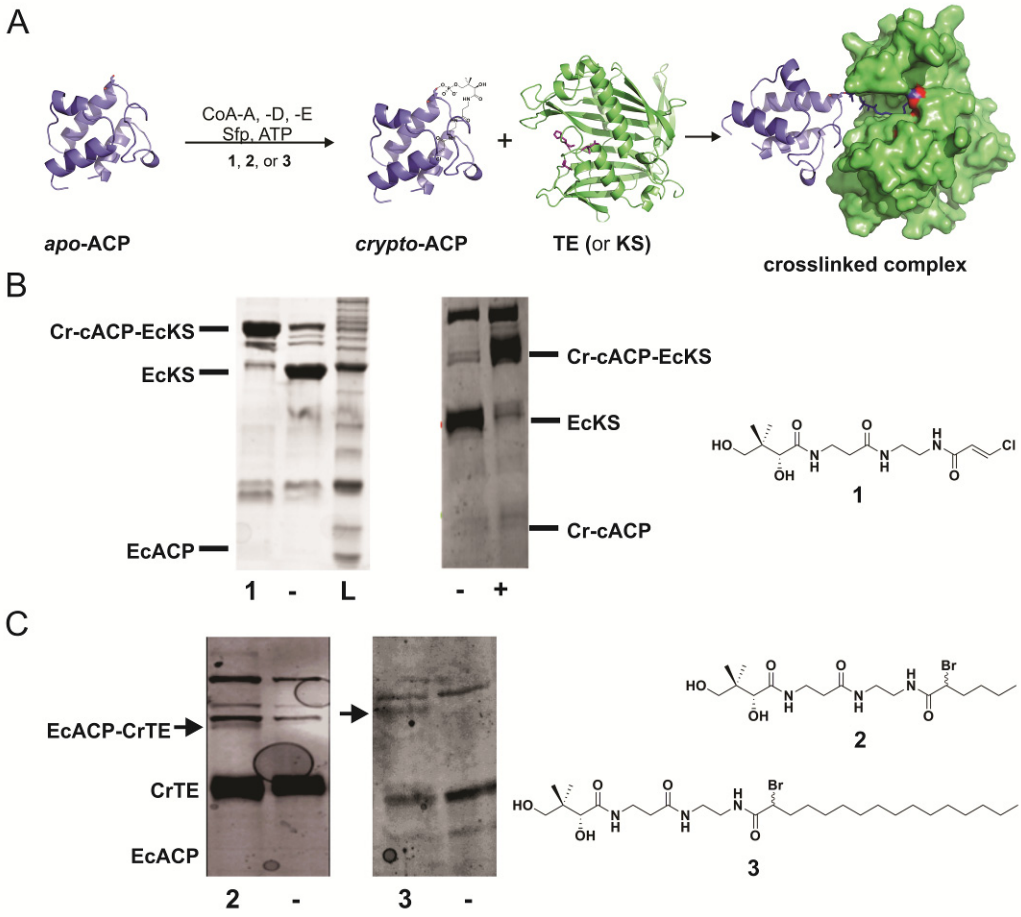


Figure 3-5 In vitro activity-based crosslinking between ACP and FAS enzymes

A) General schematic of activity-based crosslinking assay to screen for functional interactions with ACP. B) ACP-KS interactions: EcACP-EcKSII (positive control, left), Cr-cACP-EcKSII (right); C) ACP-TE interactions: EcACP-CrTE. L = Benchmark ladder.

In vivo ACP complementation assay

The results of the *in vitro* crosslinking assay between the *C. reinhardtii* ACP and *E. coli* KSII demonstrate a strong functional interaction between the two proteins,

indicating permissive ACP-KS interactions from bacteria to algae (Figure 3-5). To test whether the algal ACP is able to complement bacterial fatty acid biosynthesis and functionally interact with all seven of the bacterial FAS enzymes *in vivo*, we used an ACP complementation assay developed by Cronan et al. [49]. Whereas the ACP from gram-negative bacterium *V. harveyi* complements *E. coli* fatty acid biosynthesis both on plates and in liquid culture at all tested IPTG concentrations ranging from 1 mM to 0.01 µM, Cr-cACP did not complement *E. coli* fatty acid synthase *in vivo* (Figure 3-24).

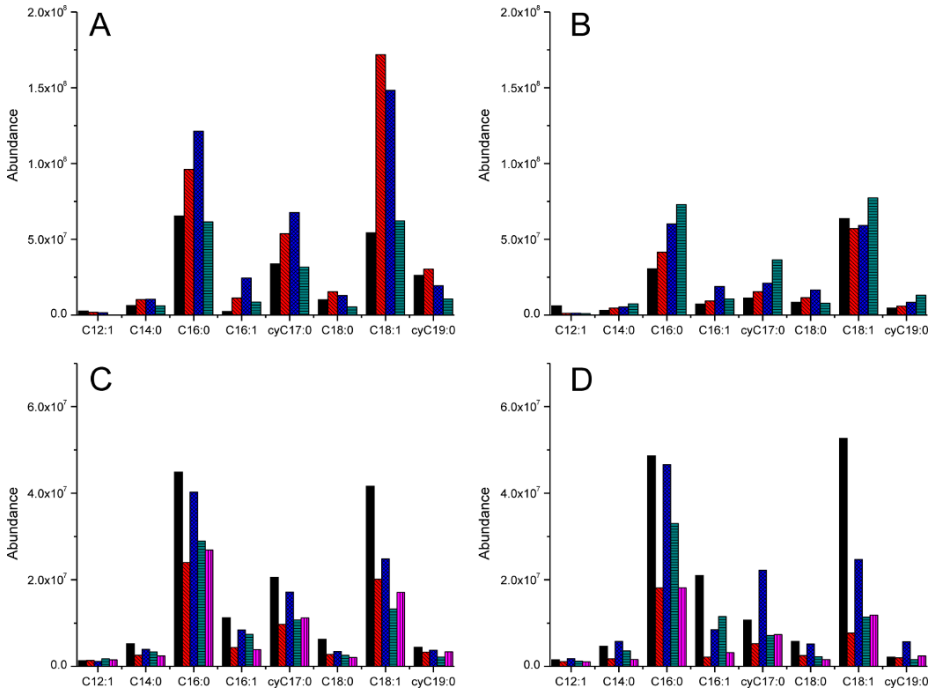


Figure 3-6 Acyl-ACP thioesterases and acyl carrier protein activity in *E. coli* strain BL21

(A) Uninduced acyl-ACP thioesterases: UcTE (solid black), ChTE (striped red), CrTE (crossed blue) and control (horizontal striped green). (B) Induced acyl-ACP thioesterases: UcTE (solid black), ChTE (striped red), CrTE (crossed blue) and control (horizontal striped green). (C) Uninduced acyl carrier proteins: Cr-cACP (solid black), Cr-mACP (striped red), Ec-ACP (crossed blue), Vh-ACP (horizontal striped green) and control (vertical striped pink). (D) Induced acyl carrier proteins: Cr-cACP (solid black), Cr-mACP (striped red), Ec-ACP (crossed blue), Vh-ACP (horizontal striped green) and control (vertical striped pink). Each data point is obtained in triplicate.

Growth of transgenic E. coli strains engineered with ACPs, KS, TEs and GC/MS analysis of fatty acid content

From the observation that the algal (Cr) ACP and bacterial (Ec) KSII demonstrate strong functional binding *in vitro* (Figure 3-5), we tested the effect of overexpressing these FAS proteins in *E. coli* on its fatty acid profile (Figure 3-25). Compared to wildtype BL21 cells, expression of the algal chloroplastic ACP (Cr-cACP) in *E. coli* drastically increased levels of 18:1 (Δ11), 16:0, and 16:1, and 14:0 and 18:0 levels were also slightly elevated (Figure 3-6). This result is similar to the overexpression of native EcACP (Figure 3-6). In contrast, the algal mitochondrial ACP (Cr-mACP) did not result in an alteration the *E. coli* FA profile (Figure 3-6).

ACP-TE *in vitro* crosslinking experiments indicate protein-protein interactions between the algal TE and bacterial ACP. No interactivity is observed *in vitro* between plant TEs and the bacterial ACP. Thus, we next investigated the *in vivo* effect of expressing the algal (Cr) TE in bacteria (*E. coli*) and compared it to plant TE expression. All induced TEs decreased fatty acid content compared to wildtype BL21s, yet uninduced TEs resulted in a significant change in fatty acid content (Figure 3-6). Basal expression of CrTE in *E. coli* increases levels of 14:0 and 16:0 two-fold and 16:1 is elevated three-fold. Furthermore, 18:1 (Δ11) acid production as well as 8-(2-hexylcyclopropyl)octanoic acid (cyC17:0, a cyclopropane containing fatty acid derived from C16:1) synthesis is significantly enhanced (Figure 3-6). Similar changes are observed in the *E. coli* strain expressing the plant ChTE. For instance, 18:0 and 18:1 (Δ11) are enhanced 3-fold (Figure 3-6). All undinduced algal and plant TEs increase cyC17:0 levels in *E. coli* (Figure 3-6).

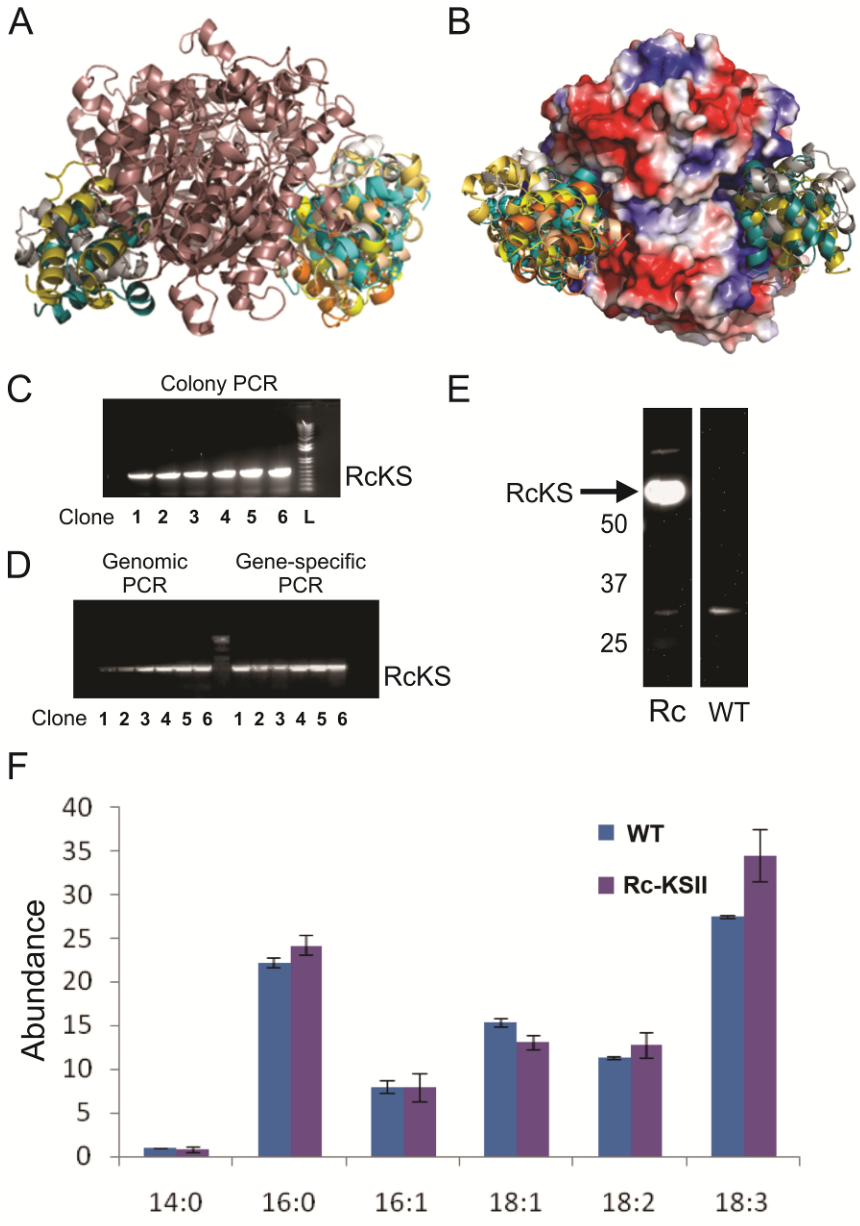


Figure 3-7 Algal ACP-plant KS II interactions in vivo

A) Docking simulation of Cr-cACP to RcKSII. B) Surface representation of A. C) Colony PCR to amplify RcKSII from Cr genomic DNA. D) PCR screen to identify insertion of RcKSII gene in Cr chloroplast genome (genomic) and to confirm homoplasmicity (gene-specific). E) Western blot analysis to detect expression of RcKSII protein in algal (Cr) chloroplast. F) Fatty acid analysis of Cr strain expressing RcKSII by GC/MS.

Induction of plant TE expression (UcTE and ChTE) in *E. coli* changed the fatty acid profile corresponding to their substrate specificity, consistent with studies

previously reported [27]. However, CrTE overexpression did not significantly alter fatty acid production in *E. coli* (Figure 3-6).

Engineering of a plant KS into the algal (Cr) chloroplast and fatty acid analysis of transgenic algal strain

To determine whether plant KSs functionally interact with the algal ACP *in vivo*, the KSII from *Ricinus communis* (RcKSII) was engineered into the *C. reinhardtii* chloroplast and the transgenic *C. reinhardtii* strain expressing RcKSII analyzed for fatty acid content (Figure 3-7). Compared to wildtype, no significant deviation in fatty acid content or composition was detected in transgenic *C. reinhardtii* strains expressing the plant KSII, but an increase in 18-carbon fatty acids was observed (Figure 3-7). As KSII functions to elongate 16:0-ACP to 18:0-ACP, this increased flux towards C18 fatty acids is an indication that the plant KS was able to functionally interact with Cr-cACP *in vivo*.

Discussion

Engineering fatty acid biosynthesis is an attractive platform for creating designer oils in heterologous hosts for the production of renewable biodiesels that comply with the existing infrastructure [50] [7]. In two recent examples [8] [9], specific plant FAS enzymes have been engineered into algal plastids as a mechanism to alter the oil content and fatty acid composition, but unpredictable results and low yields has illuminated a lack of fundamental understanding of algal fatty acid synthase. We have recently shown that protein-protein interactions between the ACP and TE mediate the hydrolysis of fatty acids from ACP in green alga *Chlamydomonas*

reinhardtii [9]. Additionally, we were able to show that plant TEs from *Cuphea hookeriana* (Ch) and *Umbellularia californica* (Uc) do not functionally interact with the ACP present in the *C. reinhardtii* chloroplast both *in vitro* and *in vivo* [9]. However, when these same plant TEs were engineered into bacteria (*E. coli*) and cyanobacteria (*S. elongatus*), they were able to siphon fatty acids and alter the fatty acid profile [51] [33]. These findings, along with similar engineering efforts, illuminate an apparent directionality in ability of FAS enzymes to interact with ACPs across species that remains to be elucidated.

Hypothesizing that the evolutionary origin of FAS enzymes dictates this interchangeability, we first sought to probe ACP-TE interactions across species, beginning with a thorough bioinformatic investigation of all known acyl-ACP TEs. The evolutionary origin of the plant TE is currently unknown, but the common assumption is that it occurred after the first photosynthetic eukaryotes, originating in the cyanobacterial symbiont [43]. Fat1 TEs were recently discovered in green algae and diatoms [9] [21], and BLAST analysis against sequenced algal genomes indicates that similar TEs are present in all algal species. An exhaustive search did not detect fatty acid TEs in any cyanobacterial genome. Thus, we conducted phylogenetic analyses of all known TEs, which reveals the presence of TEs in anaerobic bacteria, likely the ancestral origin of algal and plant TEs (Figure 3-3, Figure 3-13). Multiple sequence alignment of algal Fat1 TEs, plant FatA and FatB TEs, and anaerobic bacterial TEs illustrates a high degree of sequence conservation and high sequence identity around the active site residues (Figure 3-12). Therefore, we postulate that TEs evolved prior to the cyanobacterial engulfing event which led to the formation of

chloroplasts. In this scenario, the fatty acid biosynthetic machinery present in cyanobacteria and bacteria evolved to accommodate the TEs of their eukaryotic host, and thus bacterial and cyanobacterial FAS can recognize algal and plant TEs. This is most likely the explanation for the observation that plant and algal TEs function in bacteria and cyanobacteria *in vivo*. Interestingly, TE phylogeny closely matches those assembled for enzymes involved in anaerobic energy metabolism [52], such as hydrogenases [53] [54] [55], and cyclopropane fatty acid synthases [56]. As in these cases, it is probable that TEs were obtained by horizontal gene transfer between an early eukaryote (prior to the cyanobacterial symbiont) and an anaerobic bacterium [41]. Whereas hydrogenase and cyclopropane synthase expression in ancient anaerobic bacteria is well understood, the reason why they express acyl-ACP thioesterases remains a mystery.

Although aerobic bacteria and cyanobacteria do not have dedicated fatty acid TEs, there are thioesterases sequences annotated in their genomes. For example, TesA (TAP) [57], a thioesterase in *E. coli*, is a versatile enzyme, acting as a TE, phospholipase, and AT [58]. It is capable of releasing C8-18 fatty acids from acyl-ACP in *E. coli* [59], but at very low levels in the cell. In contrast to *in silico* ACP-TE docking between the bacterial (Ec) ACP and algal (Cr) TE (Figure 3-4, Figure 3-22), which shows a positive virtual EcACP/CrTE interaction – albeit slightly less favorable than the native CrACP/CrTE interaction (Figure 3-4) – docking TesA to EcACP results in very poor *in silico* interactions (unpublished results), suggesting that TesA may act via fatty acid substrate recognition. In contrast, docking the EcACP to CrTE (Figure 3-4, Figure 3-22) is favorable, although not as robust as the native *in silico* interaction

between the CrACP and CrTE (Figure 3-4). Because plant TEs, and likely algal TEs, also recognize the acyl chain attached to ACP [22], it is possible that this is the mechanism by which plant and algal TEs alter the fatty acid profile in TE-less prokaryotes, such as *E. coli* and *S. elongatus*, *in vivo*. To test this hypothesis, we used two activity-based TE probes of different lengths (6 carbons and 16 carbons) previously used to demonstrate specificity of the CrTE (**2** and **3**). Probes **2** and **3** were loaded onto the *E. coli* ACP to form *crypto*-EcACPs, which were incubated with CrTE, UcTE and ChTE. A crosslinked species was detected in reactions between EcACP and CrTE using C16 probe **2** and C6 probe **3** (Figure 3-5), but the slightly higher yield using C16 probe 3 suggests the acyl chain factors into the crosslinking reaction. As these yields are both lower than CrACP-CrTE complex formation, we conclude that the CrTE does slightly interact with the *E. coli* ACP *in vitro*, but also that CrTE recognizes the acyl chain attached to EcACP, as 16-carbon analogue **3** effectively traps the complex (Figure 3-5). The plant TEs do not crosslink to EcACP using either C6 probe **2** or C16 probe **3** (Figure 3-29), suggesting fatty acid substrate recognition as the mechanism by which they alter the bacterial fatty acid profile *in vivo* (Figure 3-6).

The algal TE (CrTE) and plant TEs (UcTE and ChTE) are functional in bacteria (*E. coli*) *in vivo* (Figure 3-6). Interestingly, TEs expressed in *E. coli* without induction dramatically alters the *E. coli* fatty acid profile (Figure 3-6), and induced TEs result in a decrease of fatty acid species compared to wildtype BL21s (Figure 3-6). Precedence for this phenomenon has been observed in *E. coli*, where TesA concentration relates to the activity of FAS in vivo [60]. Moderate TesA expression

enhances fatty acid production, but high overexpression suppresses the formation of fatty acids [61].

In general, overexpression of algal (Cr) and plant (Uc and Ch) TEs show a decrease in total fatty acid content (Figure 3-6). However, the Ec strain overexpressing the plant UcTE increased levels of C12 fatty acids as previously reported [27], and the strain overexpressing CrTE increased 16:1, according to its predicted natural activity. In contrast, modest expression of CrTE in *E. coli* reveals a dramatic shift in fatty acid composition (Figure 3-6). The levels of 14:0 and 16:0 increase two-fold and 16:1 is elevated by three-fold, respectively, and 18:1 (Δ11) production as well as cyC17:0 synthesis is significantly enhanced (Figure 3-6), indicating that CrTE functionally interacts with EcACP *in vivo* and is able to siphon fatty acids. Similar changes are observed in the *E. coli* strain modestly expressing the plant ChTE (Figure 3-6), and these changes are not do to hindered *E. coli* growth (Figure 3-25). All undinduced algal and plant TEs increase cyC17:0 levels in engineered *E. coli* (Figure 3-6), perhaps due to increased flux through the FAS or upregulation of a cyclopropane synthase.

Whereas the expression of the 12:0-specific plant UcTE in transgenic plants (*Arabidopsis thaliana*) resulted in the accumulation of 12:0 fatty acids at levels up to 24 and 60% compared to wildtype [62], the expression of the plant UcTE in microalgae (*Chlamydomonas reinhardtii* and *Phaeodactylum tricornutum*) only slightly altered the FA profile [9]. However, the overexpression of UcTE in bacteria (*E. coli*) shifted the fatty acid profile towards the accumulation of C12 fatty acids up to a 6-fold increase (Figure 3-6) corresponding to the specificity of UcTE. A change corresponding to substrate preference of plant TEs has also been demonstrated when

plant TEs were engineered into bacteria (*E. coli* strain K27 and BL21) [27] [63]. Similarly, expression of a FatA TE (18:1-ACP specific) from *Arabidopsis thaliana* in *E. coli* resulted in lipid accumulation and a shift towards 16:1 [64]. Previously, we demonstrated that plant TEs from Uc and Ch do not interact with the algal ACP *in vivo* [9]. Here, we have shown that the algal TE does slightly interact with the bacterial ACP *in vitro* using C16 probe **3** designed to mimic the specificity of the algal TE [9]. Further, we have shown that the algal TE (CrTE) functionally siphons fatty acids in bacteria *in vivo* to alter the fatty acid profile (Figure 3-6), which is strengthened by the observation that PtTE also changes the fatty acid content of *E. coli* according to its specificity [21]. Using *in vitro* activity-based crosslinking as a proxy for functional binding [48], we did not observe crosslinking between plant TEs and the bacterial EcACP *in vitro* (using either the C6 or C16 TE probes **2** or **3**) (Figure 3-29). However, plant TEs did act to hydrolyze fatty acids in bacteria *in vivo* to alter the fatty acid profile corresponding to their specificity (Figure 3-6).

In a recent study, thirty-one TEs from various subfamilies were expressed in *E. coli* strain K27 and the fatty acid profile analyzed [63], and the authors conclude that phylogeny alone cannot predict TE specificity. However, these results could be biased by the assay itself, which relies on the native production of fatty acids in *E. coli* and their hydrolysis from the acyl carrier protein by the overexpressed TEs. Since protein-protein and substrate-protein interactions are important in this reaction, it is the question what these results signify. Alternatively, we extracted from literature the published fatty acid profiles of various algal and plant species and compared those with the phylogenetic tree of the acyl-ACP thioesterases from Jing et al. (Figure 3-14)

[63]. Extracting from literature fatty acid profiles of the wildtype algal and plant species shows profiles that group extremely well along the lines of the TE phylogram, indicating that the non-native TEs have a different effect in the *E. coli* K27 strain than in the wildtype strain, and that in fact, phylogeny can be used predict the substrate specificity of TEs (Figure 3-14, Figure 3-15, Figure 3-16).

Taken together, the ACP-TE interaction appears to be selective, as over time TEs evolved to be much more specific, their function correlating with their structure, which presumably affects interactions with ACP. The discovery of the TE ancestor in anaerobic bacteria provides insight into the evolutionary relationship between ACPs and TEs, suggesting a possible scenario in which prokaryotic ACPs adapted to interact with TEs present in their associated host organism. Here, we have demonstrated that protein-protein interactions do occur between the bacterial ACP and algal Fat1 TE, but the bacterial ACP is unable to recognize the further evolved plant TEs. We speculate that algal and plant TEs are acting to recognize the acyl chain attached to ACP, likely the mechanism by which plant/algal TEs alter the fatty acid profile of TE-less prokaryotes such as *E. coli* and *Synechocystis*.

In contrast to TEs, the KS domain is a derivative of prokaryotic FAS [44]. Thus, we next sought to elucidate compatibility between KSs and ACPs across species. Whereas the TE only functions in the final step of fatty acid biosynthesis, the KS is involved in all cycles of fatty acid elongation (Figure 3-1). The TE domain intercepts the acyl chain attached to ACP and primes the thioester bond for attack by water, and the KS domain facilitates the carbon-carbon bond-forming reaction as a fatty acid is elongated by two-carbon units. In contrast to the ACP-TE interaction,

which is selective and unidirectional, the KS domain must have permissive interactivity with ACP for the survival of an organism. For example, in *E. coli*, KAS I, responsible for the elongation of C6-ACP to C16-ACP can take over for KASII (forms C16-ACP from C14-ACP) *in vivo* [65]. In 2006, Worthington *et al*. found that the ACP-KS interaction is permissive between prokaryotic FAS systems, but unrelated ACPs (i.e. from NRPS systems) do not functionally interact with KSs from FAS [47]. Aligning the sequences of KSIs, KSIIs, and KSIIIs from a broad spectrum of bacteria, algae, and plants show a high degree of sequence conservation across species (Figure 3-8, Figure 3-9, Figure 3-10), and phylogenetic analysis shows that all KSs have a common ancestor (Figure 3-2, Figure 3-11) [44]. Therefore, in contrast to the ACP-TE interaction, we propose that ACP-KS interactions are interchangeable across species.

In silico modeling of protein-protein interactions illustrates a similar binding motif between ACPs and KSs in bacteria, algae, and plants (Figure 3-4, Figure 3-7). As expected from the high homology between ketosynthases across species and the considerable degree of interrelatedness (Figure 3-2), the algal ACP docks to the bacterial KSII in a very similar fashion as the bacterial ACP-KS *in silico* interaction (Figure 3-4). This computational result is validated *in vitro* using activity-based crosslinking to trap the ACP-KS complex. Upon loading Cr-cACP with chloroacrylamide KSII probe **1** [36] and incubating *crypto*-Cr-cACP with KSII from *E. coli*, Cr-cACP/EcKSII complex formed just as readily as EcACP/EcKSII formation (Figure 3-5), indicating a strong functional binding between the algal ACP and bacterial KS. Notably, extracting the *E. coli* ACP from its co-crystal structure with a partner p450 enzyme (PDB ID: 3EJB), and docking the extracted EcACP to the *E. coli*

KS shows a much tighter interaction than simply docking the *E. coli* ACP and *E. coli* KS independently (Figure 3-4, Figure 3-19). This result emphasizes the role of ACP and protein-protein interactions in fatty acid biosynthesis.

Cr-cACP alters the fatty acid profile when overexpressed in *E. coli* (Figure 3-6), demonstrating the algal ACP interacts with the bacterial KS *in vivo*. Additionally, we observe predominantly *holo*-Cr-cACP when expressed in *E. coli* (Figure 3-23), indicating that Cr-cACP is recognized by the bacterial biosynthetic machinery and undergoes post-translational modification by the *E. coli* ACPS *in vivo* [66]. Moreover, overexpression of the algal ACP (Cr-cACP) in *E. coli* results in a large, unpredictable shift in fatty acid profile (Figure 3-6), demonstrating the ability of the algal ACP to functionally interact, in part, with *E. coli* FAS. However, despite the high sequence identity between ACPs across species (Figure 3-17), the algal Cr-cACP is not able to complement *E. coli* FAS in an ACP-knockout strain (CY1877) [49] indicating it cannot functionally interact with all of the requisite catalytic FAS enzymes (Figure 3-24). Cronan et al. previously demonstrated that the ACP from *Spinacia oleracea*, highly homologous to Cr-cACP, also failed to complement *E. coli* fatty acid biosynthesis [49]. Interestingly, Cronan et al. found that the spinach-ACP loaded with a 16:0 fatty acid in *E. coli* does not inhibit acetyl CoA carboxylase (ACCase), whereas EcACP loaded with any fatty acid does [67], suggesting either different FAS regulatory mechanisms in bacteria, algae, and plants or a lack of protein-protein interactions between the ACP and ACCase.

Overexpressing native FAS enzymes has provided insight towards the rate limiting steps and redundancy of the system. For example, simultaneous expression of

FabA (DH) and FabB (KS1) in *E. coli* decreases the amount of saturated fatty acids and increases the amount of C18:1, but does not change the amount of lipids produced. In contrast, overexpression of FabA increases the amount of saturated fatty acids, but overexpressing KS1 negates this effect, suggesting that FabA does not catalyze a rate-limiting step in fatty acid biosynthesis, but that KS1 does [68]. Notably, overexpression of FabF (KSII) in *E. coli* hinders cell growth because it is toxic, presumably by affecting the malonyl-CoA pools [69]. Here, we expressed FabF in *E. coli* for 24 hours at 25 °C and analyzed the fatty acid profile, which resulted in an increase in 18-carbon fatty acids (data not shown).

We next sought to test whether the ACP-KS interaction is favorable in the opposite evolutionary direction using an algal ACP and plant KSII. Examining the virtual model of the algal ACP (Cr-cACP) docked to the plant KSII from *Ricinus communis* displayed very favorable *in silico* binding (Figure 3-7), almost identical to the CrACP-CrKSII model (Figure 3-20). Engineering RcKSII into the algal (Cr) chloroplast resulted in an increase in 18 carbon fatty acids (Figure 3-7), consistent with its enzymatic activity (extending 16:0-ACP to 18:0-ACP), demonstrating RcKSII functionally interacts with Cr-cACP *in vivo*. This result is analogous to the effect of overexpressing FabF in *E. coli* (data not shown). Furthermore, it was recently found that engineering the Δ5-elongase gene, also a thiolase derivative, (Figure 3-2) from algae (*Pavlova* sp.) into moss (*Physcomitrella patens*) resulted in high level production of adrenic acid (22:4) corresponding to the enzymatic activity of the algal enzyme [70]. These examples of successfully engineering a plant KSII into algae and an algal elongase into moss, coupled to the results of our *in silico* ACP-KS modeling

and *in vitro* activity-based ACP-KS crosslinking assay, demonstrate interchangeable ACP-KS interactions across species as a function of a common evolutionary ancestor.

The aim of these efforts was to answer the question of whether evolution can predict FAS engineering. Our results suggest that this is possible. Based on phylogenetic and structure-based analyses, we hypothesized and demonstrated permissivity between KS domains and ACPs of bacteria, algae, and plants. Further, we have unveiled an evolutionary ancestor of algal and plant TEs, and as such, proposed and demonstrated selective ACP-TE interactions. Here, we have probed the directionality of KS and TE domains using phylogenetic analyses coupled to sequence-structure-function analysis, and validate these results using *in vitro* activity-based biochemical assays and *in vivo* genetic engineering. We propose evolution as a new means to decipher compatibility between FAS enzymes across species to facilitate metabolic engineering of fatty acid biosynthesis.

Materials and methods

Sequence alignments

Sequence-based alignments of bacterial, algal, and plant ACPs, KSs, and TEs were produced using TCoffee [71] and MUSCLE 3.7 [72], and the ACP structure-based alignment figure (Figure 3-17) was created using ESPript [73]. Full KS, ACP, and TE sequence alignments are provided in Figure 3-8, Figure 3-9, Figure 3-10, Figure 3-12, and Figure 3-17.

Phylogenetic analyses

Phylogeny of KS and TE domains was analyzed initially using Phylogeny.fr [74]. Detailed phylogenetic analyses of the thiolase superfamily and acyl-ACP thioesterase family were conducted with sequences obtained from the UniprotKB database using Bayesian inference [75]. Sequences were aligned with MUSCLE (3.8.31) [72] and Bayesian trees constructed using MrBayes (3.2.0) [75] and Mega (5.0) [76] (Figure 3-2, Figure 3-3). Full KS and TE phylogenetic trees are provided in Figure 3-11 and Figure 3-13.

Protein modeling and docking

Proteins were modeled using Swissmodel [77] and docked using the Cluspro server [78] (Figure 3-4) (Figure 3-18, Figure 3-19, Figure 3-20, Figure 3-21, Figure 3-22). Models were manually examined using PyMol [79].

Protein expression and purification

Ec-ACP, Cr-cACP, CrTE, ACP-H, and EcKSII proteins were expressed in *E. coli* and purified according to standard protocols (*SI*). ACP hydrolase (ACPH) catalyzed the formation of *apo*-Cr-cACP and *apo*-Ec-ACP from *holo*-ACPs *in vitro* [80] (*SI,* Figure 3-23).

In vitro activity-based crosslinking assay

Biochemical reconstitution of a truncated CoA biosynthetic pathway formed reactive CoA-analogues *in vitro* [47], and Sfp, a promiscuous phosphopantetheinyl transferase (PPTase) from *Bacillus subtillus* [81], was used to load each ACP with reactive phosphopantetheine analogues **1, 2** and **3** [47]. Reactions contained

phosphate buffer, pantetheine analogue (**1**, **2** or **3** in DMSO), $MgCl_2$, ATP, ACP, CoA-A, CoA-D, CoA-E and Sfp [47]. Following the one-pot formation of *crypto*-ACPs for 1h at 37 °C, EcKSII or CrTE/ UcTE/ ChTE was added and the reactions were left to incubate for 3h (KS) or overnight (TE). In negative control reactions, pantetheine analogue was omitted. To detect crosslinked species, anti-FLAG resin or Ni-NTA resin was added directly to each reaction and incubated for 1h at 25 °C. After washing the resin with TBS, proteins were eluted with 1 M arginine (pH 3.5) or 100 mM imidazole and run on an 8% SDS-PAGE gel stained with Coomassie to visualize crosslinked complexes (*SI*).

In vivo ACP complementation assay

To test whether the algal ACP functionally interacts with bacterial FAS *in vivo*, we used an ACP knockout strain (CY1877) of *E. coli* harboring *acpp* on an arabinose-inducible plasmid [49]. Competent *E. coli* strain CY1877 cells were transformed with IPTG-inducible plasmids harboring ACP genes from green alga *Chlamydomonas reinhardtii* (chloroplastic, ACP2) and gram-negative bacterium *Vibrio harveyi* [82]. Transformants were selected on LB-agar plates supplemented with spectinomycin (50 μg/mL), ampicillin (100 μg/mL) and 0.2% arabinose following overnight incubation at 37 °C. A single colony was used to inoculate a 5 mL LB culture containing spectinomycin, ampicillin and arabinose, which incubated at 37 °C overnight. Cells were diluted and plated on arabinose deplete LB-glucose-agar plates supplemented with spectinomycin and ampicillin. Filter discs with 1 mM, 0.1 mM, 10 μM, 1 μM, and 0.1 μM IPTG were added to each plate and incubated at 25 °C and 37 °C

overnight to screen for ACP complementation (*SI*, Figure 3-24). The assay was repeated in liquid culture according to the procedure outlined in ref. [82].

Fatty acid analysis of transgenic E. coli BL21 strains overexpressing ACPs, KSs, and TEs using GC/MS

E. coli BL21 cells harboring FAS enzymes were cultured in 5 mL LB media supplemented with the appropriate antibiotics at 37 °C overnight. These starter cultures were used to inoculate 5 mL cultures, and upon reaching an OD of 0.8, 500 μM IPTG was added to the cells to induce FAS enzyme expression (Figure 3-25). Cells were induced for 4 hours at 37 °C and harvested by centrifugation. For fatty acid analysis, the bacterial cell pellet was resuspended in methanolic acid, incubated for 30 minutes at 65 °C, and fatty acid methyl esters extracted using hexanes. Fatty acid composition of *E. coli* strains expressing FAS enzymes was determined by GC/MS analysis (Figure 3-6). *E. coli* growth curves can be found in the *SI* (Figure 3-25).

In vivo expression of plant KSII in algal chloroplast

KSII from the castor-oil plant (*Ricinus communis*) was engineered into the *C. reinhardtii* chloroplast using methods previously described [83] (*SI*, Figure 3-26, Figure 3-27)**.** The plant RcKSII gene was codon-optimized for the Cr chloroplast and subcloned into a chloroplast expression vector bearing a FLAG epitope at the carboxy terminus. Biolistics transformed *C. reinhardtii* strain csc137c with the RcKSII gene [84], integration of the RcKSII gene into the chloroplast genome was validated by PCR (Figure 3-7, Figure 3-26), and expression of the RcKSII protein in the Cr

chloroplast was detected by Western blot using M2 anti-FLAG antibody (Figure 3-7, Figure 3-27) [85].

Growth of transgenic C. reinhardtii strain expressing RcKSII and GC/MS analysis of fatty acid content

Cr strains were grown on TAP media [86] agar plates supplemented with carbendazim, ampicillin, kanamycin, and cefotaxime [87] under constant illumination. Liquid TAP media cultures were inoculated and grown under constant shaking in a greenhouse for 3 days. Cultures were harvested, the cell pellet resuspended in methanolic acid, and the reaction proceeded for 30 minutes at 65 °C to form fatty acid methyl esters (FAMEs) from algal lipids. FAMEs were extracted using hexanes and analyzed by GC/MS analysis (Figure 3-7, Figure 3-28).

Supplemental Information

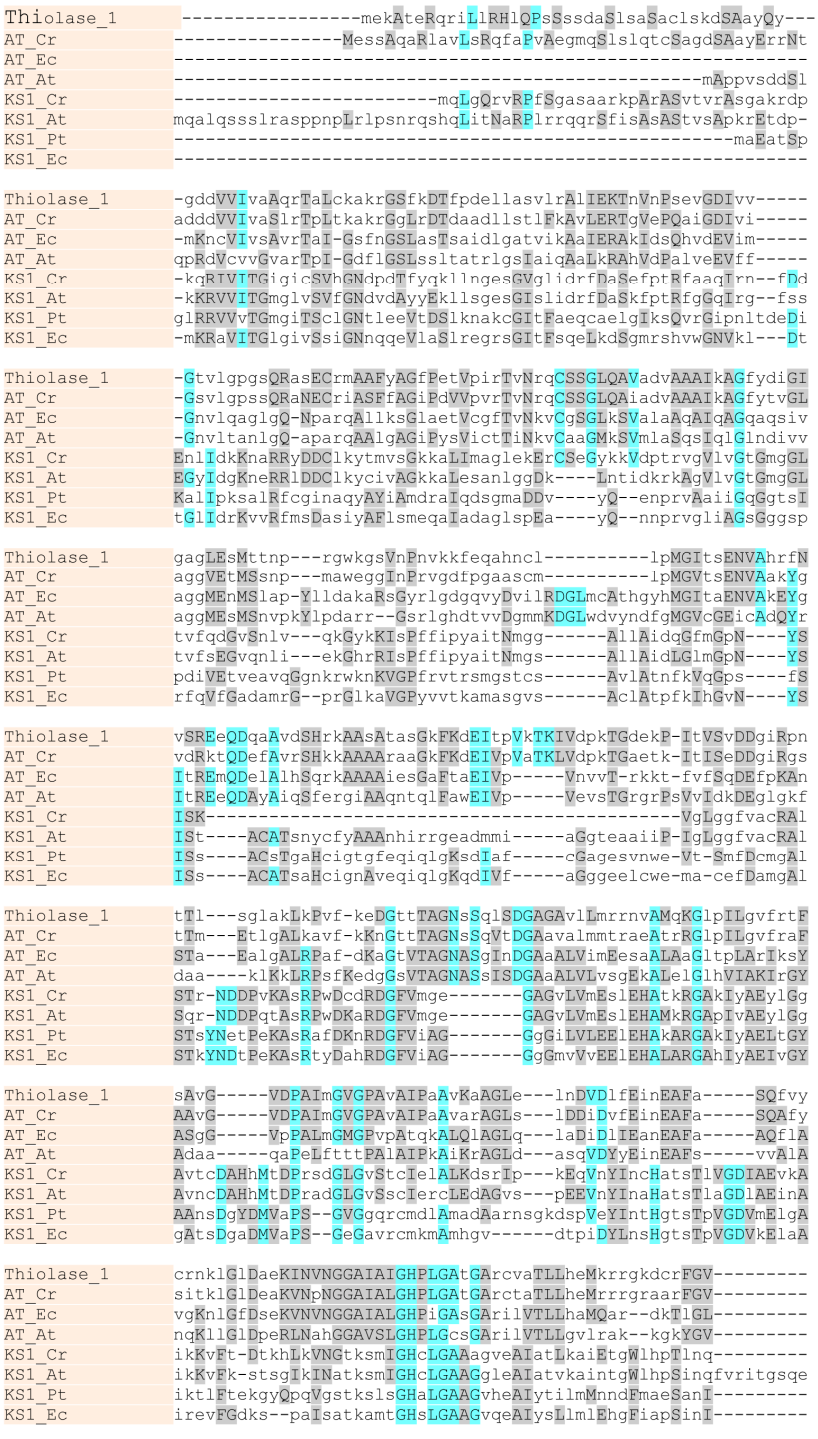


Figure 3-8 Sequence alignment of bacterial, algal, and plant KSIs

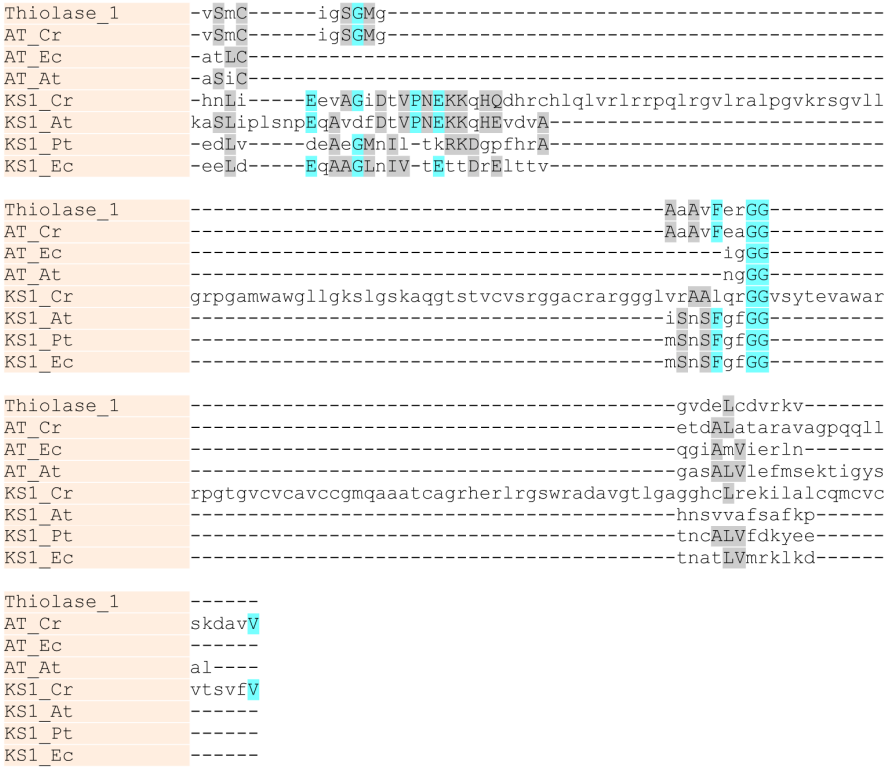


Figure 3-8 Sequence alignment of bacterial, algal, and plant KSIs (continued)

Amino acid sequence alignment of KSIs from the bacterium *Escherichia coli* (KSI_Ec), the diatom *Phaeodactylum tricornutum* (KSI_Pt), the green alga *Chlamydomonas reinhardtii* (KSI_Cr), and the plant *Arabidopsis thaliana* (KSI_At). Acyl transferase (AT) sequences and thiolase sequences were also aligned to illustrate the conservation over time between the enzymes. Sequence alignment was created using MUSCLE 3.7 [48]. Conserved residues are highlighted in grey and identical residues appear in blue boxes. Similar residues are colored as the most conserved according to BLOSUM62. Average BLOSUM62 score: Max, 3.0 (light blue), Mid, 1.5 (dark blue), Low, 0.5 (gray) [88].

Figure 3-9 Sequence alignment of bacterial, algal, and plant KSIIs

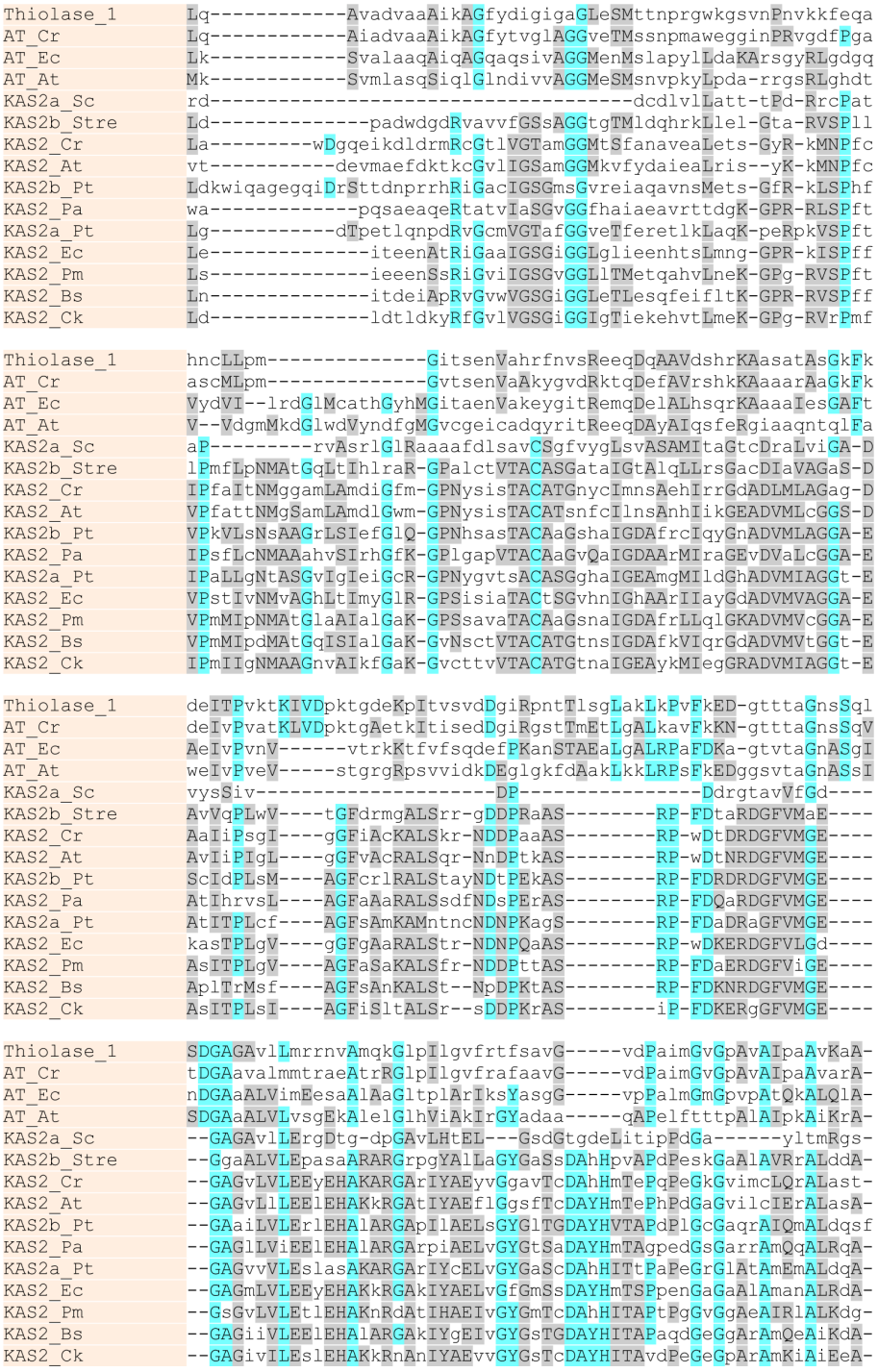


Figure 3-9 Sequence alignment of bacterial, algal, and plant KSIIs (continued)

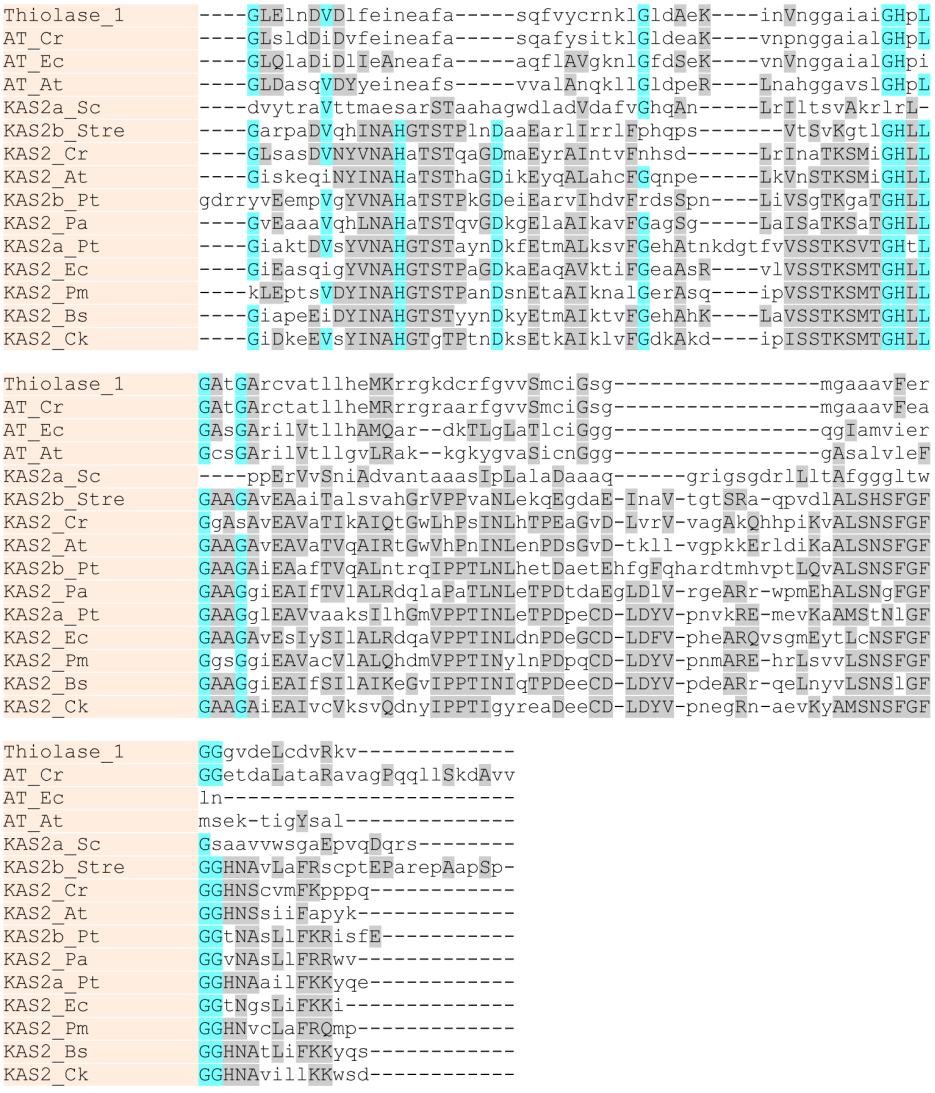


Figure 3-9 Sequence alignment of bacterial, algal, and plant KSIIs (continued)

Amino acid sequence alignment of KSIIs from the bacteria *Escherichia coli* (KAS2_Ec), *Bacillus subtilis* (KAS2_Bs), *Pseudomonas aeruginosa* (KAS2_Pa), *Streptomyces coelicolor* (KAS2_Sc), *Clostridium kluyveri* (KAS2_Ck), *Prochlorococcus marinus* (KAS2_Pm), the diatom *Phaeodactylum tricornutum* (KAS2_Pt), green alga *Chlamydomonas reinhardtii* (KAS2_Cr), and plant *Arabidopsis thaliana* (KAS2_At). Acyl transferase (AT) sequences and thiolase sequences were also aligned to illustrate the conservation over time between the enzymes. Sequence alignment was created using MUSCLE 3.7 [48]. Conserved residues are highlighted in grey and identical residues appear in blue boxes. Similar residues are colored as the most conserved according to BLOSUM62. Average BLOSUM62 score: Max, 3.0 (light blue), Mid, 1.5 (dark blue), Low, 0.5 (gray) [88].

Figure 3-10 Sequence alignment of bacterial, algal, and plant KSIIIs

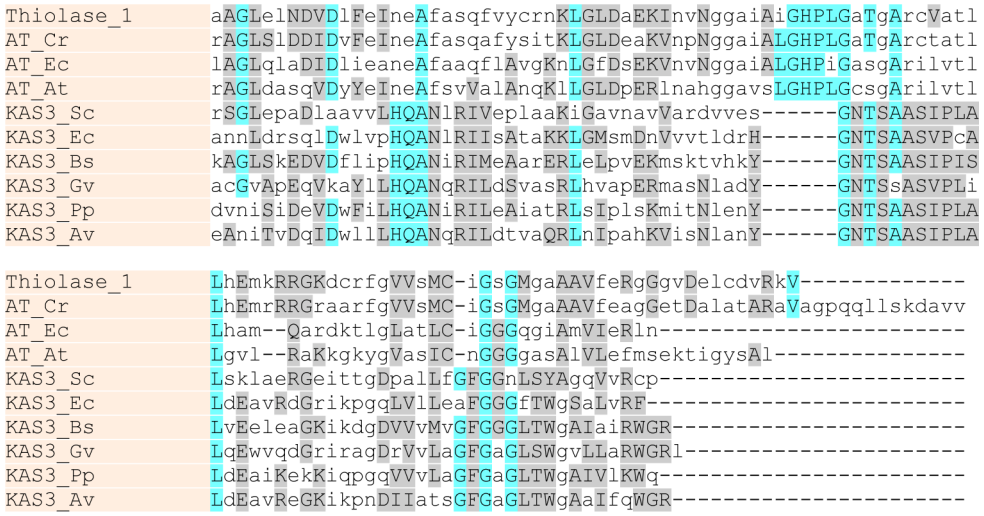


Figure 3-10 Sequence alignment of bacterial, algal, and plant KSIIIs (continued)

Amino acid sequence alignment of KSIIIs from the bacteria *Streptomyces coelicolor* (KAS3_Sc), *Escherichia coli* (KAS3_Ec) and *Bacillus subtilis* (KAS3_Bs), algae *Porphyra purpurea* (KAS3_Pp), and *Anabaena variabilis* (KAS3_Av), and *Gloeobacter violaceus* (KAS3_Gv). Acyl transferase (AT) and thiolase sequences were added to the alignment to illustrate the conservation over time between the enzymes. Sequence alignment was created using MUSCLE 3.7 [48]. Conserved residues are highlighted in grey and identical residues appear in blue boxes. Similar residues are colored as the most conserved according to BLOSUM62. Average BLOSUM62 score: Max, 3.0 (light blue), Mid, 1.5 (dark blue), Low, 0.5 (gray) [88].

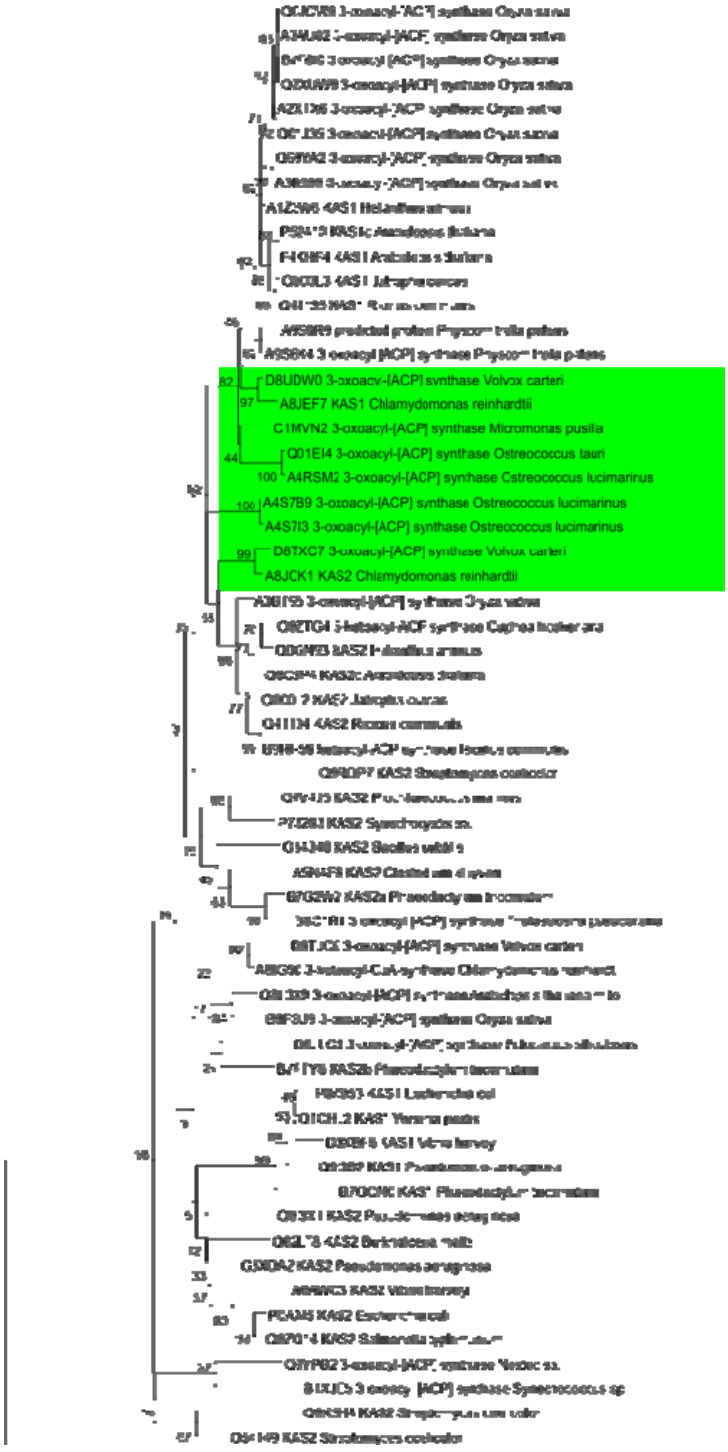


Figure 3-11 KS phylogeny

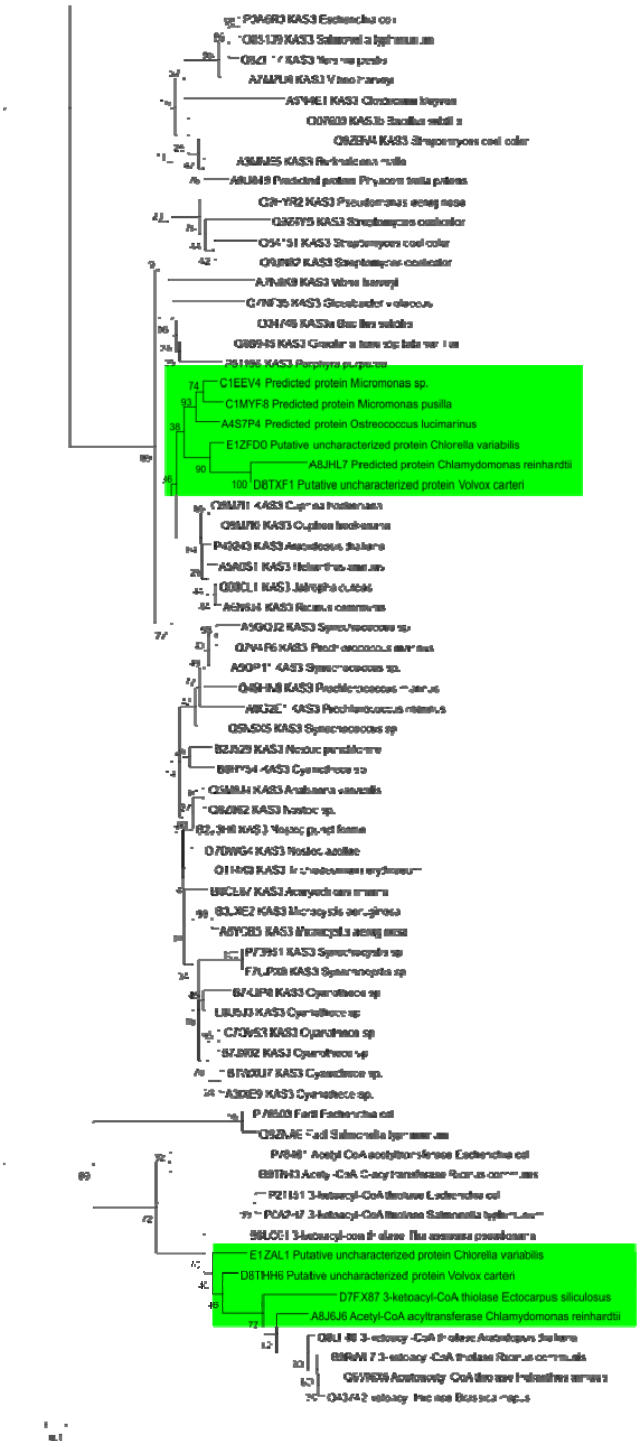


Figure 3-11 KS phylogeny (continued)

Detailed phylogenetic analysis of the thiolase superfamily was conducted with sequences obtained from the UniprotKB database using Bayesian inference [51]. Sequences were aligned with MUSCLE (3.8.31) [48] and Bayesian trees constructed using MrBayes (3.2.0) [51].

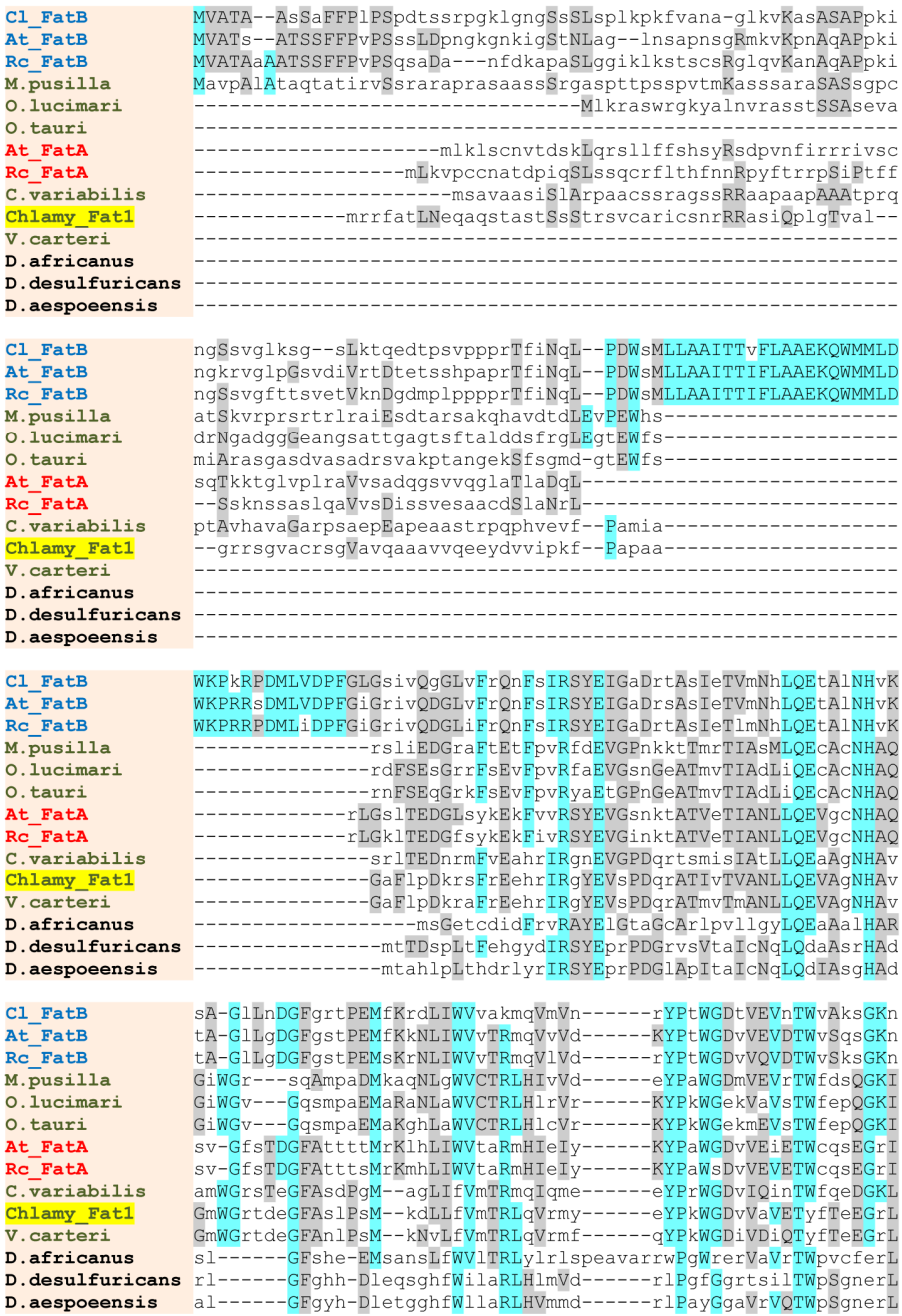


Figure 3-12 Sequence alignment of bacterial, algal, and plant TEs

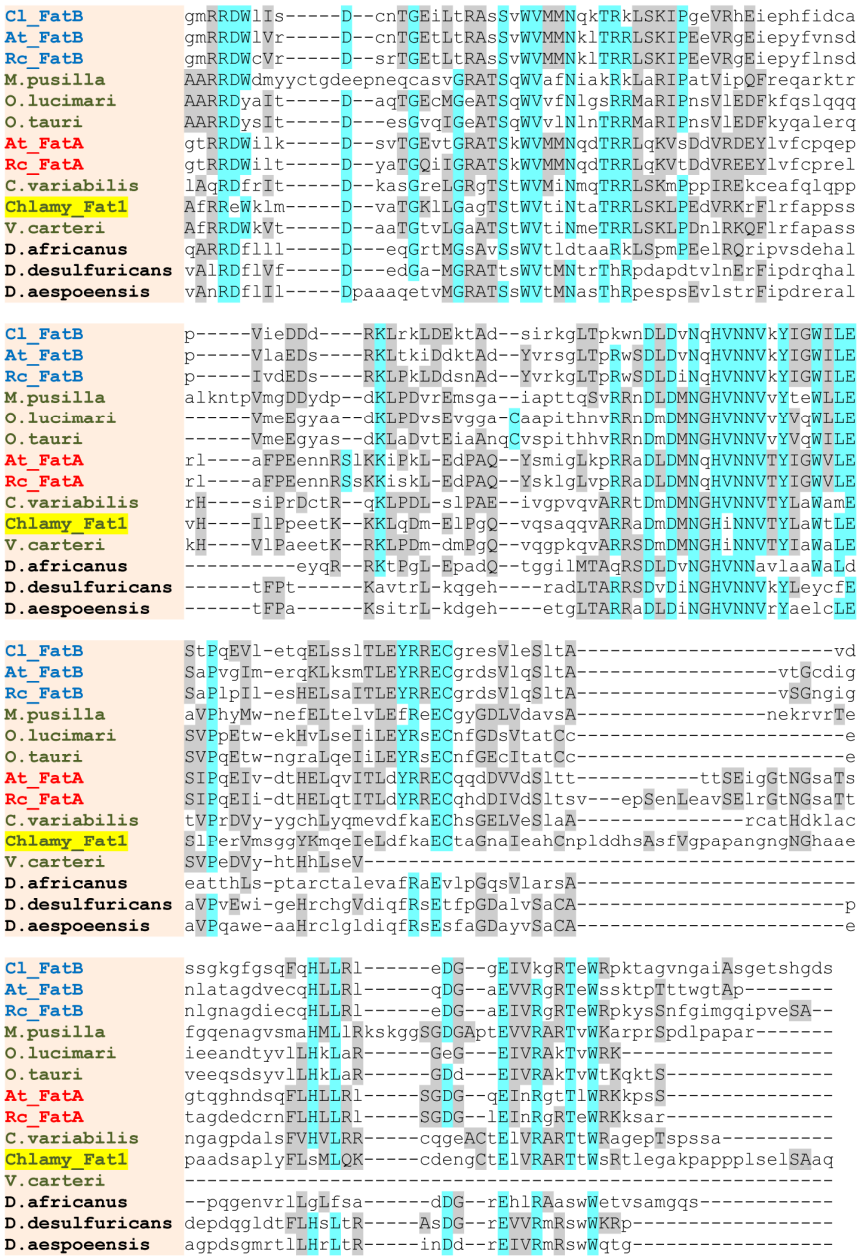


Figure 3-12 Sequence alignment of bacterial, algal, and plant TEs (continued)

Sequence alignment of plant FatA TEs from *Arabidopsis thaliana* (At) and *Ricinus Communis* (Rc), plant FatB TEs from *Arabidopsis thaliana* (At), *Ricinus communis* (Rc), and *Cuphea lanceolata* (Cl), algal TEs from *Micromonas pusilla*, *Ostreococcus lucimari*, *Ostreococcus tauri*, *Chlorella variabilis*, *Chlamydomonas reinhardtii* (Cr, Chlamy), and *Volvox carteri* and anaerobic bacteria *Desulfovibrio africanus*, *Desulfovibrio desulfuricans*, and *Desulfovibrio aespoeensis*. Sequence alignment was created using MUSCLE 3.7 [48]. Conserved residues are highlighted in grey and identical residues appear in blue boxes. Similar residues are colored as the most conserved according to BLOSUM62. Average BLOSUM62 score: Max, 3.0 (light blue), Mid, 1.5 (dark blue), Low, 0.5 (gray) [88].

A

Figure 3-13 TE phylogeny

A

Figure 3-13 TE phylogeny (continued)

B

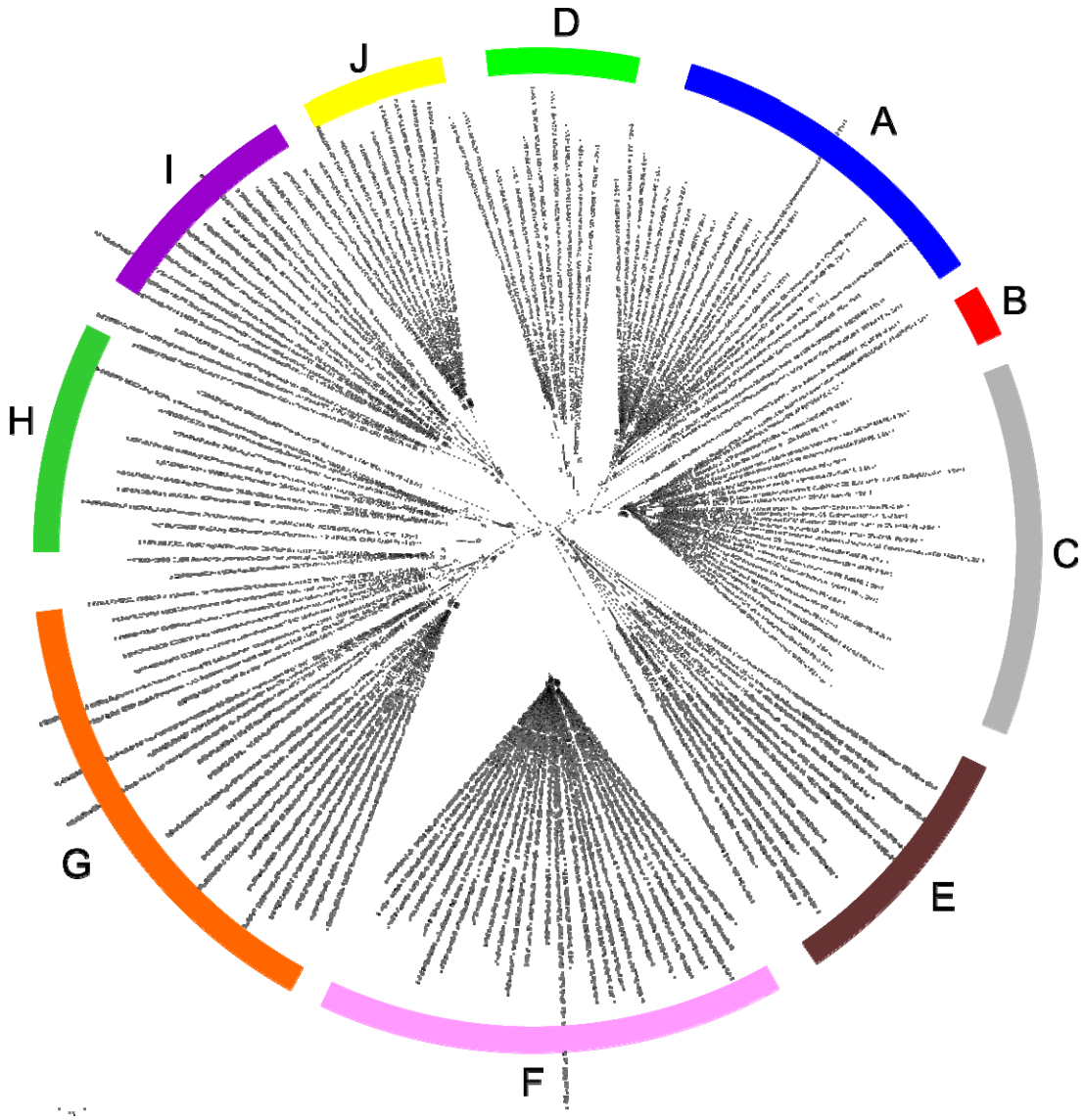


Figure 3-13 TE phylogeny (continued)

Detailed phylogenetic analyses of the acyl-ACP thioesterase family was conducted with sequences obtained from the UniprotKB database, using Bayesian inference [51]. Sequences were aligned with MUSCLE (3.8.31) [48] and Bayesian trees constructed using MrBayes (3.2.0) [51]. A) Phylogram of the relationship between thioesterases; B) Radial representation of the phylogram shown in A. Letters represent TE families, according to Cantu et al. [80].

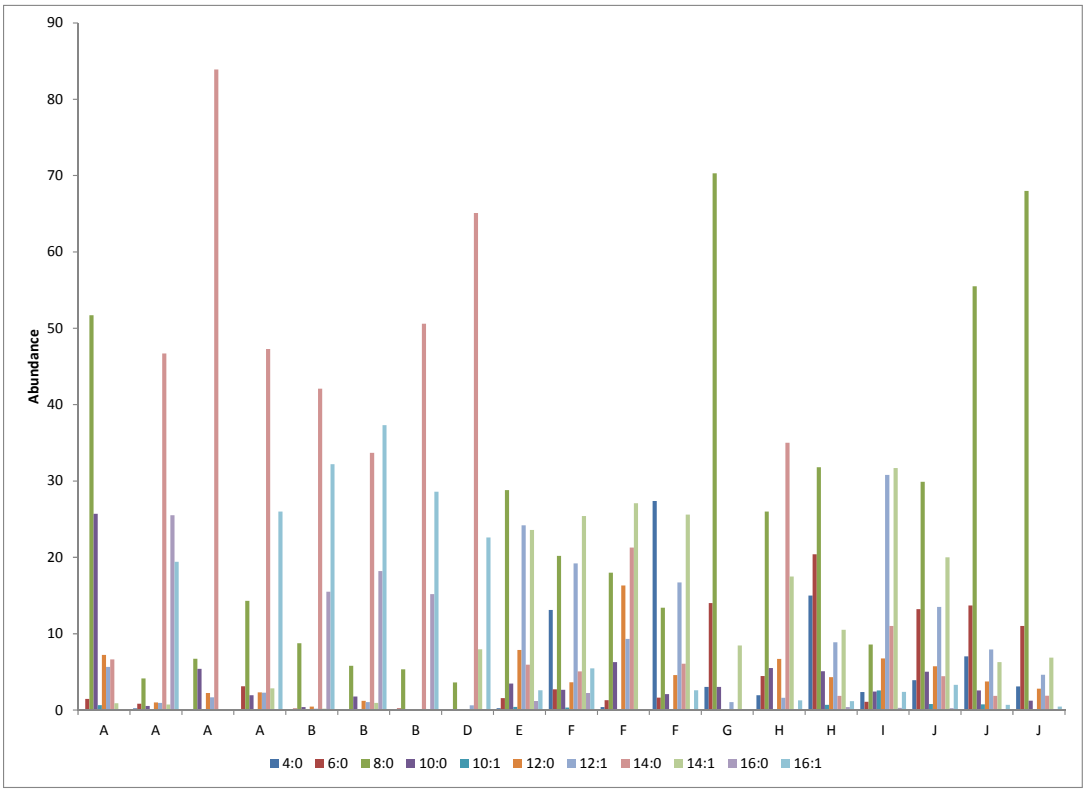


Figure 3-14 GC/MS data of TEs extracted from literature - A

A) Data from Jing et al. plotted in bar graph style, showing acyl-ACP thioesterase subfamilies on the bottom and fatty acid profiles of K27 *E. coli* strains transformed with representative TEs from each class [80].

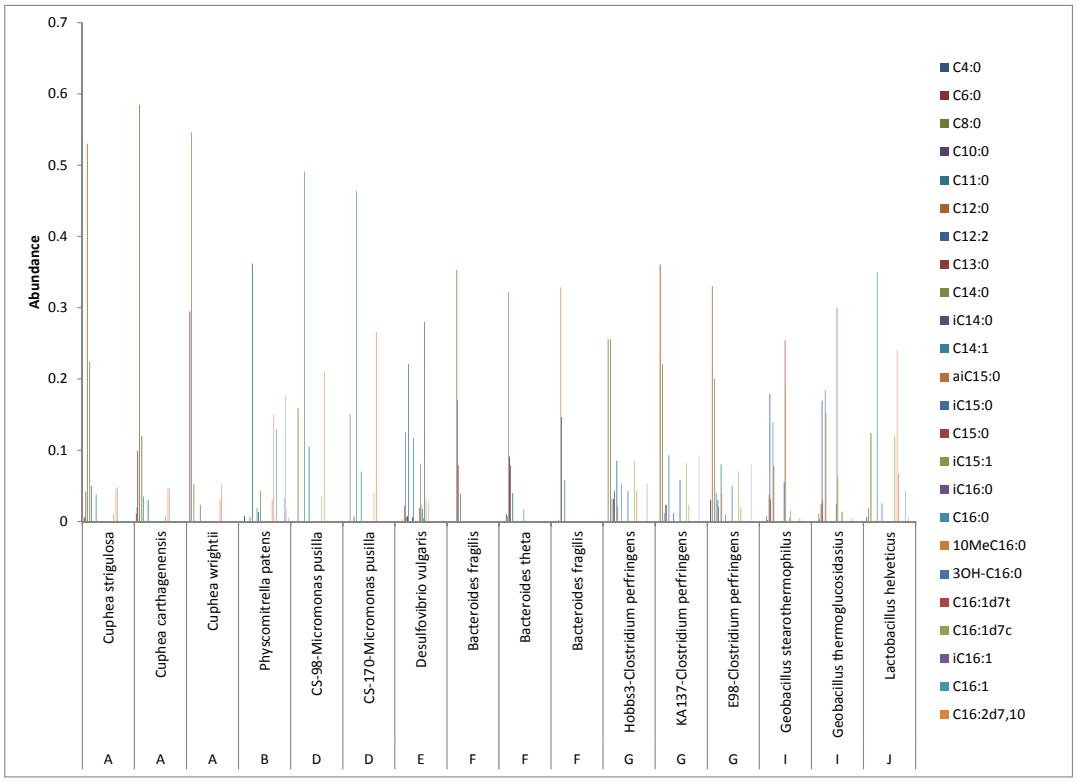


Figure 3-15 GC/MS data of TEs extracted from literature - B

B) Extracted GCMS profiles of various species, which matches the TEs Jing et al. [80] expressed in *E. coli* strain K27.

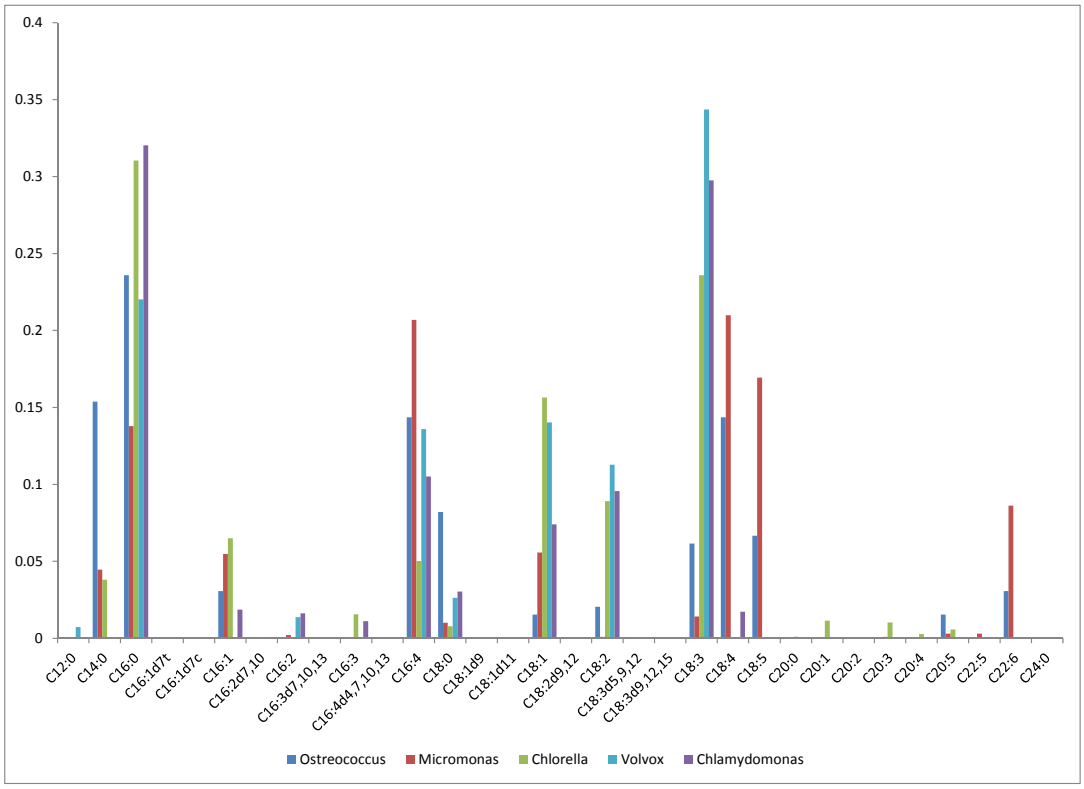


Figure 3-16 GC/MS data of TEs extracted from literature - C

C) Fatty acid profiles of sequenced algae

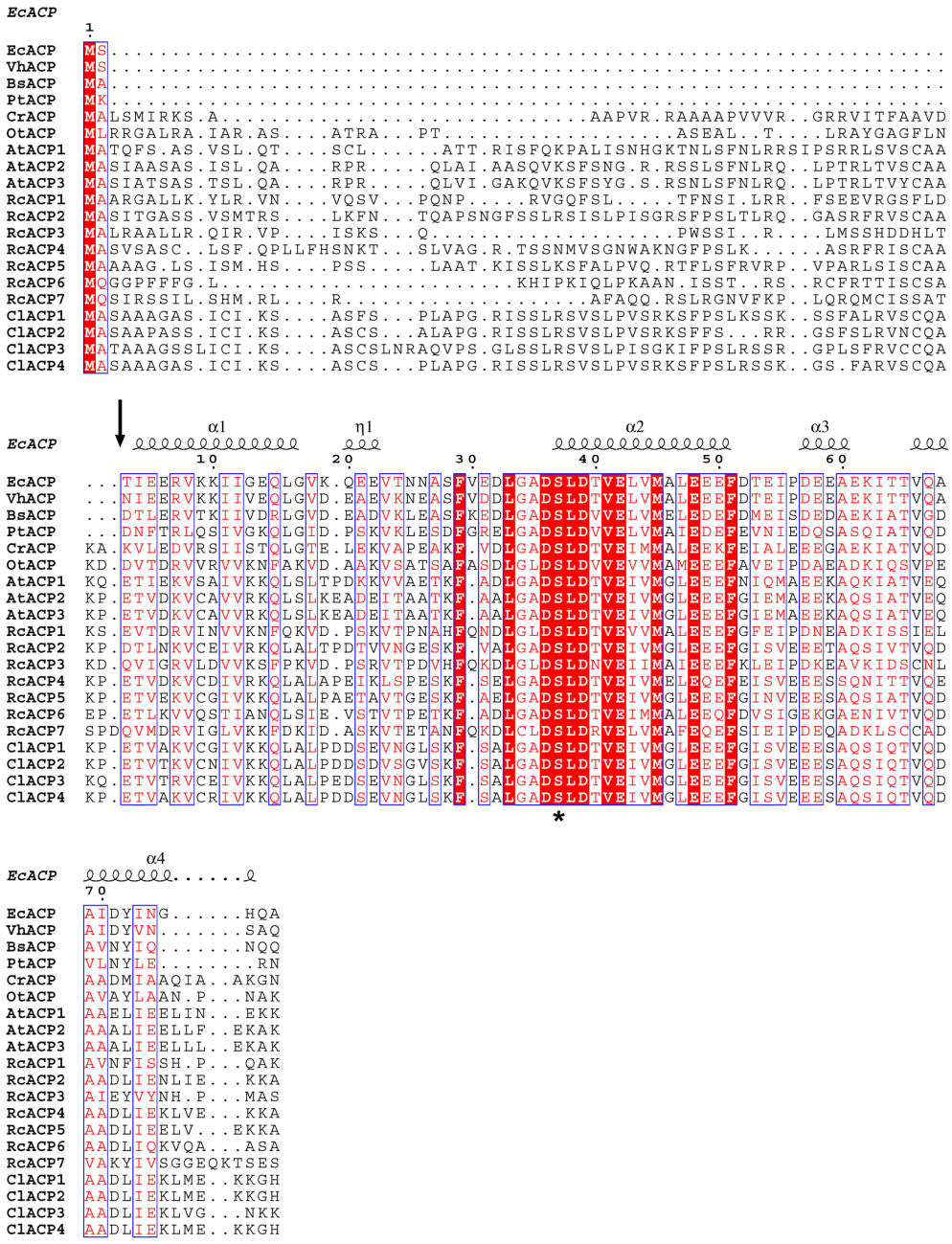


Figure 3-17 Sequence alignment of bacterial, algal, and plant ACPs

Structure-based sequence alignment of ACPs from *Escherichia coli* (Ec), *Vibrio harveyi* (Vh), *Bacillus subtillus* (Bs), *Phaeodactylum tricornutum* (Pt), *Chlamydomonas reinhardtii* (Cr), *Ostreococcus tauri* (Ot), *Arabidopsis thaliana* (At) (3), *Ricinus communis* (Rc) (7), and *Cuphea lanceolata* (Cl) (4). Conserved residues are highlighted in red and similar residues appear in blue boxes. A black arrow indicates the end of transit peptide sequences. Conserved serine is indicated by an asterix under residue. Symbols are denoted as: α, α-helices; β, β-strands; η, 310 helices; TTT, strict α-turn; TT, strict β-turn. Sequence-based alignments were produced using TCoffee [47]. The figure was created using ESPript [49].

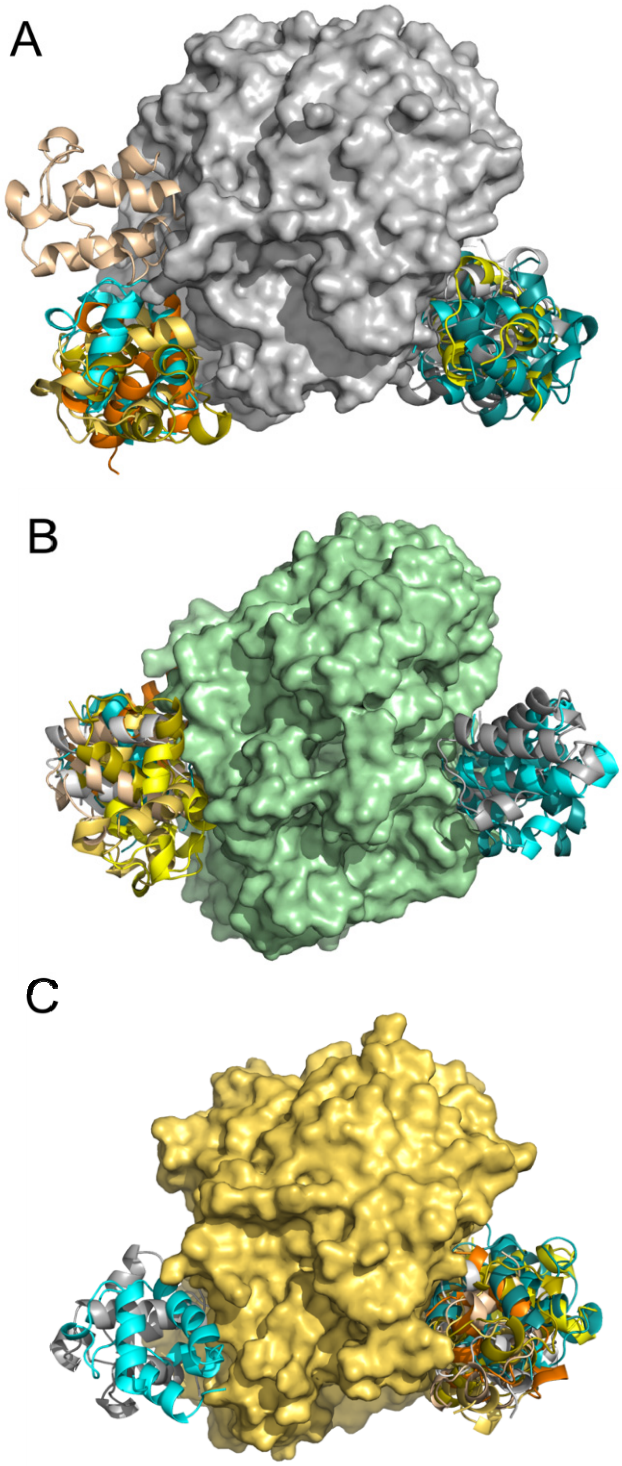


Figure 3-18 Ensemble representation of protein-protein docking of EcACP with KSs

A) EcKSII, B) CrKSII, and C) RcKSII

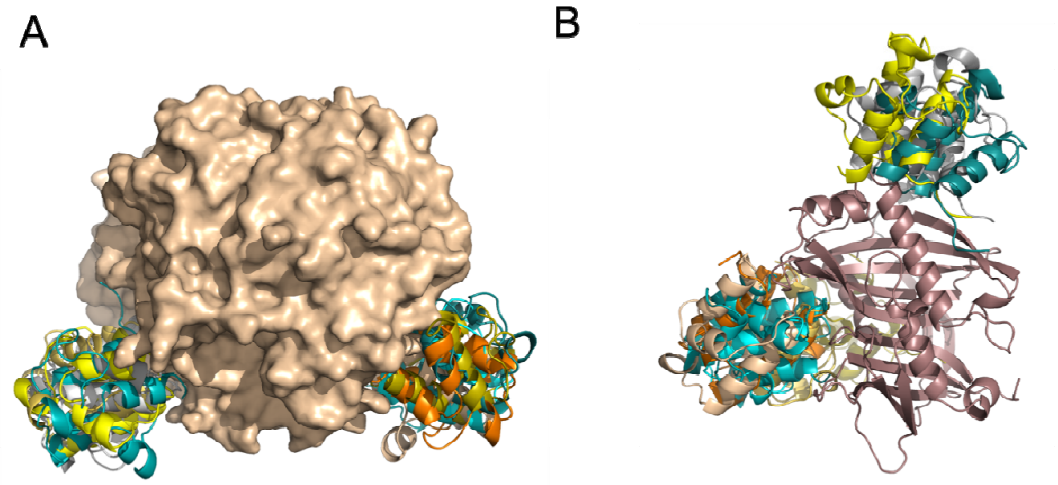


Figure 3-19 Ensemble representation of protein-protein docking with extracted EcACP

Extracted EcACP docked to A) EcKSII and B) CrTE

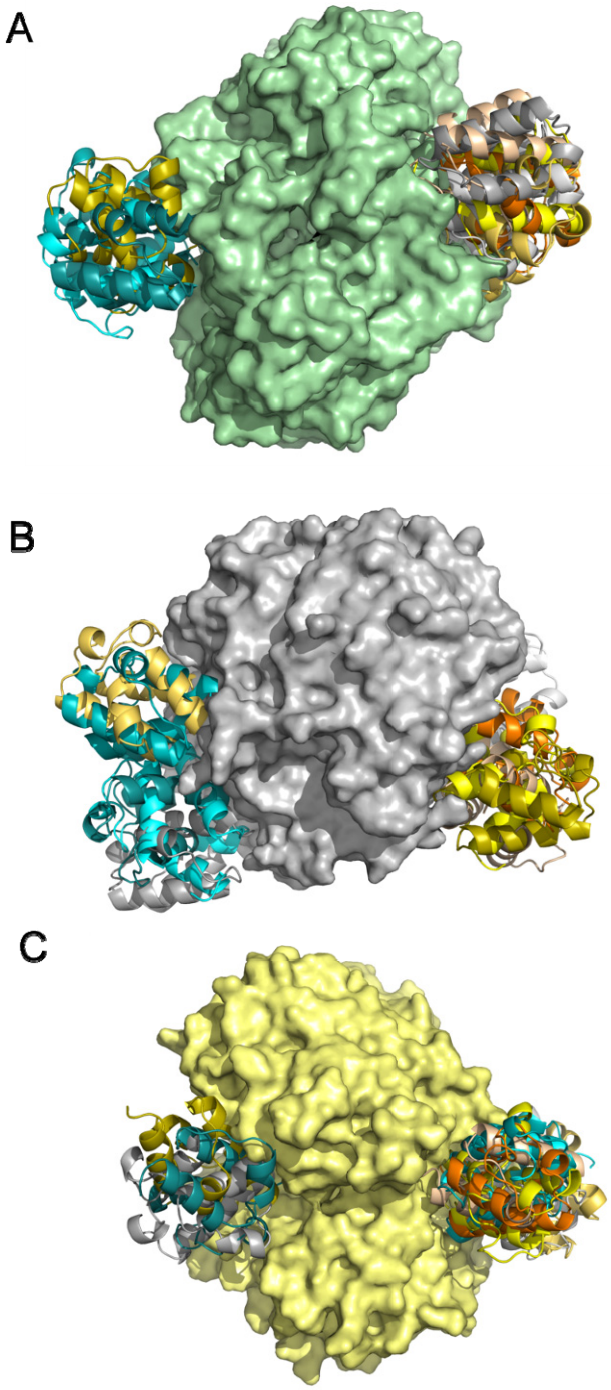


Figure 3-20 Ensemble representation of protein-protein docking of Cr-cACP with KSs

A) CrKSII, B) EcKSII and C) RcKSII

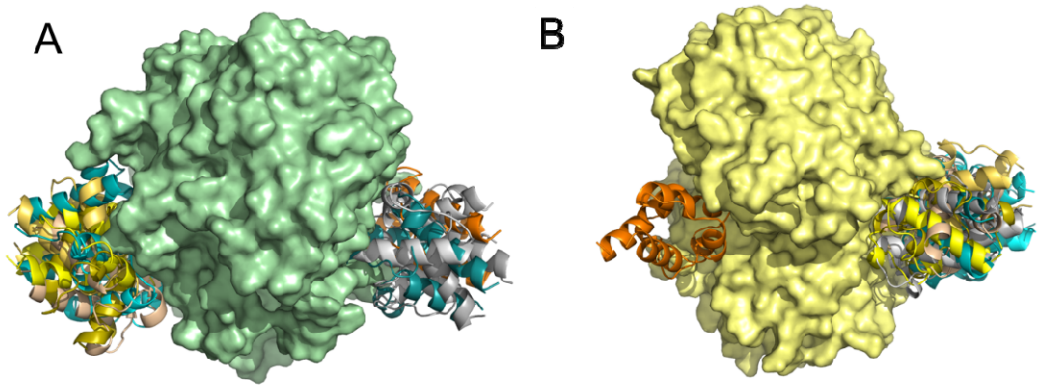


Figure 3-21 Ensemble representation of protein-protein docking of RcACP with KSs

A) RcKSII and B) CrKSII

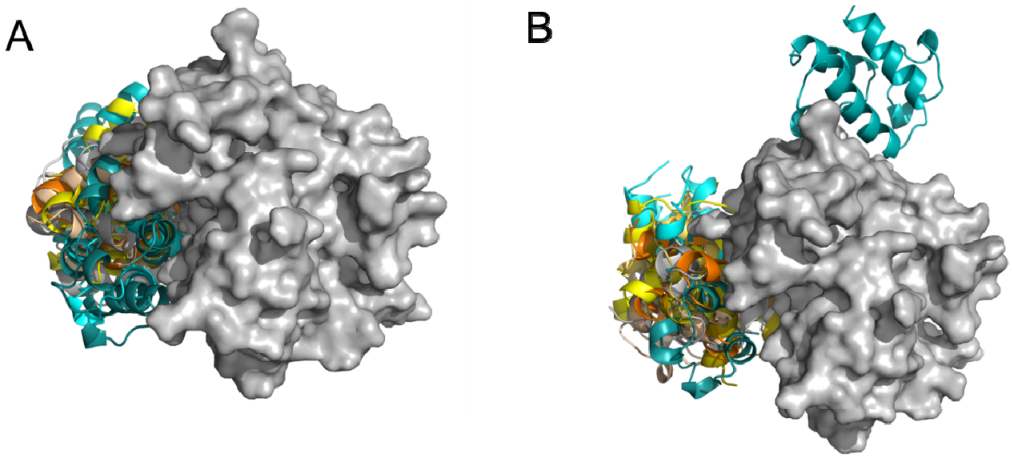


Figure 3-22 Ensemble representation of protein-protein docking of EcACP with CrTE

Protein-protein docking of CrACP- to A) CrTE and B) EcACP to CrTE.

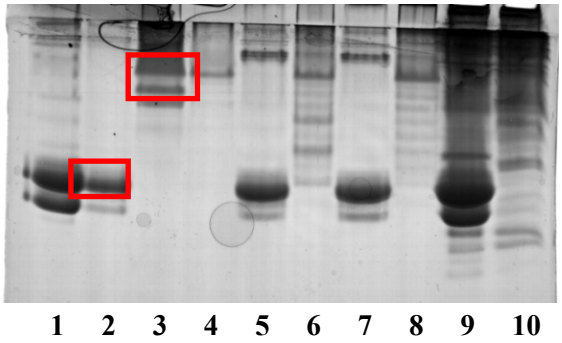
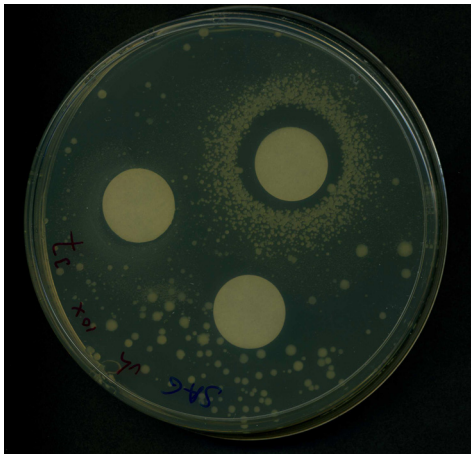
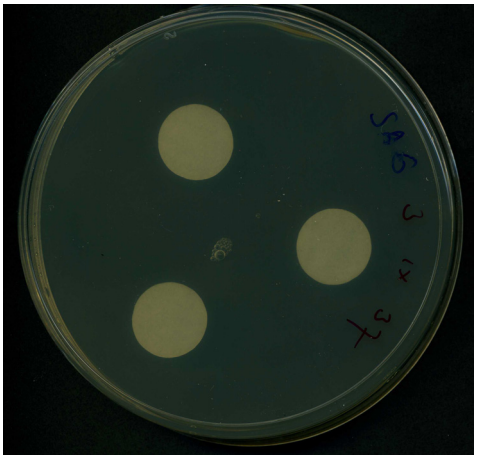


Figure 3-23 Expression of Cr-cACP in *E. coli*

A) Urea gel showing expression of Cr-cACP and Ec-ACP in *E. coli*. 1) Ec-ACP, elution 1, mixture of *holo*- (bottom band) and *apo*- (top band); 2) *apo*-Ec-ACP (red box) formed after treatment with ACPH; 3) Cr-cACP, elution 1, mixture of *holo*- (bottom band) and *apo*- (top band); 4) *apo*-Cr-cACP (red box) formed after treatment with ACPH; 5) Ec-ACP undergoing conversion to *apo*-ACP after 36 hours; 6) Cr-cACP undergoing conversion to *apo*-ACP after 36 hours; 7) Ec-ACP undergoing conversion to *apo*-ACP after 18 hours; 8) Cr-cACP undergoing conversion to apo-ACP after 18 hours; 9) Ec-ACP lysate; 10) Cr-cACP lysate. Samples were loaded on a 15% Urea-PAGE gel and visualized by staining with Coomassie; Ec = *E. coli*; Cr = *C. reinhardtii*.



A) *Vibrio harveyi* ACP complements *E. coli* strain CY1877



B) *Chlamydomonas reinhardtii* ACP does not complement *E. coli* strain CY1877

Figure 3-24 In vivo ACP complementation assay results

A) *Vibrio harveyi* ACP shows complementation both with undiluted and diluted cells (shown here is a 10x dilution) at 0.1 μM IPTG. B) *C. reinhardtii* ACP does not show complementation with undiluted and diluted cells (shown here is a 1x dilution) [59].

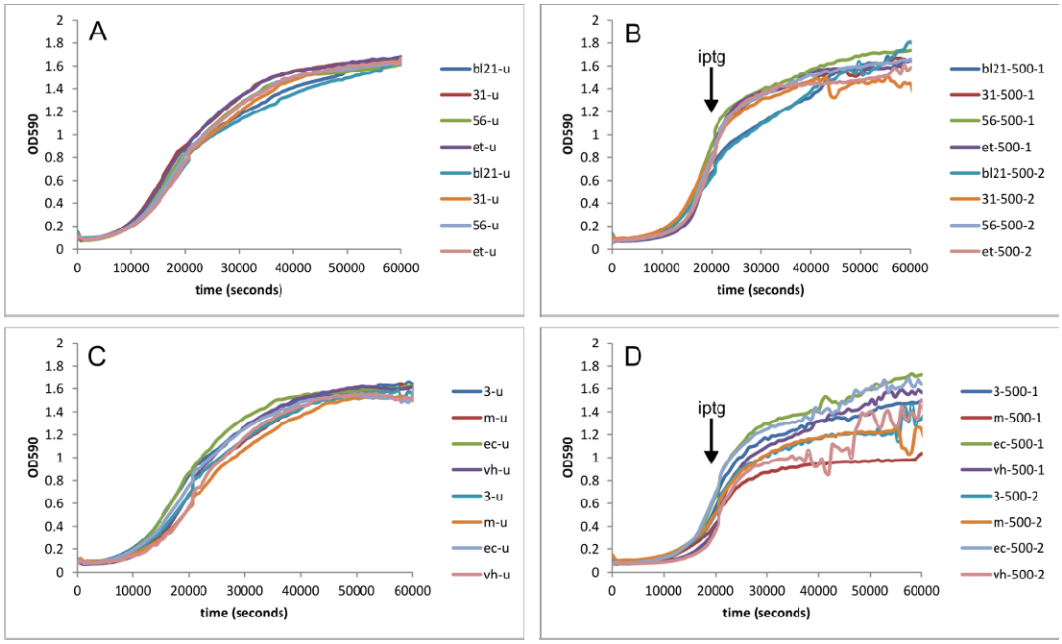


Figure 3-25 *E. coli* growth curves

Growth curves in a 96-well plate at 30 °C of *E. coli* BL21 strains transformed with various thioesterases (A and B) or acyl carrier proteins (C and D). Media was supplemented with the appropriate antibiotic and the wells inoculated with a fresh overnight culture cell suspension. Strains were grown for 6h under continuous shaking and the optical density at 590 nanometers measured. After 6h, protein expression was induced with 500 µM IPTG (B and D) and the OD measured for 12h. Experiments were carried out in duplicate.

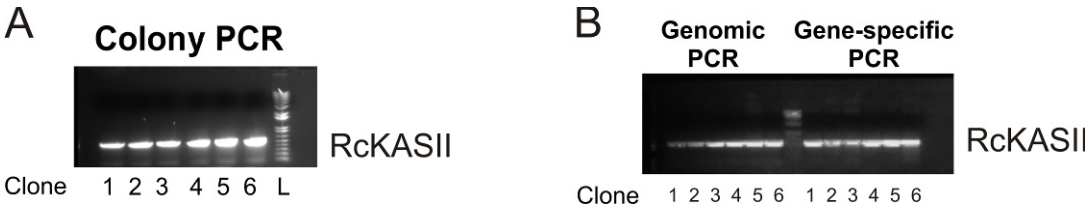


Figure 3-26 PCR screen to identify *C. reinhardtii* transformants harboring KSII from *R. communis*

Initial screening of primary Cr transformants was carried out using PCR to determine which clones have proper insertion of exogenous RcKSII gene. A) Colony PCR on Cr clones showing proper integration of RcKSII gene into the Cr chloroplast genome. B) Left: PCR analysis of genomic chloroplast DNA to determine degree of homoplasmy in Cr transformed with RcKSII (genomic PCR). Right: PCR on inserted nucleotide sequence of Cr transformed with Rc-KSII (gene-specific PCR).

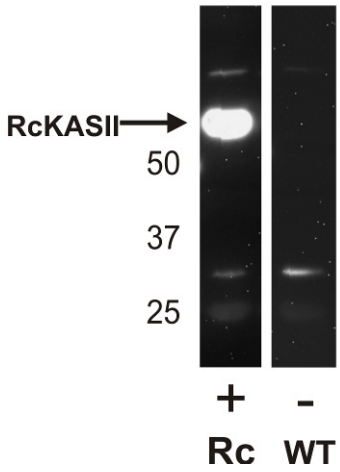


Figure 3-27 Western blot analysis validating expression of RcKSII in *C. reinhardtii* chloroplast

To confirm expression of RcKSII protein in the Cr chloroplast, Western blot analysis was conducted on a transgenic Cr strain containing the RcKSII gene. The RcKSII protein was purified via affinity purification using M2 FLAG resin and the Western blot was carried out using anti-FLAG antibody. Gel images show the detection of RcKSII protein expression in *C. reinhardtii* chloroplast by Western blot analysis. Left: Cr strain expressing RcKSII. Right: Wildtype Cr strain csc137c. Numbers to the left of gel indicate molecular weight in kDa.

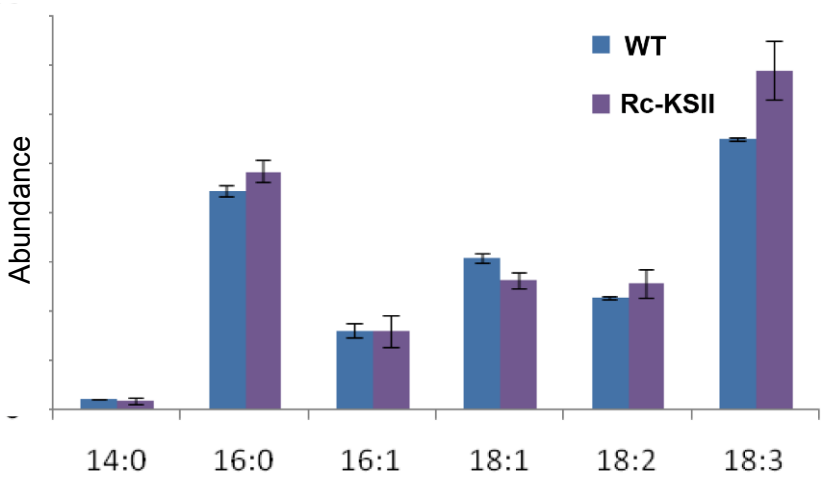


Figure 3-28 Fatty acid composition analysis of transgenic algae (Cr) expressing plant KSII (Rc)

Transgenic Cr strains expressing the RcKSII protein were analyzed for fatty acid content and composition using GC/MS. Wildtype Cr and transgenic Cr expressing RcKSII were grown in TAP media under constant illumination until they reached an OD of 0.7. Cells were harvested and fatty acids extracted as their methyl ester derivatives. Blue bars show the fatty acid composition of wildtype *C. reinhardtii* (csc137c) and purple bars represent transgenic Cr expressing RcKSII. Numbers on the horizontal axis represent fatty acid identity (carbon chain length: #unsaturations) and the vertical axis quantifies the amount fatty acid produced in each strain (% composition).

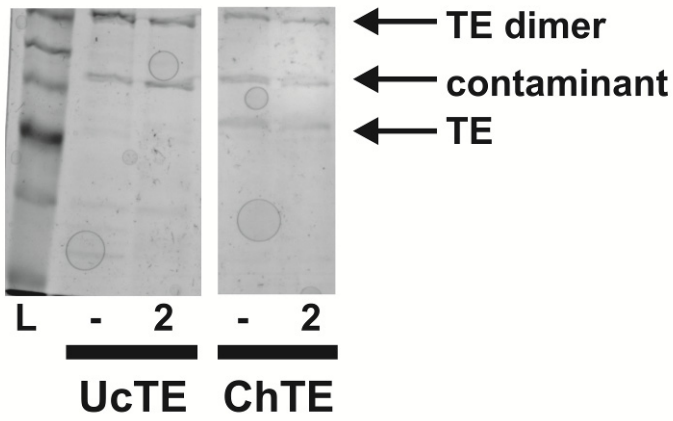


Figure 3-29 In vitro activity-based crosslinking between plant TEs and *E. coli* ACP

Activity-based crosslinking was carried out between the *E. coli* ACP and plant TEs from *Umbellularia californica* (Uc) and *Cuphea hookeriana* (Ch) according to methods described in the experimental section. Following the crosslinking reaction, FLAG resin was added to the mixture and affinity purification yielded TEs or TE complexes, which were analyzed by SDS-PAGE analysis on an 8% gel stained with Coomassie. L = Benchmark ladder.

Materials and Methods

Materials

CoA -A, -D, -E and Sfp were expressed and purified using previously published methods [57]. Protein concentrations were determined using Bradford reagent. Buffers and media were prepared with ultrapure water. Anti-FLAG M2 affinity gel and anti-FLAG antibody were obtained from Sigma-Aldrich. Genomic DNA purification kit was purchased from Promega.

Phylogenetic analyses

Phylogenetic trees comprising various bacterial, algal, and plant KSs (Figure 3-11) and TEs (Figure 3-13) were created using web-based tools. Initially, all KS and TE protein sequences were obtained from the UniProt protein database and submitted for analysis using the one-click mode with default settings of the Phylogeny.fr server [50].

Sequences of putative algal, cyanobacterial, bacterial, plant and archaeal thioesterases, as well as candidates from the thiolase superfamily, were obtained from the UniprotKB database. Some enzymes (for example, ketoacyl synthase III) are not annotated in any algal genome. These putative enzymes were identified by psi-blasting against the microalgal genomes. Initial phylogenetic trees were constructed using the 'one-click' phylogeny.fr server [50]. The thioesterase tree clearly shows that the thioesterases present in cyanobacteria and bacteria (e.g. TesA, TesB) are not part of the acyl-ACP thioesterase family. The thiolase tree shows a similar result to the tree

published by Jiang et al. [44], in which thiolases and ketoacyl synthases are closely related.

Detailed phylogenetic analyses of the acyl-ACP thioesterase family and the thiolase superfamily were conducted with the entire sequences obtained from the UniprotKB database, using Bayesian inference [51]. Sequences were aligned with Muscle (3.8.31) [48] and Bayesian trees constructed using MrBayes (3.2.0) [51]. The 'mixed' amino acid analysis model was used, and Markov chain Monte Carlo analysis was performed for 1 million generations with four independent chains, sampling every 500 generation. At the end of the run, average standard deviation (< 0.001) and potential scale reduction factor (PSRF, approaching 1.0000) were used as a measure for convergence of the process. Tree representations were made using FigTree 1.3.1 or Mega 5.1 [52].

Protein docking

Protein sequences of ACPs, TEs and KSs were obtained from the UniProt database (see details in Table 3-1). Protein models were prepared using the automated Swissmodel server [53], which predicts protein structure based on homology modeling. The four steps used by Swissmodel are template superposition, target-template alignment, model building and energy minimization. In the latter, deviations in the protein structure, which have been introduced by the modeling algorithm when joining stiff fragments, are regularized by using the steepest descent energy minimization of the GROMOS96 force field. Models were compared with the models obtained from the i-Tasser server [89] and no significant deviations between the two

homology modeling algorithms were observed. Since the ketoacyl synthase II is dimeric, dimer models (prior to protein-protein docking) were predicted using the Cluspro server. Proteins were docked using the Cluspro server [54]. The top 10 balanced models were manually examined using PyMol [55].

Table 3-1 Protein modeling

Protein	Uni Prot	Models after	Protein	Organism	RMSD **	Identity	Q-state	Q-mean	Comments
RcKS2	Q41134	3HO9*	FabF	*E. coli*	0.079 A	45%	Dimer	0.59	Does not model with leader peptide.
RcACP	B9RR02	1X3O	ACP	*T. thermophiles* HB8	0.083 A	44%	Mono	0.8	*Ricinus communis* contains 8 ACPs. Cr-cACP was blasted against the Rc genome to identify this ACP.
VhKS2	P55338	3HO9	FabF	*E. coli*	0.078 A	75%	Dimer	0.63	
VhACP	P0A2W3	2L0Q	ACP	*V. harveyi*	0 A	97%	Mono	0.86	
CrKS2	A8JCK1	1J3N	KASII	*T. thermophiles* HB8	0.227 A	47%	Dimer	0.52	
Cr-cACP	Q6UKY5	2EHT	ACP	*A. aeolicus*	0.08 A	53%	Mono	0.7	
Cr-mACP	Q6UKY4	2X2B	ACP	*B. subtilis*	0.09 A	52%	Mono	0.3	
CrTE	A8HY17	2OWN	TE	*L. plantarum*	0.47 A	18%	Mono	0.4	
EcACP		1T8K				Xray	Mono		
EcACP		3EJB				Xray	Mono		Extracted from p450biol
EcKS2		2GFW				Xray	Dimer		

*With i-tasser [89] after 1e5m **RMSD values obtained by 'align' algorithm in PyMol [55]

Table 3-2 Strains, plasmids, and restriction sites

Enzyme	Vector	Organism	Restriction Sites
Cr-cACP	pGEX-5X-3	*C. reinhardtii*	BamH1, Xho1
Cr-cACP	pET28b	*C. reinhardtii*	Nde1, Xho1
Cr-mACP	pET28b	*C. reinhardtii*	Nde1, Xho1
Vh-ACP	pGEX-5X-3	*V. harveyi*	BamH1, Xho1 [82]
Ec-KSII	PCA24N	*E. coli*	Nco1, HindIII [36]
ACPH	pET24b	*P. aerigunosa*	Nde1, Xho1
-	pGEX-5X-3	(empty vector, - control)	

Table 3-3 *C. reinhardtii* strains

Name	Type	Enzyme	Organism	Sequence
Cr_wt	Strain	Cr 137c (mt+)	*C. reinhardtii*	Wildtype strain [95]
Cr_RcKSII	Strain	Cr 137c (mt+) RcKSII	*C. reinhardtii*	Strain transformed with Rc-KSII [85]

Subcloning of Cr-cACP

The *E. coli* codon-optimized gene for Cr-cACP was subcloned into the multiple cloning site of the pGEX-5X-3 vector using BamH1 and Xho1 restriction sites.

In vivo ACP complementation assay

Competent cells were generated from mutant *E. coli* ACP-knockout strain CY1877 [59]. A Cr-cACP/pGEX-5X-3 plasmid was transformed into CY1877 competent cells, streaked onto LB plates supplemented with spectinomycin (50 μg/mL), ampicillin (100 μg/mL) and 0.2% arabinose, and incubated at 37 °C overnight. The same experiment was carried out with a VhACP/pGEX-5X-3 as a positive control [60]. For each construct (Vh and Cr), a colony was picked and grown overnight in 5 ml LB with spectinomycin (50 μg/mL), ampicillin (100 μg/mL) and 0.2% arabinose, and incubated at 37 °C overnight. Cells were diluted to 1x, 10x,

100x, and 1000x and plated on LB-agar plates containing ampicillin, spectinomycin, and glucose. Filter discs supplemented with 1 mM, 0.1 mM, 10 μM, 1 μM, and 0.1 μM IPTG were added to each plate as a screen for ACP expression for complementation of the mutant *E. coli* strain and plates were incubated overnight at 37 °C or 25 °C.

***In vivo* complementation assay in liquid culture**

The *in vivo* ACP complementation assay was also carried out in liquid culture. Single colonies of CY1877 transformed with pGEX plasmids harboring either VhACP or Cr-cACP were resuspended in 1 ml LB medium supplemented with 50 ug/ml spectinomycin and 100 μg/ml ampicillin. Fifty microliters of this solution was used to inoculate 2 ml of LB containing spectinomycin and ampicillin, additionally supplemented with various concentrations of arabinose or glucose/IPTG in a 24-well plate. The absorbance at 590 nanometers was monitored for 2 days and the plate incubated at room temperature continuously shaking. Cells harboring VhACP were able to grow in 0.2% glucose and 1 mM IPTG (albeit much slower than in the presence of arabinose), whereas cells harboring Cr-cACP did not grow at all.

Fatty acid analysis of transgenic *E. coli* BL21 strains expressing ACPs, KSs, and TEs using GC/MS.

E. coli BL21 cells harboring FAS enzymes were cultured in 5 mL LB media supplemented with the appropriate antibiotics at 37 °C overnight. These starter cultures were used to inoculate 5 mL cultures, and upon reaching an OD of 0.8, 500 μM IPTG was added to the cells to induce FAS enzyme expression. Cells were induced for 4 hours at 37 °C and harvested by centrifugation. For fatty acid analysis,

the bacterial cell pellet was resuspended in methanolic acid, incubated for 30 minutes at 65 °C, and fatty acid methyl esters extracted using hexanes. Fatty acid composition of *E. coli* strains expressing FAS enzymes was determined by GC/MS analysis.

Expression of Cr-cACP

The Cr-cACP/pET28b plasmid was transformed into chemically competent BL21 (DE3) *E. coli* cells. 5 ml cultures in LB media, containing 50 μg/ml kanamycin, were grown from a single colony at 37 °C overnight. 1 liter cultures of LB media containing 50 mg/L kanamycin were inoculated with these 5 ml overnight cultures and grown for 4h at 37 °C until they reached an OD_{600} of 0.8, at which point they were induced with 500 μM IPTG (final concentration) and grown overnight at 25 °C. Cells were harvested by centrifugation at 4 °C for 30 minutes at 2000 rpm, and the obtained pellets were frozen at -80 °C.

Expression of ACP hydrolase

The *Pseudomonas aerigunosa* ACP hydrolase (ACPH) [80] in either pET24b (MBP-LVPRGSH-ACPH construct) or pMALc2x (MBP-ENLYFQ-HHHHHH-ACPH construct) plasmid was transformed into chemically competent BL21(DE3) *E. coli* cells and 5 ml LB cultures containing 50 μg/ml kanamycin or 100 μg/ml ampicillin were grown from a single colony at 37 °C overnight. 1 liter cultures of LB, in the presence of 50 mg/l kanamycin or 100 mg/l ampicillin, were inoculated with these 5 ml overnight cultures and grown for 4h at 37 °C. When they reached an OD_{600} of 0.8, cultures were induced with 500 μM IPTG (final concentration) and grown overnight at 25 °C. Cells were harvested by centrifugation at 4 °C for 30 minutes at 2000 rpm, and the obtained pellets frozen at -80 °C.

Purification of ACPH

The pellet from a 1 liter culture was suspended in 20 ml of lysis buffer (50 mM potassium phosphate pH 7.4, 500 mM sodium acetate, 0.1 mM EDTA, 20% glycerol) and lysed using a French pressure cell at 4 °C. Cell debris was spun down at 10,000 rpm at 4 °C for 1h. The MBP-tagged protein was purified by amylose affinity chromatography and eluted from the resin with 100 mM maltose.

Purification of *holo*-Cr-cACP

The pellet from a 1 liter culture was resuspended in 20 ml of lysis buffer (50 mM potassium phosphate pH 7.4, 500 mM sodium acetate, 0.1 mM EDTA, 20% glycerol) and lysed using a French pressure cell at 4 °C. Cell debris was spun down at 10,000 rpm at 4 °C for 1h. The His_6-tagged protein was purified by Ni-NTA affinity chromatography. ACP was identified as *holo*-Cr-cACP by high resolution MS.

Apofication of Cr-cACP and purification of *apo*-Cr-cACP

The purified *holo*-Cr-cACP (0.5 mg/ml) was transferred to dialysis tubing (3000 kDa MWCO) and ACPH, Tris-HCl (pH 8), NaCl, $MgCl_2$, and $MnCl_2$ added to final concentrations of 1 mg/ml and 50, 100, 15 and 1 mM, respectively. This mixture was dialyzed overnight at room temperature against 1 L of 50 mM Tris-HCl (pH 8), 100 mM NaCl, 15 mM $MgCl_2$, and 1 mM $MnCl_2$. The mixture containing the apofied CrACP was purified by Ni-NTA affinity chromatography, concentrated, and further purified by size-exclusion chromatography (FPLC). ACP was identified as *apo*-Cr-cACP by high resolution MS.

Expression of EcKSII

The EcKSII/PCA24N [36] plasmid was transformed into chemically competent BL21 (DE3) *E. coli* cells. 5 ml LB cultures containing 100 μg/ml ampicillin were grown from a single colony at 37 °C overnight. 1 liter cultures of LB containing 100 mg/L ampicillin were inoculated with these 5 ml overnight cultures and grown for 24h at 25 °C (OD_{600} of 0.8), induced with 500 μM IPTG (final concentration), and grown overnight at 25 °C. Cells were harvested by centrifugation at 4 °C for 30 minutes at 2000 rpm. Cell pellets were frozen at -80 °C.

Purification of EcKSII

The pellet from a 1 liter culture was suspended in 20 ml of lysis buffer (50 mM potassium phosphate pH 7.4, 500 mM sodium acetate, 0.1 mM EDTA, 20% glycerol) and lysed using a French pressure cell at 4 °C. Cell debris was spun down at 10,000 rpm at 4 °C for 1h. The His_6-tagged protein was purified by Ni-NTA affinity chromatography.

Activity-based crosslinking experiments

Crosslinking reactions contained 50 mM phosphate buffer (pH 8), 2 mM crosslinker (**1**, **2**, or **3** in DMSO), 12.5 mM $MgCl_2$, 8 mM ATP, 0.4 μg/ml ACP, 0.1 μg/ml CoA-A, 0.1 μg/ml CoA-D, 0.1 μg/ml CoA-E and 0.1 μg/ml Sfp. After 1h incubation at 37 °C, thioesterase (final concentration 0.1 μg/ml) was added and the reaction mixture further incubated. After overnight incubation, a 10 μl suspension of anti-FLAG resin (or Ni-NTA resin) was directly added to the mixtures, incubated for 30 minutes at room temperature, centrifuged, washed with sterile pH 7.4 TBS buffer, and the proteins eluted with 3.5 M arginine (pH 3.5) or 100 mM imidazole prior to

being loaded onto 8% SDS-PAGE gels. Gels were fixed overnight and stained with Coomassie to visualize protein complexes. Crosslinked species were confirmed by high resolution mass spectrometry.

Chlamydomonas reinhardtii chloroplast engineering

Table 3-4 *C. reinhardtii* primers

Number	Type	Enzyme	Organism	Sequence
Primer 1	Primer	3' rev for Rc-KSII	Rc	CCTGACATTACACGCTAGGT
Primer 2	Primer	3HB fwd wt	Cr	CGCCACTGTCATCCTTTAAGT
Primer 3	Primer	3HB fwd	Cr	TGTTTGTTAAGGCTAGCTGC
Primer 4	Primer	psbA 5' UTR fwd	Cr	GTGCTAGGTAACTAACGTTTGATTTTT
Primer 5	Primer	Control fwd homoplasty	Cr	CCGAACTGAGGTTGGGTTTA
Primer 6	Primer	Control rev homoplasty	Cr	GGGGGAGCGAATAGGATTAG
Primer 7	Primer	psbA 5' UTR fwd	Cr	GGAAGGGGACGTAGGTACATAAA
Primer 8	Primer	psbA 3' rev	Cr	TTAGAACGTGTTTTGTTCCCAAT
Primer 9	Primer	psbC 5' UTR fwd	Cr	TGGTACAAGAGGATTTTTGTTGTT
Primer 10	Primer	psbD 5' UTR fwd	Cr	TGGTACAAGAGGATTTTTGTTGTT
Primer 11	Primer	atpA 5' UTR fwd	Cr	CCCCTTACGGGCAAGTAAAC

Plant KS (Rc-KSII) plasmid construction for transformation into *C. reinhardtii*

All DNA and RNA manipulations were carried out using standard methods previously described [61]. The codon-optimized Rc-KASII gene was cloned into a chloroplast expression vector, in which the insert is under the control of the 5'UTR and promoter sequence for the *psbD* gene and 3'UTR for the *psbA* gene. The kanamycin resistance gene (*aphA6*) is regulated by the 5'UTR and promoter sequence for *atpA* and 3'UTR for the *rbcL* gene. The transgene cassette is flanked by 5' and 3' homology regions that target the entire cassette to the 3HB locus of *C. reinhardtii.*

Transformation of plant KS (Rc-KSII) plasmid into *C. reinhardtii 137c (mt+)*

C. reinhardtii cells were grown to late log phase (7 days) in TAP media in the presence of 0.5 mM 5-fluorodeoxyuridine at 23 °C, under constant illumination of 450 lux on a rotary shaker at 100 rpm. Cells (50 ml) were harvested by centrifugation (4000 x g, 23 °C, 5 minutes) and the supernatant was decanted and the cells resuspended in 4 ml TAP media prior to transformation by particle bombardment [62]. All transformations were performed under kanamycin (100 μg/ml) selection.
PCR was used to identify the transformed strains (Figure 3-26). For PCR analysis, 10^6 cells from plate or liquid culture are suspended in 10 mM EDTA and heated to 95 °C for 10 minutes and then cooled to 23 °C. This mixture was added to a typical PCR cocktail, containing $MgCl_2$, dNTPs, PCR primer pairs, DNA polymerase and water. $MgCl_2$ and EDTA concentrations were carefully controlled. Additionally, genomic DNA (extracted with a Promega Genomic DNA purification kit) was used as a template if crude algal lysates yielded unclear results.

To identify strains containing the plant ketosynthase gene, a primer pair was used in which one anneals to a site in the 5'UTR and the other to the ketosynthase coding segment. Desired clones are those that yield the desired size PCR fragment. To determine the degree to which the endogenous gene locus has been displaced (heteroplasmic versus homoplasmic), a PCR reaction containing two sets of primers was used. The first pair amplifies the endogenous locus targeted by the expression vector. The second pair amplifies a constant, or control, region that is not targeted by the expression vector. The number of cycles was increased to >30 to enhance

sensitivity. The most desired clones are those that yield a product from the constant region but not from the endogenous region.

To confirm expression of the plant ketosynthase (Rc-KSII) protein in transgenic *C. reinhardtii* strains, Western blot analysis was performed (Figure 3-27). $5x10^8$ cells were collected by centrifugation from TAP media by centrifugation. Cells were resuspended in cold lysis buffer (25 mM Tris, 137 mM NaCl, 2.7 mM KCl, pH 7.4) supplemented with an EDTA-free protease inhibitor cocktail (Roche). Cells were lysed by sonication, followed by centrifugation. Proteins were purified by FLAG affinity chromatography at 4 °C and the protein eluted off the resin using 0.1 M glycine pH 3.0. Proteins are separated by SDS-PAGE and transferred to PVDF membrane. The membrane is blocked with blocking solution for 30 minutes at room temperature, followed by incubation with HRP-anti-FLAG antibody for 12h at 4 °C, washed, and visualized by HRP-substrate and chemiluminescent detection [63].

***Chlamydomonas reinhardtii* growth and fatty acid analysis**

Chlamydomonas reinhardtii strains were maintained on TAP media [64], supplemented with 40 μg/ml carbendazim, 500 μg/ml ampicillin and 100 μg/ml cefotaxime (TAP+C+A+C) [65] (and in the case of the plant KS, an additional 100 μg/ml kanamycin (+K)), agar plates under constant illumination. Sterile 150 ml flasks containing 50 ml TAP+C+A+C (+K) media were inoculated with single colonies and grown under constant shaking in a greenhouse. Cell density of the cultures was measured. After 3 days, cultures were centrifuged at 4000 rpm at 4 °C, and the cell pellet resuspended in 2 ml of a 1 M methanolic acid solution. The suspension was incubated for 30 minutes at 65 °C and the fatty acid methyl esters extracted using 2 ml

of hexanes. The fatty acid composition was determined by GC/MS analysis (Figure 3-28) on an Agilent 7890A GC system connected to a 5975C VL MSD quadrupole MS (EI). Helium was used as a carrier gas. A method employing a gradient of 110 °C to 200 °C at 15 °C /min followed by 20 minutes at 200 °C on a 60 meter DB23 column was used to accurately measure fatty acid methyl ester content and composition.

References

1. Cooper JE (1962) Fatty acids in recent and ancient sediments and petroleum reservoir waters. Nature 193: 744-746.

2. Montgomery R (2004) Development of biobased products. Bioresource Technology 91: 1-29.

3. Sheehan J, Dunahay T, Benemann J, Roessler P (1998) A Look Back at the U.S. Department of Energy's Aquatic Species Program—Biodiesel from Algae. Golden, Colorado: National Renewable Energy Laboratory.

4. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. Plant Journal 54: 621-639.

5. Chisti Y (2007) Biodiesel from microalgae. Biotechnology Advances 25: 294-306.

6. Knothe G, Van Gerpen J, J K (2005) The Biodiesel Handbook. AOCS Press: Champaign, Illinois.

7. Radakovits R, Jinkerson RE, Darzins A, Posewitz MC (2010) Genetic engineering of algae for enhanced biofuel production. Eukaryotic Cell 9: 486-501.

8. Radakovits R, Eduafo PM, Posewitz MC (2011) Genetic engineering of fatty acid chain length in *Phaeodactylum tricornutum*. Metabolic Engineering 13: 89-95.

9. Blatti JL, Beld J, Behnke C, Mendez M, Mayfield S, Burkart MD (2012) Manipulating fatty acid biosynthesis in microalgae for biofuel through protein-protein interactions. Submitted.

10. Dunahay TG, Jarvis EE, Roessler PG (1995) Genetic transformation of the diatoms *Cyclotella cryptica* and *Navicula saprophila*. Journal of Phycology 31: 1004-1012.

11. Gebicki JM, Hicks M (1973) Ufasomes are stable particles surrounded by unsaturated fatty acid membranes. Nature 243: 232-234.

12. Gebicki JM, Hicks M (1976) Preparation and properties of vesicles enclosed by fatty acid membranes. Chemistry and Physics of Lipids 16: 142-160.

13. Hargreaves WR, Deamer DW (1978) Liposomes from ionic, single-chain amphiphiles. Biochemistry 17: 3759-3768.

14. Apel CL, Deamer DW, Mautner MN (2002) Self-assembled vesicles of monocarboxylic acids and alcohols: conditions for stability and for the encapsulation of biopolymers. Biochimica et Biophysica Acta - Biomembranes 1559: 1-9.

15. Deamer D, Dworkin JP, Sandford SA, Bernstein MP, Allamandola LJ (2002) The first cell membranes. Astrobiology 2: 371-381.

16. Zornetzer GA, Fox BG, Markley JL (2006) Solution structures of spinach acyl carrier protein with decanoate and stearate. Biochemistry 45: 5217-5227.

17. Weissman KJ, Mueller R (2008) Protein-protein interactions in multienzyme megasynthases. ChemBioChem 9: 826-848.

18. Ohlrogge J, Browse J (1995) Lipid biosynthesis. Plant Cell 7: 957-970.

19. Koo AJK, Ohlrogge JB, Pollard M (2004) On the Export of Fatty Acids from the Chloroplast. Journal of Biological Chemistry 279: 16101-16110.

20. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, et al. (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318: 245-251.

21. Gong Y, Guo X, Wan X, Liang Z, Jiang M (2011) Characterization of a novel thioesterase (PtTE) from *Phaeodactylum tricornutum*. Journal of Basic Microbiology 51: 666-672.

22. Salas JJ, Ohlrogge JB (2002) Characterization of substrate specificity of plant FatA and FatB acyl-ACP thioesterases. Archives of Biochemistry and Biophysics 403: 25-34.

23. Graham SA (1989) Cuphea - a new source of medium-chain fatty acids. Critical Reviews in Food Science and Nutrition 28: 139-173.

24. Dehesh K, Jones A, Knutzon DS, Voelker TA (1996) Production of high levels of 8:0 and 10:0 fatty acids in transgenic canola by overexpression of Ch FatB2, a thioesterase cDNA from *Cuphea hookeriana*. Plant Journal 9: 167-172.

25. Voelker T (1996) Plant acyl-ACP-thioesterases: chain length determining enzymes in plant fatty acid biosynthesis. In: Setlow JK (ed) Genetic Engineering, Vol. 18. Plenum Press: New York, pp. 111-133.

26. Thelen JJ, Ohlrogge JB (2002) Metabolic engineering of fatty acid biosynthesis in plants. Metabolic Engineering 4: 12-21.

27. Voelker TA, Davies HM (1994) Alteration of the specificity and regulation of fatty acid synthesis of *Escherichia coli* by expression of a plant medium-chain acyl-acyl carrier protein thioesterase. Journal of Bacteriology 176: 7320-7327.

28. Dormann P, Kridl JC, Ohlrogge JB (1994) Cloning and expression in *Escherichia coli* of a cDNA coding for the oleoyl-acyl carrier protein thioesterase from coriander (*Coriandrum sativum*). Biochimica et Biophysica Acta - Lipids and Lipid Metabolism 1212: 134-136.

29. Dormann P, Voelker TA, Ohlrogge JB (1995) Cloning and expression in *Escherichia coli* of a novel thioesterase from *Arabidopsis thaliana* specific for long-chain acyl-acyl-carrier proteins. Archives of Biochemistry and Biophysics 316: 612-618.

30. Cahoon EB, Mills LA, Shanklin J (1996) Modification of the fatty acid composition of *Escherichia coli* by coexpression of a plant acyl-acyl carrier protein desaturase and ferredoxin. Journal of Bacteriology 178: 936-939.

31. Rehm BHA, Steinbüchel A (2001) Heterologous expression of the acyl-acyl carrier protein thioesterase gene from the plant *Umbellularia californica* mediates polyhydroxyalkanoate biosynthesis in recombinant *Escherichia coli*. Applied Microbiology and Biotechnology 55: 205-209.

32. Huynh TT, Pirtle RM, Chapman KD (2002) Expression of a *Gossypium hirsutum* cDNA encoding a FatB palmitoyl-acyl carrier protein thioesterase in *Escherichia coli*. Plant Physiology and Biochemistry 40: 1-9.

33. Liu X, Sheng J, Curtiss III R (2011) Fatty acid production in genetically modified cyanobacteria. Proceedings of the National Academy of Sciences, USA 108: 6899-6904.

34. Zhang X, Li M, Agrawal A, San KY (2011) Efficient free fatty acid production in Escherichia coli using plant acyl-ACP thioesterases. Metabolic Engineering 13: 713-722.

35. Rock CO, Cronan JE (1996) *Escherichia coli* as a model for the regulation of dissociable (type II) fatty acid biosynthesis. Biochimica et Biophysica Acta - Lipids and Lipid Metabolism 1302: 1-16.

36. Worthington AS, Rivera H, Torpey JW, Alexander MD, Burkart MD (2006) Mechanism-based protein cross-linking probes to investigate carrier protein-mediated biosynthesis. ACS Chemical Biology 1: 687-691.

37. Weissman KJ, Hong H, Popovic B, Meersman F (2006) Evidence for a protein-protein interaction motif on an acyl carrier protein domain from a modular polyketide synthase. Chemistry and Biology 13: 625-636.

38. Tran L, Broadhurst RW, Tosin M, Cavalli A, Weissman KJ (2010) Insights into protein-protein and enzyme-substrate interactions in modular polyketide synthases. Chemistry and Biology 17: 705-716.

39. Magnuson K, Jackowski S, Rock CO, Cronan Jr JE (1993) Regulation of fatty acid biosynthesis in *Escherichia coli*. Microbiological Reviews 57: 522-542.

40. Ohlrogge JB, Jaworski JG (1997) Regulation of fatty acid synthesis. Annual Review of Plant Physiology and Plant Molecular Biology 48: 109-136.

41. Reyes-Prieto A, Weber APM, Bhattacharya D (2007) The origin and establishment of the plastid in algae and plants. Annual Review of Genetics 41: 147-168.

42. Ryall K, Harper JT, Keeling PJ (2003) Plastid-derived type II fatty acid biosynthetic enzymes in chromists. Gene 313: 139-148.

43. Jones A, Davies HM, Voelker TA (1995) Palmitoyl-acyl carrier protein (ACP) thioesterase and the evolutionary origin of plant acyl-ACP thioesterases. Plant Cell 7: 359-371.

44. Jiang C, Kim SY, Suh D-Y (2008) Divergent evolution of the thiolase superfamily and chalcone synthase family. Molecular Phylogenetics and Evolution 49: 691-701.

45. Chen Y, Kelly EE, Masluk RP, Nelson CL, Cantu DC, Reilly PJ (2011) Structural classification and properties of ketoacyl synthases. Protein Science 20: 1659-1667.

46. Siggaard-Andersen M (1993) Conserved residues in condensing enzyme domains of fatty acid synthases and related sequences. Protein Sequences and Data Analysis 5: 325-335.

47. Armougom F, Moretti S, Poirot O, Audic S, Dumas P, Schaeli B, Keduas V, Notredame C (2006) Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-coffee. Nucleic Acids Research 34: W604-W608.

48. Edgar RC (2004) MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797.

49. Gouet P, Courcelle E, Stuart DI, Metoz F (1999) ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics 15: 305-308.

50. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Research 36: W465-W469.

51. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754-755.

52. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28: 2731-2739.

53. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22: 195-201.

54. Comeau SR, Gatchell DW, Vajda S, Camacho CJ (2004) ClusPro: An automated docking and discrimination method for the prediction of protein complexes. Bioinformatics 20: 45-50.

55. PyMol (v1.4) Schrödinger, LLC.

56. Murugan E, Kong R, Sun H, Rao F, Liang Z-X (2010) Expression, purification and characterization of the acyl carrier protein phosphodiesterase from *Pseudomonas aeruginosa*. Protein Expression and Purification 71: 132-138.

57. Worthington AS, Rivera H, Jr., Torpey JW, Alexander MD, Burkart MD (2006) Mechanism-based protein crosslinking probes to investigate carrier protein-mediated biosynthesis. ACS Chemical Biology 1: 687-691.

58. Quadri LEN, Weinreb PH, Lei M, Nakano MM, Zuber P, Walsh CT (1998) Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. Biochemistry 37: 1585-1595.

59. De Lay NR, Cronan Jr JE (2007) In vivo functional analyses of the type II acyl carrier proteins of fatty acid biosynthesis. Journal of Biological Chemistry 282: 20319-20328.

60. Volkmann G, Murphy PW, Rowland EE, Cronan Jr JE, Liu XQ, Blouin C, Byers DM (2010) Intein-mediated cyclization of bacterial acyl carrier protein stabilizes its folded conformation but does not abolish function. Journal of Biological Chemistry 285: 8605-8614.

61. Mayfield SP, Franklin SE, Lerner RA (2003) Expression and assembly of a fully active antibody in algae. Proceedings of the National Academy of Sciences, USA 100: 438-442.

62. Boynton JE, Gillham NW, Harris EH, Hosler JP, Johnson AM, Jones AR, Randolph-Anderson BL, Robertson D, Klein TM, Shark KB, et al. (1988) Chloroplast transformation in Chlamydomonas with high velocity microprojectiles. Science 240: 1534-1538.

63. Mendez M, O'Neill B, Burkart MD, Behnke C, Lieberman S, Bielinski V, Poon Y (2010) Production of fatty acids by genetically modified photosynthetic organisms. US Patent No. WO 2010/019813.

64. Gorman DS, Levine RP (1965) Cytochrome F and plastocyanin - their sequence in photosynthetic electron transport chain of *Chlamydomonas reinhardtii*. Proceedings of the National Academy of Sciences, USA 54: 1665-1669.

65. Kan Y, Pan J (2010) A one-shot solution to bacterial and fungal contamination in the green alga *Chlamydomonas reinhardtii* culture by using an antibiotic cocktail. Journal of Phycology 46: 1356-1358.

66. Worthington AS, Porter DF, Burkart MD (2010) Mechanism-based crosslinking as a gauge for functional interaction of modular synthases. Organic and Biomolecular Chemistry 8: 1769-1772.

67. Gressel J (2008) Transgenics are imperative for biofuel crops. Plant Science 174: 246-263.

68. Liu T, Khosla C (2010) Genetic engineering of *Escherichia coli* for biofuel production. Annual Reviews Genetics 44: 53-69.

69. Hampl V, Stairs CW, Roger AJ (2011) The tangled past of eukaryotic enzymes involved in anaerobic metabolism. Mobile Genetic Elements 1: 71-74.

70. Horner DS, Heil B, Happe T, Embley TM (2002) Iron hydrogenases - Ancient enzymes in modern eukaryotes. Trends in Biochemical Sciences 27: 148-153.

71. Meyer J (2007) [FeFe] hydrogenases and their evolution: A genomic perspective. Cellular and Molecular Life Sciences 64: 1063-1084.

72. Ludwig M, Schulz-Friedrich R, Appel J (2006) Occurrence of hydrogenases in cyanobacteria and anoxygenic photosynthetic bacteria: Implications for the phylogenetic origin of cyanobacterial and algal hydrogenases. Journal of Molecular Evolution 63: 758-768.

73. Yu XH, Rawat R, Shanklin J (2011) Characterization and analysis of the cotton cyclopropane fatty acid synthase family and their contribution to cyclopropane fatty acid synthesis. BMC Plant Biology 11: 97.

74. Spencer AK, Greenspan AD, Cronan Jr JE (1978) Thioesterases I and II of *Escherichia coli*. Hydrolysis of native acyl-acyl carrier protein thioesters. Journal of Biological Chemistry 253: 5922-5926.

75. Lee YL, Lee LC, Shaw JF (2012) Multifunctional enzyme thioesterase I/protease I/lysophospholipase L 1 of *Escherichia coli* shows exquisite structure for its substrate preferences. Biocatalysis and Agricultural Biotechnology 1: 95-104.

76. Lo YC, Lin SC, Shaw JF, Liaw YC (2005) Substrate specificities of *Escherichia coli* thioesterase I/protease I/lysophospholipase L1 are governed by its switch loop movement. Biochemistry 44: 1971-1979.

77. Yu X, Liu T, Zhu F, Khosla C (2011) In vitro reconstitution and steady-state analysis of the fatty acid synthase from *Escherichia coli*. Proceedings of the National Academy of Sciences, USA 108: 18643-18648.

78. Liu T, Vora H, Khosla C (2010) Quantitative analysis and engineering of fatty acid biosynthesis in *E. coli*. Metabolic Engineering 12: 378-386.

79. Voelker TA, Worrell AC, Anderson L, Bleibaum J, Fan C, Hawkins DJ, Radke SE, Davies HM (1992) Fatty acid biosynthesis redirected to medium chains in transgenic oilseed plants. Science 257: 72-74.

80. Jing F, Cantu DC, Tvaruzkova J, Chipman JP, Nikolau BJ (2011) Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. BMC Biochemistry 12: 44.

81. Cheng T, Yang J, Liu H, Zhang Y, Song Q, Xian M (2011) Expression of *Arabidopsis thaliana* thioesterase gene (*atfata*) in *Escherichia coli* and its influence on biosynthesis of free fatty acid. Chinese Journal of Applied and Environmental Biology 17: 568-571.

82. Garwin JL, Klages AL, Cronan Jr JE (1980) β-Ketoacyl-acyl carrier protein synthase II of *Escherichia coli*. Evidence for function in the thermal regulation of fatty acid synthesis. Journal of Biological Chemistry 255: 3263-3265.

83. Flugel RS, Hwangbo Y, Lambalot RH, Cronan Jr JE, Walsh CT (2000) Holo-(Acyl Carrier Protein) Synthase and Phosphopantetheinyl Transfer in *Escherichia coli*. Journal of Biological Chemistry 275: 959-968.

84. Davis MS, Cronan Jr JE (2001) Inhibition of *Escherichia coli* acetyl coenzyme A carboxylase by acyl-acyl carrier protein. Journal of Bacteriology 183: 1499-1503.

85. Clark DP, DeMendoza D, Polacco ML, Cronan Jr JE (1983) β-Hydroxydecanoyl thio ester dehydrase does not catalyze a rate-limiting step in *Escherichia coli* unsaturated fatty acid synthesis. Biochemistry 22: 5897-5902.

86. Subrahmanyam S, Cronan Jr JE (1998) Overproduction of a functional fatty acid biosynthetic enzyme blocks fatty acid synthesis in *Escherichia coli*. Journal of Bacteriology 180: 4596-4602.

87. Kaewsuwan S, Bunyapraphatsara N, Cove DJ, Quatrano RS, Chodok P (2010) High level production of adrenic acid in *Physcomitrella patens* using the algae Pavlova sp. Δ5-elongase gene. Bioresource Technology 101: 4081-4088.

88. Eddy SR (2004) Where did the BLOSUM62 alignment score matrix come from? Nature Biotechnology 22: 1035-1036.

89. Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a unified platform for automated protein structure and function predicition. Nature Protocols 5: 725-738.

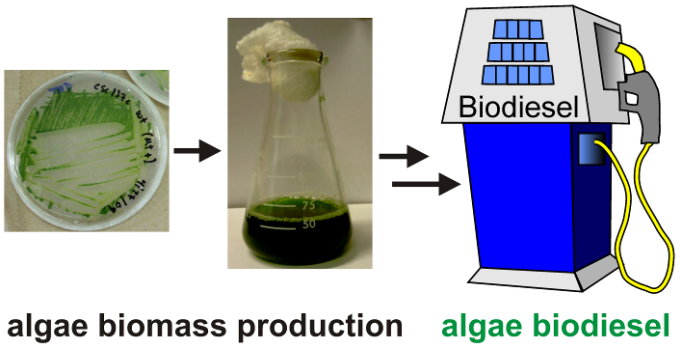
Chapter 3 is adapted from a manuscript currently being prepared for publication in Algal Research, on which I am the primary author. Joris Beld assisted with the *in silico* work and *in vitro* crosslinking experiments. Craig Behnke and Mike Mendez assisted with engineering of KS into the *C. reinhardtii* chloroplast. Professor John Cronan provided *E. coli* strain C1877 and Professor Byers supplied the *Vibrio harveyi* ACP construct.

Chapter 4 : Chemical Education: an exercise in inquiry-based learning

Releasing stored solar energy within pond scum: biodiesel from algal lipids

Abstract

Microalgae have emerged as an attractive feedstock for the mass production of renewable transportation fuels due to their fast growth rate, flexible habitat preferences, and substantial oil yields. To develop a laboratory as an educational tool that mimics emerging algal biofuel technology, we describe the extraction of algal lipids and explore their transesterification into biodiesel. Students are first introduced to algae as an energy feedstock and taught how to cultivate and harvest algal biomass. Next, students learn how to convert algal biomass to a transportation fuel using transesterification and extraction, including an experimental comparison of base and acid catalyzed transesterification methods. The energy content of homemade algae biodiesel is determined and compared to soybean biodiesel and petroleum diesel using a homemade calorimeter.



Introduction

Dwindling petroleum reserves and climate change caused by combustion-related CO_2 emissions illustrate the necessity for a sustainable, carbon-neutral fuel source. Non-renewable petroleum hydrocarbons currently meet 37% of energy needs in the United States [1], and transportation is the second largest consumer of energy in the US, accounting for nearly 60% of our nation's petroleum use, an amount equivalent to all of the oil imported into the country [1]. Therefore, creating a renewable transportation fuel such as biodiesel has the potential to dramatically impact the future energy outlook of the US [2].

Biodiesel is a high energy, lightweight fuel derived from lipids and oils through a chemical transformation called *transesterification.* In terms of energy transport and storage, biodiesel offers several advantages over competing transportation fuels such as hydrogen [3] and corn ethanol. Biodiesel has an energy density comparable to that of petroleum, while emitting less greenhouse gases, and if produced by a carbon-sequestering organism such as plants or algae, it can be rapidly produced on scale in a carbon-neutral and sustainable fashion.

Microalgae have emerged as an attractive feedstock for the mass production of renewable transportation fuels due to their fast growth rate, flexible habitat preferences, and substantial oil yields [4]. Algae efficiently store solar energy and sequester CO_2 during photosynthesis, alleviating climate change caused by the emission of greenhouse gases. Biodiesel created from microalgae does not require fertile land, the main drawback of corn-based ethanol and plant oils [5]. In fact, algae can grow in a myriad of environments, including brackish water, wastewater, near

power plants, and in both extreme cold and warm environments. In addition to their fast growth rate, microalgae are capable of producing 1000-4000 gallons of oil/acre/year [6], a much higher yield than the 48 gallons of oil/acre/year from soybeans, currently the leading biodiesel feedstock [6]. Microalgae store excess energy as fatty acid triacylglycerides (TAGs) in lipid bodies, which can be liberated and converted to their methyl ester derivatives via transesterification to synthesize biodiesel. Last discussed in *JCE* in 1985 to demonstrate the isolation of natural products [7], algae have recently attracted wide interest due to their potential as a biofuel feedstock. As new algal technology emerges, education will be necessary to sustain the growing industry [8].

The production of algae biodiesel can be used to visualize chemical and biological concepts and techniques in a classroom setting. This laboratory aims to expose students to important facets of algal biofuel technology. Students are introduced to algae as an energy feedstock and taught how to cultivate and harvest algal biomass. Next, students learn how to convert algal biomass to a transportation fuel using transesterification and extraction. Finally, the energy content of homemade algae fuel is determined using calorimetry.

Experimental Overview

This laboratory was conducted over a semester in an introductory pre-college science course. The first part of the lab involved the culturing and scale-up of model algal species *Chlamydomonas reinhardtii* over a six-week period, followed by harvesting algal biomass. The second part of the laboratory involved the synthesis,

characterization, and utilization of algae biodiesel. Since industrial scale biodiesel production utilizes base catalysis due to rapid reaction rates and high yields, base-catalyzed transesterification of soybean oil was carried out and compared to a recently developed acid-catalyzed method that converts lipids within algal biomass to biodiesel [9]. Students quantified the amount of biodiesel obtained from algae and compared it to that prepared from soybean oil, a plant-based feedstock. Energy content of each fuel was measured using calorimetry and compared to petroleum diesel. Finally, students were able to observe the efficacy of their homemade algae biodiesel, combusting it to power a diesel engine.

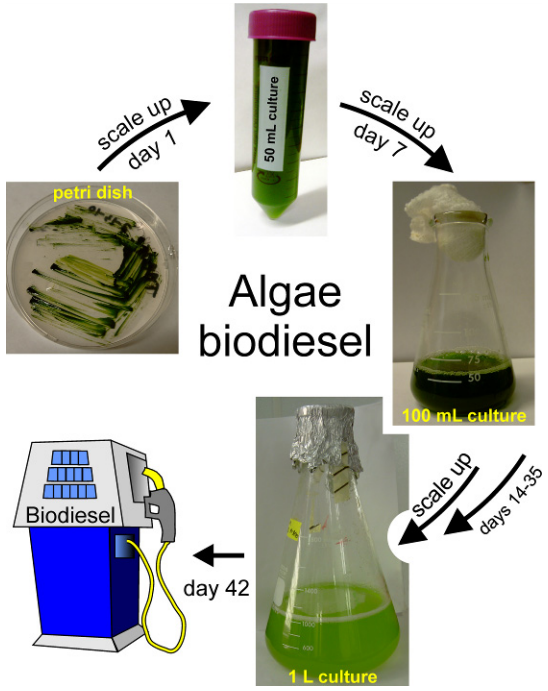


Figure 4-1 Timeline of algal growth

Algal biomass was cultivated over a 6-week period, beginning from a single algal colony on a TAP-agar plate (Day 1). Students monitored the algal culture daily and when cells were grown to high density, they were used to inoculate a larger volume of TAP media. The process continued until students had obtained a culture volume of 1 L. Centrifugation yielded a concentrated algal biomass pellet. Acid catalyzed esterification converted algal lipids to biodiesel.

Experimental Details

Algal culturing, timeline, and harvesting of biomass

Chlamydomonas reinhardtii strain *csc137c* was grown initially on an agar plate enriched with Tris-Acetate-Phosphate media (TAP) [10]. A sterile 50 mL liquid TAP starter culture was inoculated from a single, robust algal colony. After 7 days of agitation under constant white light illumination, the algal culture was used to inoculate a flask containing a larger volume of media. Cells were inoculated every 7 days or when the culture became a dark green color. This process was repeated until a 1 L culture was obtained (Figure 4-1). When 1 L of a dense, dark green *C. reinhardtii* culture was observed, algal biomass was harvested via centrifugation. The algal paste was transferred to a 50 mL conical tube and stored at -20 °C.

Acid-catalyzed transesterification of algal biomass

To convert algal lipids to their methyl ester derivatives [9], 10 mL of a 1 M methanolic HCl solution was added to the dry algal biomass pellet. The algae cell pellet was resuspended to homogeneity in the methanolic acid solution and the mixture incubated at 65 °C for 30 minutes (Figure 4-2).

Liquid-liquid extraction

Fatty acid methyl esters were extracted from the mixture using hexanes. 2 mL of hexanes was added directly to the reaction vessel immediately upon cooling and the mixture was vortexed or shaken vigorously for one minute. To break emulsions, a one minute centrifugation step was carried out. The top organic layer was removed and transferred to a pre-weighed vial. The hexane extraction was repeated six times, each

with 2 mL hexanes. The hexane extract was analyzed for fatty acid methyl ester content and composition by GCMS (Figure 4-3).

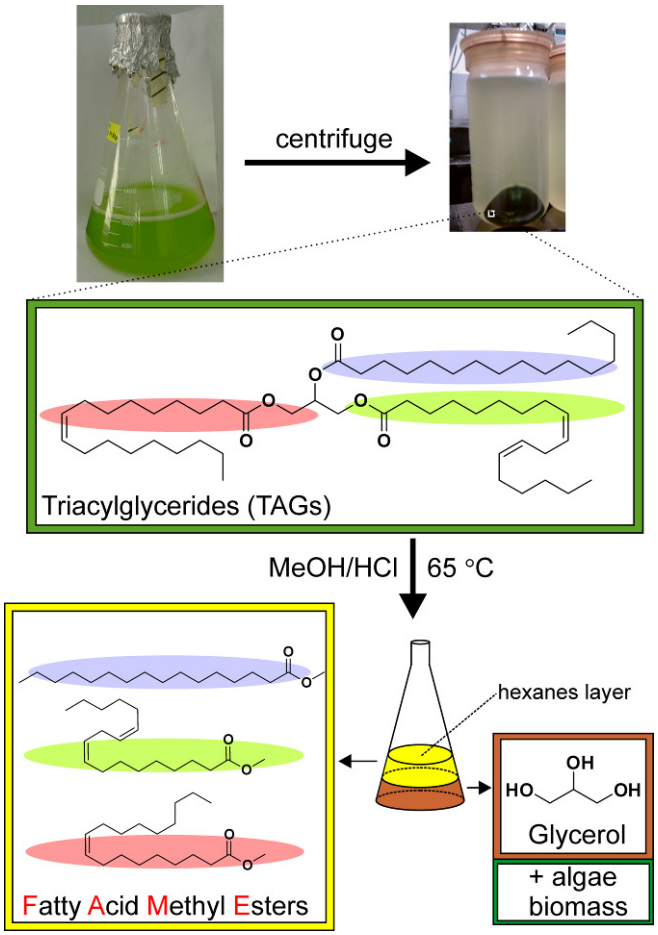


Figure 4-2 Biodiesel from algal lipids

After harvesting algal biomass, acid-catalyzed transesterification was performed directly on concentrated algal biomass to convert triacylglycerides to their methyl ester derivatives (biodiesel). Fatty acid methyl esters were extracted from the mixture with hexanes to afford biodiesel.

Base-catalyzed transesterification of soybean oil and soybean biomass

To compare methods and feedstocks, base-catalyzed transesterification of soybean oil was performed [11] [12] [13] [14] [15] [16]**.** The temperature of the

stir/hotplate was raised to 65 °C and 5 mL of soybean oil was added to a flask containing 5 mL of a sodium methoxide solution. The reaction proceeded for 30 minutes, stirring magnetically at 65 °C. The biodiesel was washed with distilled water to remove all basic residue.

In parallel, base-catalyzed transesterification was carried out directly on soybean biomass as a more equivalent comparison to algae biomass. 2 g of soybean biomass was added to a 50 mL conical tube and resuspended in 5 mL sodium methoxide. The reaction was placed in a heat bath set to 65 °C for 30 minutes. Biodiesel was extracted from the mixture using 12 mL hexanes (6 X 2 mL extractions).

Calorimetry

A calorimeter was built from simple materials [17] to measure the energy density of three fuels: algal biodiesel, plant biodiesel, and petroleum diesel. In parallel, the length of time that each fuel remained lit in an ethanol lamp was recorded (Table 4-1).

Viscosity

The viscosity of each biodiesel fuel was measured and compared to the oil from which it was synthesized using a Pasteur pipette and bulb as a makeshift viscometer (Table 4-2). The time each fuel drained from the pipette was measured and recorded.

Gas Chromatography / Mass Spectrometry (GCMS)

The composition of each fuel was determined using GCMS. Soybean and algae biodiesel were compared to commercially available B-20 biodiesel.

Results

During the first part of this laboratory, students grew and cultivated algae biomass over a six-week period beginning from a single colony of model algae *Chlamydomonas reinhardtii*. This required daily monitoring in which students took detailed notes and observations about the color of the algae, the temperature, and the light source. Students made sure the cells were agitated, taking turns to shake the flasks each day. At the end of six weeks, students had acquired a substantial amount of algal biomass, approximately 2 g/L of culture. Using a homemade centrifuge from a salad spinner, students harvested one liter of algae biomass via centrifugation and prepared it for a one-step conversion to biodiesel.

The second part of this laboratory focused on the synthesis and characterization of biodiesel from algal biomass. Direct acid-catalyzed transesterification of 2 g of algae biomass yielded 100 mg of biodiesel. In parallel, base-catalyzed transesterification of soybean oil and soybean biomass afforded 12% and 6% soybean biodiesel, respectively. Students built a calorimeter, burned each fuel, and measured heat released. Both biodiesel from algae and soybean biomass gave comparable energy changes to that of conventional diesel (Table 4-1). Compared to petroleum diesel, biodiesel burned cleaner, longer, and yielded more heat energy. In the qualitative energy density test, each fuel was burned in an ethanol lamp and the length of time the wick stayed lit measured (Table 4-1).

Table 4-1 Energy density comparison of algal biodiesel, plant biodiesel and diesel

Biodiesel Source	Algae	Soybean (Oil)	Soybean (biomass)	Diesel
Yield of biodiesel produced from 2 g biomass (%)	5	12	6	NA
Energy Density (temp. change, °C)[a]	12±3	13±3	11±2	16±3
Time fuel burned in EtOH lamp (min)	19	22	20	21
Heat released upon combustion (kJ)[b]	5.0	5.4	4.6	6.7

[a] Final water temperature – initial water temperature; [b] $q_v = C \times dT \times m$,
where q_v = heat flow, or change in internal energy (J), C = heat capacity of water, C = 4.18 J/ g · °C,
dT = temperature change = $T_f - T_i$ (°C), m = mass of water (g). n = 50 students.

Table 4-2 Viscosity measurements of synthesized biodiesel fuels and comparison to oil standards

Fuel/ Oil	Viscosity
Motor oil	6:21 min
Soybean oil	3:21 min
Soybean biodiesel (oil)	0:26 min
Soybean biodiesel (biomass)	0:03 min
Algae biodiesel	0:03 min
Standard biodiesel	0:13 min
Diesel	0:11 min

Viscosity measurements and GCMS analysis confirmed that indeed biodiesel was synthesized from soybean oil, soybean biomass, and algae biomass (Table 4-2 and Figure 4-3).

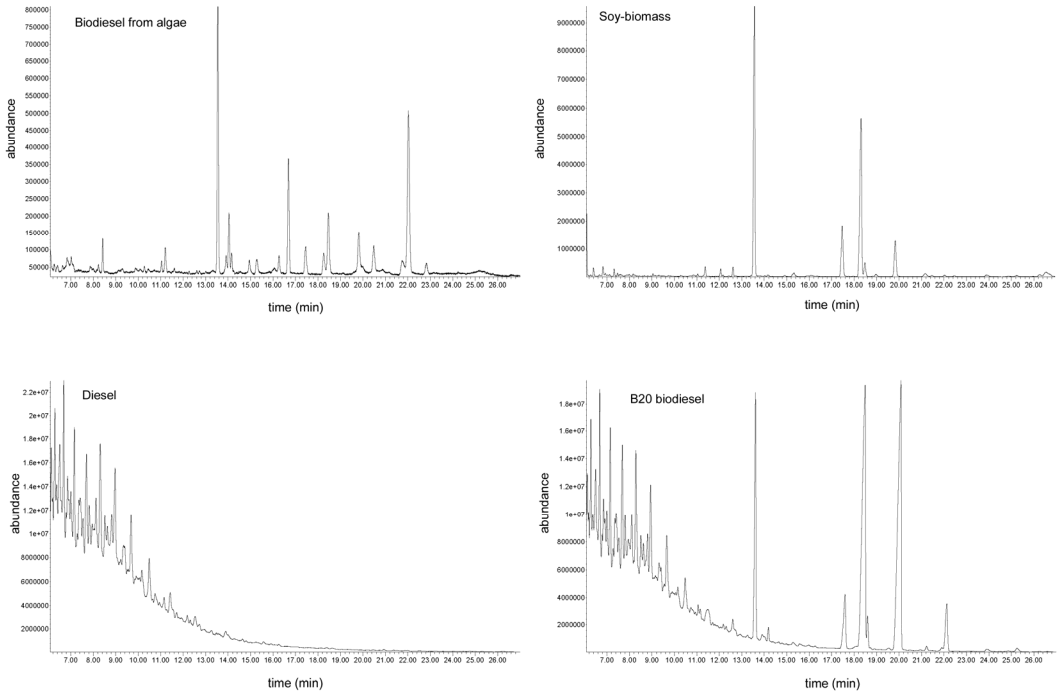


Figure 4-3 GC/MS chromatograms: validation of biodiesel synthesis

Discussion

This laboratory introduces students to the current research of algal biofuel technology and teaches basic chemical and biological techniques involved. This series of experiments is unique and powerful in that students are able to observe and play an active role in the entire flow of energy, emphasizing the Law of Conservation of Energy. Students are trained in the art of culturing algae, as solar energy is harvested by algae via photosynthesis. Students learn the chemical process by which triacylglycerides are converted into biodiesel as they extract the chemical energy stored within the algal biomass through acid-catalyzed transesterification. At the end of the project, students can use the algal fuel to power a small propeller equipped with

a diesel engine, demonstrating the conversion of chemical energy to mechanical energy.

Biodiesel from soybean is commercially produced using catalytic base since it is faster than the acid-catalyzed pathway, requires lower temperatures and pressures, and is the most economical process to treat vegetable oils [18]. However, base-catalysis produces soap as a side reaction, reducing the overall yield of biodiesel and complicating the subsequent purification process [19]. Soap formation can be avoided using strong acid catalysts to convert free fatty acids into their corresponding methyl esters through acid-mediated esterification. It has recently been shown that fatty acid methyl esters can be recovered in high yields through direct acid-catalyzed methanolysis of dried microbial biomass [9]. Therefore, traditional base-catalyzed methodology was used to synthesize biodiesel from soybean oil and the more recent direct methanolysis procedure was conducted on algae biomass. By demonstrating both acid and base-catalyzed transesterification, students realize that the same biodiesel product can be obtained through two different chemical pathways.

Commercial production of algal biodiesel is in its infancy. Following hexane extraction and evaporation, algal biodiesel can mixed with petroleum diesel, such as in current B-20 biodiesel blends, or it can be deoxygenated to produce conventional diesel in a process called hydrotreatment. Algal biodiesel can also be used as a feedstock for traditional petroleum refineries, allowing it to be easily incorporated into the existing infrastructure [20].

Conclusion

Exposing the next generation of scientists and citizens to renewable energy issues is critical, especially when they are developing basic inquiry and observation skills. Following completion of this laboratory, students acquire an appreciation of algae biodiesel, develop awareness about current renewable energy issues, and are exposed to future energy prospects. Throughout this exercise, students are provided with many opportunities to formulate new ideas about the concept of energy. These experiments evoke scientific inquiry, foster the development of laboratory skills, and elicit a novel perspective about the energy challenges we face in the future.

Supplemental Information

Instructional materials

A student handout; an instructor protocol and notes; an instructional Powerpoint presentation; full list of materials; protocol for base-catalyzed preparation of biodiesel; protocol for direct acid-catalyzed transesterification of biodiesel from algal biomass; calorimetry design and protocol; salad spinner centrifuge design; TAP media recipe and algae culture conditions is available in the Supplemental Information of *J. Chem. Educ.* 2012, *89* (2), 239-242.

Supplemental Notes

The soybean biodiesel yield synthesized from soybean oil was calculated on a mass-to-mass basis using the known information that soybean biomass contains 20% oil / gram biomass [4]. *Chlamydomonas reinhardtii* strain *csc*137 c, a common laboratory strain of this model species, contains ~20% oil.

The yields reported in the main text for biodiesel synthesized from both algae and soybean biomass represents the yield from 6 X 2 mL hexane extractions. The yield increases if additional hexane extractions are performed. The comparison between the two biomass sources is relative since in both cases the same amount of hexanes was used.

In the calorimetry experiments, the values for combusting hexanes were measured and subtracted from each measured biofuel value.

Experiments were facilitated by an academic collaboration with the San Diego Center for Algal Biotechnology (SD-CAB) [21].

References

1. US Annual Energy Review, Report No. DOE/EIA-0384 (2009). Released: August 19, 2010. US Dept. of Energy, US Primary Energy Consumption by Source and Sector: http://www.eia.doe.gov/emeu/aer/pecss_diagram.html.

2. Balat, H (2010) Prospects of biofuels for a sustainable energy future: a critical assessment. Energy Education Science and Technology Part A – Energy Science and Research 24: 85-111.

3. Eberle U, Felderhoff M, Schuth F (2009) Practical solutions of hydrogen storage. Angewandte Chemie - International Edition 48: 2-25.

4. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A (2008) Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. Plant Journal 54: 621–639.

5. Dismukes GC, Carrieri D, Bennette N, Ananyev G, Posewitz MC (2008) Aquatic phototrophs: efficient alternatives to land-based crops for biofuels. Current Opinion in Biotechnology 19: 235-240.

6. Chisti Y (2007) Biodiesel from microalgae. Biotechnology Advances 25: 294-306.

7. Porter LA (1985) Use of brown algae to demonstrate natural products techniques. Journal of Chemical Education 62: 635-636.

8. US DOE (2010) National Algal Biofuels Technology Roadmap. U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Biomass Program.

9. Liu B, Zhao ZB (2007) Biodiesel production by direct methanolysis of oleaginous microbial biomass. Journal of Chemical Technology and Biotechnology 82: 775-780.

10. Harris E (1989) The Chlamydomonas sourcebook: a comprehensive guide to biology and laboratory use. Academic Press: San Diego, CA.

11. Stout R (2007) Biodiesel from used oil. Journal of Chemical Education 84: 1765.

12. Bucholtz EC (2007) Biodiesel synthesis and evaluation: an organic chemistry experiment. Journal of Chemical Education 84: 296–299.

13. Clarke NR, Casey JP, Brown ED, Oneyma E, Donaghy K (2006) Preparation and viscosity of biodiesel from new and used vegetable oil – an inquiry-based environmental chemistry laboratory. Journal of Chemical Education 83: 257–260.

14. Akers SM, Conkle JL, Thomas SN, Rider KB (2006) Determination of the heat of combustion of biodiesel using bomb calorimetry: a multidisciplinary undergraduate chemistry laboratory. Journal of Chemical Education 83: 260–263.

15. Wagner EP, Koehle MA, Moyle TM, Lambert PD (2010) How green is your fuel? Creation and comparison of automotive biofuels. Journal of Chemical Education 87: 711–713.

16. Bladt D, Murray S, Gitch B, Trout H, Liberko C (2011) Acid-catalyzed preparation of biodiesel from waste vegetable oil: an experiment for the undergraduate organic chemistry laboratory. Journal of Chemical Education 88: 201-203.

17. Brouwer HJ (1991) Small-scale thermochemistry experiment. Journal of Chemical Education 68: A178.

18. Lang X, Dalai AK, Bakhshi NN, Reaney MJ, Hertz PB (2001) Preparation and characterization of bio-diesels from bio-oils. Bioresource Technology 80: 53-62.

19. Alcantra R, Amores J, Canoira L, Fidalgo E, Franco MJ, Navarro A (2000) Catalytic production of biodiesel from soyben oil, using frying oil and tallow. Biomass and Bioenergy 18: 515-527.

20. Scott SA, Davey MP, Dennis JS, Horst I, Howe CJ, Lea-Smith DJ, Smith AG (2010) Biodiesel from algae: Challenges and perspectives. Current Opinion in Biotechnology 21: 277-286.

21. San Diego Center for Algal Biotechnology (SD-CAB)

Chapter 4 is adapted from an article in the Journal of Chemical Education, on which I was the primary and corresponding author. The curriculum was piloted and implemented at Abraham Lincoln High School in San Diego in the classroom of Janelle Javier, and the results published as a laboratory experiment. The project was supported by an NSF-GK12 fellowship (NSF Grant 0742551, Principal Investigator Maarten J. Chrispeels), the Department of Chemistry and Biochemistry at the University of California San Diego, and SD-CAB. Stephen P. Mayfield (University of California, San Diego) provided *C. reinhardtii* strain *csc*137c.

Chapter 5 : Outlook

Conclusions

In efforts to mitigate CO_2 emissions and harness new forms of energy, microalgae have emerged as a promising biodiesel feedstock. Metabolic engineering of algal fatty acid biosynthesis promises to create strains capable of synthesizing large amounts of oil with select charcteristics, while continuing to grow at fast rates. In this thesis work, a biochemical investigation of fatty acid biosynthesis within the algal chloroplast has illuminated distinct features of algal FAS. A novel strategy was employed in which activity-based chemical probes were used to trap protein-protein interactions between algal FAS enzymes, including the KS and TE, with the ACP.

Sequence alignment of the CrTE with known TEs identifies a Cys-His-Asn catalytic triad, and homology modeling illustrates a hydrophobic binding cleft on the surface of the TE. A C16 α-bromopalmitamide analogue was thus envisioned, chemically synthesized, and attached to the algal ACP using *in vitro* chemoenzymatic biosynthesis to form *crypto*-ACP. Indeed, protein-protein interactions triggered a site-specific crosslinking reaction between *crypto*-ACP and CrTE, effectively trapping the ACP-TE complex. In contrast, plant TEs did not result in ACP-TE complex formation due to a lack of intermolecular interactions. Importantly, *in vivo* engineering of TEs into the *C. reinhardtii* chloroplast confirmed *in vitro* results.

Looking to Nature to provide indight into the observed selectivity in ACP-TE interactions, extensive bioinformatic and phylogenetic analyses were carried out to decipher the evolutionary origin of the algal TE. From this work, anaerobic bacteria were identified as the algal TE ancestor, providing an explanation to the observation that eukaryotic TEs function to siphon fatty acids in TE-less prokaryotes, but not in

more closely related algae. As a comparison, as KS domains all derive from a common ancestor, they were found to be permissive across evolutionary diverse species. Evolution is thus a powerful tool to predict compatibility of FAS enzymes across species, and *in silico* modeling, *in vitro* biochemical assays and *in vivo* engineering can be used to elucidate protein-protein interactions between algal fatty acid biosynthetic enzymes. The work presented in this thesis has contributed to our understanding of algal FAS, paving the way to successful algae biofuel engineering.

Outlook

This work has provided insight into the molecular mechanisms of the algal fatty acid synthase with which to guide future engineering. The requirement of protein-protein interactions between the algal ACP and TE in the hydrolysis of fatty acids has been demonstrated using *Chlamydomonas reinhardtii* as a model. Phenomena have been illuminated in these studies, such as the importance of characterizing biosynthetic enzymes in newly sequenced genomes, as algal metabolism is quite distinct from bacteria and plants. Furthermore, these results have shifted the paradigm in algae biofuel engineering. Rather than treating algae as a 'black box' and genetically engineering single (or groups) of enzymes that may result in the desired phenotypic output, rational approaches may result in higher yields. Therefore, tools to predict protein-protein interactions within algal metabolic pathways and ways to control the regulation of algal gene products will be invaluable in rationally designed engineering approaches.

Future Directions

The results presented herein provide the foundation for many intriguing future studies in algal fatty acid metabolism and biochemistry. For instance, it would be interesting to design mutants of the algal ACP and observe how this affects the binding affinity of FAS enzymes. A CrACP mutant has been envisioned that replaces residue 70 with a tryptophan, hypothesizing that it may block the ACP hydrophobic pocket in which fatty acids reside during biosynthesis, resulting in shorter chain length fatty acids. *In vitro* crosslinking assays, *in vivo* complementation assays, isothermal calorimetry, and fluorescence anisotropy may be important tools to deduce the effects of these ACP mutations on intermolecular recognition with FAS enzymes.

The role of targeting sequences in algae is not well understood, and for nuclear expression of fatty acid biosynthetic genes, targeting to the chloroplast is essential. Initial inspection of the Cr-cACP and Cr-mACP reveals ambiguous targeting sequences (see Chapter 2). ACPs can be tagged with fluorescent or affinity labels and mass spectrometry can be used to find the active form of the algal ACP in the chloroplast, which will facilitate engineering of algal FAS.

The NMR solution structure of the CrACP will be an important reference to compare the structure and function of newly sequenced algal ACPs and to examine the dynamics of ACP upon substrate sequestration and FAS protein binding. Studies to determine the crystal structure of the CrACP/CrTE complex are ongoing and will provide valuable information about the residues involved in molecular recognition between the ACP and TE, as well as enzyme-substrate interactions.

The localization of algal FAS proteins can be deciphered using *in vivo* fluorescent approaches, and experiments are underway in the laboratory, along with proteomic profiling of FAS domains.

Furthermore, outdoor, large-scale growth of algae, such as *C. reinhardtii*, has lead to the realization that ecological phenomena drastically affect the metabolite distribution. For example, the fatty acid profile of *C. reinhardtii* changes according to light intensity, nutrient stress, media, growth phase, predators and organisms present in the culture, and other variables. Resulting from this thesis work, a quorum-sensing molecule was isolated from *C. reinhardtii*, which was previously unreported in the literature. Traditionally, *C. reinhardtii* has been cultivated in laboratory conditions using fluorescent light for photosynthesis in TAP media (acetate-containing, originally devised for photosynthetic mutants). The expression of enzymes involved in fatty acid desaturation in coordination with growth conditions is currently under investigation. The ability to manipulate polyunsaturated fatty acid biosynthesis has implications in engineering algae for high-value food products. Ecological studies to elicit the effects of each individual predator/cohabitant have been formulated, which may lead to creative methods aimed at biocontainment and crop protection.

Algae biofuel technology is in its infancy. As genetic tools become available, it would be advantageous to use them to understand the seemingly complex metabolic pathways present in microalgae and how they are regulated. This will eventually facilitate engineering of algae 'super strains' capable of biosynthesizing high yields of designer oils, ultimately for the development of an economical platform for producing renewable and sustainable biofuels using a photosynthetic microalgal biocatalyst.