Microbial Synthesis of Fuel Hydrocarbons: Enzymes and Metabolic Pathways

A DISSERTATION

SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF

THE UNIVERSITY OF MINNESOTA

BY

Janice Alina Frias

IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

Dr. Lawrence P. Wackett, Advisor

January, 2011

ACKNOWLEDGEMENTS

I am very grateful to my advisor, Larry Wackett, for first taking a chance on me and inviting me to join his lab. I am also grateful for his flexibility in allowing me to follow my interest and passion in a project that has so many questions and needs so many answers.

I thank my lab and those that I worked closely with as I explored alkane and alkene production, Jack Richman, Dave Sukovich, Jennifer Seffernick, Jasmine Erickson, and Stephan Cameron. Jack Richman has really enriched my understanding of chemistry and was always interested and willing to help out with a new chemistry challenge. Jasmine Erickson has worked very hard on several aspects of this project on top of her full her undergraduate load. I couldn't have asked for a better lab experience or better lab mates. Everyone is so willing to help out with new techniques or just to listen as I explore the trials and tribulations that make graduate school what it is.

I also thank Carrie Wilmot, Brandon Goblirsch, and the Wilmot lab for the immense training that they offered as I delved into the world of crystallography with no background other than a great interest. Brandon has been extremely patient with me and a terrific teacher.

My committee has been truly helpful during committee meetings, but also helping guide my work and offering advice when I needed someone to talk to. I thank all members: Romas Kaslauskas, Carrie Wilmot, Jeff Gralnick, Michael Sadowsky, and Sharon Murphy.

I am also grateful to Dr. Sharon Murphy and her lab member, Dr. Linda von Weymarn, for allowing me the much needed time on their frequently used HPLC with radioflow detector to make real head way in my project that involved molecules that are very difficult to detect.

I thank the mass spectrometry lab and specifically Tom Krick. I have utilized the lab resources several times during my pathway intermediate explorations and Tom especially has been critical with learning the equipment and brainstorming the many questions of my project.

Thanks to Ken Valentas, Michael Sadowsky, and Hiroshi Takagi for the unforgettable experience of the exchange program through the NIH Biotechnology Training Grant with NAIST in Nara, Japan. This was a terrific cultural experience and a great experience of learning about another lab and being a small part of it for a short time. The students in Dr. Takagi's lab, as well as previous students that have visited Minnesota from NAIST taught me a lot about Japan and helped make my experience wonderful. I also thank the students from the University of Minnesota that I traveled to and throughout Japan with, Kat Volzing, Josh Ochocki, and Chad Satori.

I have also really appreciated all that Darlene Toedter, Tami Jauert, Kristi Lecy, Lori Bubolz, Kerry Ann Hamilton-Rose, Mary Reilly, and Amado Logo have done in supporting functions during my years in graduate school.

DEDICATION

I dedicate this thesis to my husband, Kory Kolvig, and my family and friends. It has been a long and challenging road and I couldn't have made it through without their support.

ABSTRACT

Petroleum is the major source of motor fuels and commodity chemicals in modern society. Petroleum consists of hydrocarbons and is formed largely by abiotic reactions. Hydrocarbons are also biosynthesized by plants, insects, and microbes. Therefore, understanding how hydrocarbons are biosynthesized could lead to a new source of fuels and chemicals. Biohydrocarbons would be superior to ethanol and biodiesel in many ways. To produce biohydrocarbons on large scale, it is necessary to study the genes and enzymes involved in the synthesis of hydrocarbon molecules. This thesis research deals with microorganisms that synthesize alkanes or alkenes.

The bacterium *Vibrio furnissii* M1 was reported to produce considerable quantities of diesel-length alkanes. We obtained the bacterium from the institute in Japan where it was studied. Our studies reached the conclusion that the strain did not produce alkanes. Subsequent research focused on *Micrococcus luteus* ISU and *Stenotrophomonas maltophilia* that were reported to produce C_{25}-C_{31} alkenes. Previous studies indicated that the alkenes derived from fatty acids joined at or near the carboxyl carbon atoms in a process denoted as a head-to-head condensation mechanism. The genes, enzymes, and metabolic intermediates in the head-to-head biosynthetic pathway were not identified in those studies. My research showed that some *Arthrobacter* strains produced head-to-head hydrocarbons and others did not. A putative hydrocarbon biosynthetic gene cluster was identified in the genomes of the strains capable of making hydrocarbons. Research done in *Shewanella*, another organism

identified and verified to produce hydrocarbons, verified experimentally that the gene cluster was indeed involved in long-chain alkene synthesis.

The four genes identified, *oleABCD*, encode for proteins from the following superfamilies: thiolase, α/β-hydrolase, AMP-dependent ligase/synthase, and short chain dehydrogenase, respectively. The OleC protein from *Stenotrophomonas maltophilia* was expressed in *Escherichia. coli* and was purified to homogeneity. It was shown to react with ATP and activity increased in the presence of long chain β-hydroxy acids. The OleC enzyme was crystallized in the presence of 5'-AMP and synchotron diffraction data collected to 3.4 Angstrom resolution. There was significant mobility in the linker region between the N terminal domain and the C terminal domain making detailed structure elucidation impossible.

OleA was thought to catalyze the crucial first step in the biosynthetic pathway and thus provided the key to reconstituting alkene biosynthesis *in vivo*. The *Arthrobacter* OleA was not active *in vitro*. Subsequently six synthetic *oleA* genes were cloned into *E. coli* expression hosts. The *Xanthamonas campestris* OleA protein expressed well in *E. coli* and comprised approximately 50% of the soluble protein. The his-tagged protein was purified to homogeneity in one-step via Ni-column chromatography. OleA has a subunit MW of 38,800 Da and a subunit stoichiometry of 1.75. OleA was assayed with fatty acyl-CoA substrates by measuring the release of coenzyme A (CoA) using 5,5'-dithio-bis-(2-nitrobenzoic acid) (DTNB). Surprisingly, the ratio of CoA released:fatty-acyl group consumed was 1:1 whereas homologous biosynthetic thiolases show a ratio of 1:2. OleA was indicated to catalyze a Claisen

condensation of fatty acyl groups coupled with hydrolysis of the generated CoA ester to produce a β-keto acid intermediate. The β-keto acid intermediate was rigorously identified by high pressure liquid chromatography and mass spectrometry of the methyl ester.

Details of the alkene biosynthetic pathway remain to be elucidated. This thesis describes research in which two of the four enzymes have been purified to homogeneity, details of their reactions characterized, the proteins were crystallized, and details of protein structure were revealed. The OleA, OleC, and OleD proteins were combined *in vitro* and shown to produce alkenes. Moreover, the product of the OleA reaction produced synthetically was shown to be a competent intermediate for alkene synthesis by OleC and OleD. The OleB protein is not required and its role in alkene biosynthesis remains to be elucidated.

TABLE OF CONTENTS

Acknowledements……..i

Dedication……..iii

Abstract……..…iv

List of Tables……...xiii

List of Figures…….….xvi

Chapter 1: Introduction

1.1 Status of current petroleum based fuels**.....**…..1

1.2 Current alternatives……..……..1

1.3 Next generation fuel alternatives……3

1.4 Alkane and alkene biosynthesis……...…..5

1.5 Goals of this research…….8

Chapter 2: Genomic and Biochemical Studies Demonstrating the Absence of an Alkane-Producing Phenotype in *Vibrio furnissii* M1.

2.1 Introduction……...…10

2.2 Materials and Methods……...……..…12

2.2.1 Microorganisms and Cultivation……...……12

2.2.2 Chemicals……..13

2.2.3 Analytical Methods……...……14

2.2.4 Methods relating to alkane contamination reduction……15

2.2.5 *In vitro* experiments……...……...15

2.2.6 REP-PCR and pulsed-field gel electrophoresis (PFGE)……..15

2.2.7 Genome sequencing and annotation……..17

2.2.8 Functional analysis of ORF 275……....18

2.2.9 Nucleotide sequence and accession number…….....20

2.3 Results……..20

2.3.1 *V. furnissii* M1 general characteristics and comparison to other vibrios…….20

2.3.2 PFGE…….23

2.3.3 Genome annotation……...…23

2.3.4 Whole-cell studies attempting to detect alkanes……...…26

2.3.5 Cell-free enzyme assays for alkane formation……..30

2.3.6 No evidence for hydrocarbon oxidation……32

2.3.7 Other Vibrio strains…….……..33

2.4 Discussion……33

2.5 Conclusions……..37

2.6 Acknowledgements……..37

2.7 Publication of thesis work……38

Chapter 3: C_{29} Olefinic Hydrocarbons Biosynthesized by *Arthrobacter* Species

3.1 Introduction……..39

3.2 Demonstration of alkenes in cultures of *Arthrobacter* spp……..40

3.3 Rigorous assignment of structures to resolved C_{29} isomers…….44

3.4 Growth Studies…….48

3.5 Conclusions……..51

3.6 Acknowledgements……..52

3.7 Publication of thesis work……52

Chapter 4: Studies on expressing the *Arthrobacter oleA* gene

4.1 Introduction……..53

4.2 Methods……53

4.2.1 Cloning thiolase…….53

4.2.2 Expression studies and purification……...…55

4.2.3 Enzyme assays……..56

4.2.4 Recloning thiolase with C terminal his tag……...57

4.2.5 Expression with pRARE plasmid……..59

4.3 Results……..60

4.4 Discussion……69

4.5 Conclusions……..70

Chapter 5: Expressing *Stenotrophomonas maltophilia* genes, *oleABCD*

5.1 Introduction……..72

5.2 Methods……75

5.2.1 Cloning……..75

5.2.2 Expression studies…….75

5.2.3 Purification of OleA, B, C, and D……..…76

5.2.4 Improving expression of OleA……..77

5.2.5 Improving soluble expression of OleB……..78

5.2.5 Improving soluble expression of OleD……...……..79

5.3 Results and Discussion……80

5.3.1 Expression studies with spin columns……..……80

5.3.2 Purification of OleA……..84

5.3.3 OleB……..84

5.3.4 OleD……..88

5.4 Conclusions……..95

Chapter 6: Purification and crystallization of the thiolase OleA from *Xanthamonas campestris* and demonstration of a non-decarboxylative Claisen condensation

6.1 Introduction……..96

6.2 Experimental Procedures……..……100

6.2.1 Chemical synthesis and analysis……100

6.2.2 Cloning and expression of OleA……101

6.2.3 Purification of OleA……102

6.2.4 Identification and expression of active OleD……..103

6.2.5 Purification of OleC, OleD, and assay of OleD……...…...104

6.2.6 Detecting the release of CoASH thiol for assay of OleA substrate range……105

6.2.7 Hydrocarbon detection enzyme assay……105

6.2.8 Radiolabeled acyl-CoA Assay……106

6.2.9 Mass spectrometry analysis……..…107

6.2.10 Analytical gel filtration……...…..107

6.3 Results……108

6.3.1 Cloning and expression of *oleA* genes and purification of OleA protein……..108

6.3.2 General characteristics of OleA……..108

6.3.3 Initial defining of substrate specificity and reaction products of OleA……..…112

6.3.4 Identification of initial condensation product……...…..117

6.3.5 Role of OleA in olefin biosynthesis……120

6.4 Discussion……..124

6.5 Ole enzyme pathway……..129

6.6 Publication of thesis work……..129

6.7 Acknowledgements……...…….131

Chapter 7: Cloning, purification, crystallization and preliminary X-ray diffraction of the OleC protein from *Stenotrophomonas maltophilia* involved in head-to-head hydrocarbon biosynthesis

7.1 Introduction……...…….133

7.2 Experimental……...……...134

7.2.1 Cloning of the *oleC* gene……134

7.2.2 Expression and purification of OleC……..135

7.2.3 Crystallization……137

7.2.4 X-ray diffraction data collection, processing and structure solution……138

7.3 Results and discussion……...……138

7.3.1 Crystallization……..…138

7.3.2 Crystallization and X-ray diffraction data collection……..142

7.4 Acknowledgements……...…….142

7.5 Publication of thesis work……..144

Chapter 8: Conclusions and recommendation for future work

8.1 Thesis conclusions……..…145

8.2 Future work……146

8.2.1 OleA……146

8.2.2 OleD……...…….147

8.2.3 OleC……148

8.2.4 OleB……149

References……...……150

Appendix I: Thiolase (OleA) sequence from *Arthrobacter*, AAur_1998……...159

Appendix II: *Stenotrophomonas* gene sequences (as ordered through GenScript)…157

Appendix III: Detailed methods and scheme of chemical synthesis and analysis….162

Appendix IV: OleA genes (as ordered from DNA 2.0)……..168

Appendix V: OleD genes ordered through DNA 2.0……..172

Appendix VII: OleD Solubility Test……...……175

Appendix VI: OleC Preliminary Enzyme Assay Data……...….178

LIST OF TABLES

2-1. *In vitro* assay for hexadecanol reduction to hexadecane using [^{14}C]hexadecanol……31

3-1. Distribution of alkene chain lengths detected by GC-MS for a variety of *Arthrobacter* strains and comparison to their distribution in two previously studied bacteria, *S. maltophilia* and *M. luteus*……..43

6-1. Properties of OleA compared with homologous proteins in the thiolase superfamily……111

6-2. Substrate specificity of OleA as determined by CoA release……...…113

6-3. Product ratios determined by GC-MS for reactions of OleA with acyl-CoA substrates of different carbon chain lengths……..…115

6-4. Electrospray-ionization (ESI) mass spectrometry of product 1 from reaction of OleA with myristoyl-CoA and synthetic 2-myristoylmyristic acid……...…119

7-1. Data-collection statistics……..……140

VI-1. OptiSol Solubility Screen top solubility results……...……175

LIST OF FIGURES

2-1. REP-PCR of genomic DNA from *V. furnissii* M1 and control strains……..……21

2-2. Comparative PFGE of genomic DNA from *V. furnissii* M……..……...……..22

2-3. Genome region from *V. furnissii* M1 showing similarities to a gene region in *Salmonella*……...……...25

2-4. Initial gas chromatograms of *V. furnissii* M1 and a chloroform blank showing the extent of contamination……..……27

2-5. Gas chromatograms of extracts of *V. furnissii* M1 (A) and *E. coli* K-12 (B) using cleaner solvents and methods……...……..29

3-1. Gas chromatograms of nonpolar extracts from *Stenotrophomonas*, *Micrococcus*, and *Arthrobacter* strains showing the variety of alkenes produced by each strain…...…...41

3-2. Synthetic scheme for preparing iso- and anteiso-*cis*-dimethyl-13-heptacosenes (I-III, in box) and corresponding alkyne, *trans*-alkene and alkanes……...…….……45

3-3. Gas chromatograms of the C_{29} alkene region from an *Arthrobacter chlorophenolicus* A6 extract shown in comparison with synthetic standards……...….……46

3-4. Time course of culture optical density and C_{29} alkene accumulation in *Arthrobacter chlorophenolicus* A6……...……..…….49

4-1. 10% Tris-HCl SDS-PAGE of insoluble and soluble *E. coli* fractions of crude lysate containing thiolase gene from *Arthrobacter*……...……...…61

4-2. 10% Tris-HCl SDS-PAGE gel of FPLC fractions and fractions 3 and 4 with varying levels of reducing agent, DTT……..……62

4-3. Mass Spectrometry results of the processed gel slices……..……63

4-4. Mascot search results of mass spectrometry fragments of L1 protein band running between 36 and 53 kDa by SDS-PAGE…….64

4-5. Mascot search results of mass spectrometry fragments of S1 protein band running between 53 and 78 kDa by SDS-PAGE…….……65

4-6. FPLC fractions with modified elution program…….67

4-7. Purification results with BL21 Rosetta 2 cells……...68

5-1. Nickel affinity spin column results……..…….…….81

5-2. Expression study of OleA strains…….…….82

5-3. FPLC Chromatogram of BL21 (DE3) OleA purification……..83

5-4. Expression studies of OleB before and after Rosetta2 plasmid…….85

5-5. OleB expression time study at 16°C……..86

5-6. OleB purification results……87

5-7. Temperature expression studies with OleD……...89

5-8. Induction study to increase OleD soluble expression……90

5-9. Purification of OleD……...91

5-10. OleD soluble vs. insoluble fraction with chaperones……92

5-11. OleD Purification with chaperones……93

6-1. Fundamentally different condensation mechanisms have been proposed for OleA…..99

6-2. SDS-PAGE gel showing: standard molecular weight markers, and purified OleA…110

6-3. OleA reaction products with [^{14}C]-myristoyl-CoA as the substrate……116

6-4. Mass spectra for products from the reaction of diazomethane with: product 1 (40 min retention time) from the reaction of OleA with 2-myristoylmyristic acid, and synthetic 2-myristylmyristic acid……118

6-5. Gas chromatogram with mass detector showing products observed on co-incubations with OleA and OleC and OleA, OleC and OleD……..…121

6-6. Gas chromatogram and accompanying mass spectra of peaks eluting between 20.3 and 20.4 minutes…….……122

6-7. Proposed reaction cycle for OleA……127

6-8. Ole Enzyme Pathway……...130

7-1. SDS-PAGE analysis of OleC……...…136

7-2. OleC crystals grown in the presence of 5'-AMP……...…139

7-3. Section of an X-ray diffraction image from OleC crystals grown in the presence of 5'-AMP…….…141

III-1. Scheme showing chemical synthesis of 2-myristoyl myristic acid (2MMA)……..…165

VII-1. Proposed reaction mechanism of OleC in alkene biosynthesis……...……....176

VII-2. Substrate Effect on OleC Activity…….……..182

VII-3. Investigating the Nature of the Substrate…….……183

VII-4. HPLC results of standards and OleC reaction…….184

VII-5. HPLC results, revised program, standards (A-C) and OleC reaction with a hypothesized substrate…….185

CHAPTER 1

Introduction

1.1 Status of current petroleum based fuels

Scholars, government, and the public all agree that petroleum based fuels are not going to meet our long term needs. The US imports twice as much crude oil as it produces to meet our energy needs (36). The foreign dependence that the US has on crude oil is troubling in times of global unrest, but also leads to higher fluctuations as our control on pricing is limited. In recent years alone gasoline prices at the pumps have fluctuated from $2 up to almost $4 per gallon in most parts of the country. The difficulties that high prices at the pump have caused are only a sign of things to come as we reach the controversial crude oil production peak. Although the estimates of the peak year span a range up to almost 100 years depending on calculation assumptions, some sources estimate as early as 2025 (36). There is no doubt that alternatives are necessary for the current and growing energy demands.

1.2 Current alternatives

Current alternatives that are in use include ethanol and biodiesel. These biofuels are working towards the goals of reducing our dependence on foreign sources and providing carbon neutrality as they are derived largely from plant sources (34, 73)

The current renewable fuel options need to be optimized in two important areas: they are using significant croplands for the production of plants that could be used for food and the processes are not as cost effective as current crude oil sources (34).

Bioethanol is produced by yeast fermentation of predominantly corn and sugar cane (34). The higher degree of oxygenation of ethanol in comparison to petroleum based gasoline reduces the energy generated (73). Ethanol polarity and azeotropic water also cause problems with corrosion (52). Another issue that is being heavily researched is the use of food products and valuable crop land. A more ideal system could use a waste stream of biomass as a fermentation growth medium to produce ethanol to remove the choice of food versus fuel. Many researchers are working on improving the processing of cellulosic materials, enzymatically or chemically, in order to make it a more ideal substrate for ethanol fermentation (19, 98). Another challenge facing ethanol is that it costs more to produce than a comparable volume of petroleum fuel (19). The research that is being done to utilize lignocellulosic biomass as a feedstock will ultimately bring the processing and feedstock costs down, but additional engineering knowledge will need to be accrued for efficiently removing ethanol from an aqueous system (19).

Biodiesel is defined as the monoalkyl ester of long chain fatty acids and it is derived from natural triacylglycerides and phospholipids by the removal of glycerol in chemical processing (73). Currently animal fats and plant oils, such as rapeseed oil, canola oil, soybean oil, used cooking oil, and others, are used to generate biodiesel. Biodiesel can be blended with petroleum diesel or used directly with modified engines

(34). Biodiesel processing costs are still relatively high (98). The transesterification process uses significant energy, the byproduct and catalyst are difficult to recover, and the highly alkaline wastewater requires treatment (98). Improvements toward the feasibility of traditional biodiesel as a fuel alternative are being investigated. Algae sources for lipids are being explored as an alternative to edible land plant sources. Enzymatic transesterification processes are also being developed to reduce processing costs (98).

1.3 Next generation fuel alternatives

Given the ever pressing needs, significant progress has been made on new biofuel alternatives. Several avenues are being extensively explored using biotechnology of microorganisms, including short chain alcohols, isoprenoids, and fatty acid derived hydrocarbons (89). Although many of these pathways are being studied and engineered in heterotrophic bacteria grown on glucose, researchers are more recently investigating photosynthetic bacteria or algae. This has the benefit of deriving carbon for a biofuel from carbon dioxide and energy from photosynthesis.

Short-chain alcohol pathways, in addition to ethanol, have been studied for decades (25, 61), and new engineered pathways are also emerging as the search for a new biofuel becomes more pertinent. The most promising naturally-occuring fermentation pathway to n-butanol has been studied in *Clostridium acetobuylicum*, yielding 20 g/L (93). Isopropanol is also produced fermentatively in *Clostridium*

species, although only to a titer of 1.8 g/L (30). The n-butanol and isopropanol pathways have been engineered into *E. coli* in attempts to increase the yield over the native Clostrial pathways (30), but improvements are needed to obtain higher production in heterologous hosts. Another promising route to short-chain alcohols involves the metabolic engineering of bacteria to utilize 2-keto acids derived from amino acid biosynthesis to produce alcohols for use as fuel (30, 93). A decarboxylase and dehydrogenase have been engineered into the strains to convert intermediates in amino acid biosynthesis to a variety of acids, such as 1-propanol, 1-butanol, isopropanol, 2-methyl-1-butanol, 3-methyl-1-butanol, 3-methyl-1-propanol, 4-methyl-1-propanol, and 2-phenylethanol (30, 93).

Isoprenoid biosynthetic pathways have been explored intensively as a means to make C_{10} and C_{15} hydrocarbons and alcohols with excellent fuel properties. Amyris, Inc. is currently at the pilot scale with plans for commercial production in 2011 (31). Isoprenoid compounds have already been engineered into organisms effectively for the production of pharmaceuticals (22), and have more recently been applied to fuel biosynthesis by pathway modification. Isoprenoid alcohols and olefins are useful as biofuels, and produced in one step after condensing the C5 isoprenoid units, isopentenyl pyrophosphate and dimethylallyl pyrophosphate, to form geranyl pyrophosphate, farnesyl phosphate, and geranylgeranyl pyrophosphate (89).

Fatty acid derived hydrocarbons are also being explored as a potential gasoline or biodiesel alternative. The high energy density and chemical properties of long chain hydrocarbons derived from fatty-acyl CoAs could be compatible with current

infrastructure and refining processing (93). Alkane and alkene biosynthesis, although described in the literature for decades, are only recently becoming understood at the gene and enzyme level (12, 48, 49, 96, 109). Metabolic engineering in microbes is proving to be a feasible vehicle for hydrocarbon production. LS9 Inc, is scheduled for the production start-up of long chain alkanes for biodiesel in 2011 (31), and Joule Unlimited is already at pilot scale start-up of biodiesel from cyanobacteria (www.jouleunlimited.com). Gevo is currently retrofitting bioethanol plants for isobutanol production, which can be blended into gasoline for use or can be further processed to hydrocarbons that can used directly for transportation fuels (www.gevo.com).

1.4 Alkane and alkene biosynthesis

Hydrocarbon biosynthesis has been studied for decades. As early as the 1920s and 30s Chibnall and coworkers (23, 26) proposed mechanisms for the biosynthesis of long chain alkanes, ketones, and alcohols found in plants and insects. The pathway proposed in 1929 by Channon and Chibnall (23) described a head-to-head condensation of fatty acids to form hydrocarbon. They investigated a ketone intermediate by the rigorous identification of the biosynthesized ketone and this resulted in odd chain length fatty acid starting compounds (23). The lack of evidence for fatty acids with an odd number of carbons resulted in skepticism of a possible ketone intermediate (23, 90). Chibnall and Piper later published a revised mechanism

that involves the conversion of long chain fatty acids through oxidation, decarboxylation, and reduction to alkanes rather than head-to-head condensation (26). In the 1960s, Kolattukudy studied the hydrocarbon biosynthetic pathway in plants, and supported the mechanism that alkanes were biosynthesized by elongation of fatty acids to a carbon chain length up to 33 carbons followed by decarboxylation to yield an alkane (67). For years, head-to-head condensation or elongation and decarboxylation were the accepted competing mechanisms to alkanes (41, 68).

The elongation and decarboxylation mechanism was investigated in great detail by Kolattudy et al (67-69). Strong evidence for the elongation/decarboxylation mechanism in plants was presented in the work done on *Brassica oleracea* leaves, in which ^{3}H labeling studies showed a C_{29} alkane was derived from a C_{30} long chain fatty acid (69). In subsequent studies, an aldehyde was ~~found as a potential~~proposed to be a precursor to the alkane and then exogenous aldehyde was converted to alkane using the subcellular fraction of *Pisum sativum* extract (15). The activity of a partially purified aldehyde decarbonylase from *Botyrococcus braunii* was reported by the detection of 1:1 stoichiometry of heptadecane and CO from octadecanal (35), but this work has never been reproduced. A fatty acyl-CoA reductase and decarbonylase were further purified and partially characterized from *Pisum sativum* (97, 117). Other research provided evidence for an elongation/decarboxylation mechanism, such as the work by Cassgne and Lessire in *Allium porrum* L. leaves, in which labeled lignoceric acid is converted to tricosane (20).

Another alkane biosynthetic mechanism was reported for a *Vibrio* species isolated from wastewater (88). Very high levels of hydrocarbons were described, with levels of alkanes reported in *Vibrio furnisii* M1 as high as 124% of cell dry weight (88). *V. furnisii* M1 also attracted interest because it could grow on waste materials. Growth substrates were investigated in the desire to understand the sewage or waste stream that would be appropriate to use as a growth medium for this organism that produced high levels of alkanes (87). The biosynthetic mechanism was also studied with ^{14}C labeled substrates. Based on those studies, the proposed pathway followed the reduction of fatty acids to the corresponding primary alcohols with a subsequent alcohol reduction to alkane. This would make a very efficient commercial process as no carbon was lost (86). A patent was issued in Japan based on these studies.

In addition to these studies of alkane biosynthesis in plants and bacteria, bacterial alkene biosynthesis was also studied in the 1960s by Albro, Huston, and Dittmer (2-6) and by Tornabene and Oro (111, 112). Hydrocarbons produced by *Sarcinea lutea* (now reclassified as *Kocuria rhizophilia*) were identified in detail (3, 6) and the competing biosynthetic mechanisms were revisited (2-5, 23, 67). The pathway was initially investigated by the addition of diagnostic substrates. For example, branched-chain fatty acid precursors such as ^{14}C isoleucine and ^{14}C valine were used to determine that the terminal ends of the fatty acids were incorporated into the hydrocarbons (5, 113). Terminally labeled ^{14}C palmitic acid was found to incorporate at a high rate into hydrocarbons while radioactivity from 1-^{14}C palmitic acid starting compound was incorporated at a very low rate, further support for the condensation

mechanism (113). Ketones were also investigated as a substrate because labeled precursors to alkenes were also found in ketones and secondary alcohols. However, it was found that ketones would not give rise to alkenes, therefore ketones were not considered to be pathway intermediates (2).

Albro and Dittmer attempted to elucidate the pathway via *in vitro* studies. They developed a cell free assay using a dialyzed cell lysate that was centrifuged to produce a crude soluble enzyme fraction. Labelled (^{14}C) fatty acid substrate and cofactors were added to the extract to determine which cofactors increased alkene production, measured by radioactivity. The cofactors that were found to increase activity included ATP, Coenzyme A, Mg^{2+}, NADPH, and pyridoxal or pyridoxamine phosphate (4). In the next report on cell free studies, Albro and Dittmer report that palmitaldehydes are incorporated into hydrocarbons, but no mechanistic follow-up is done and it seems inconsistent with the previous studies (7). Although there was early work towards understanding the mechanism of alkene biosynthesis in microbes, no pathway or enzymes were definitively identified.

1.5 Goals of this research

The ultimate goals of this research have been to understand a hydrocarbon biosynthesis pathway that could be scaled up to produce hydrocarbons as a potential biofuel. Linear hydrocarbons (straight chain and branched) are the major constituent of gasoline from petroleum and can be fed into current infrastructure to heavily

supplement our current gasoline supplies. Reports already existed in the literature that described the biosynthesis of hydrocarbons by various organisms, but the pathways and enzymes were not understood. It was decided here to work with bacteria that can be studied in detail more readily than insects and plants. Bacterial genomes can be sequenced and mined with greater ease than the much larger genomes of plants and insects. Many bacterial genomes are already available, allowing for a comparative bioinformatics approach. Genetic engineering of bacteria is also more developed than comparable manipulations with plants and insects.

The alkane and alkene pathways described in the literature were investigated and the genes identified (49) in alkene biosynthesis. Once the genes were identified, I was able to study the enzymes *in vitro* and develop hypothetical mechanisms based on sequence homology with enzymes they are related to in the large superfamilies. The mechanisms uncovered, however, revealed differences between the new enzymes discovered here and their superfamily homologs. The ultimate goal of this research was to elucidate the pathway of hydrocarbon biosynthesis in microbes. Challenges existed in obtaining pure and active proteins in the pathway and identifying previously undescribed, and sometimes unstable, metabolites.

CHAPTER 2

Genomic and Biochemical Studies Demonstrating the Absence of an Alkane-Producing Phenotype in *Vibrio furnissii* M1.

2.1 Introduction

The need for renewable energy sources will require the development of biofuel options other than ethanol. One excellent fuel option would be bio-alkanes. Alkanes comprise the major component of current petroleum-based fuels. A biological petroleum would be renewable and completely compatible with existing fuel infrastructure. Thus, considerable interest was generated by recent reports of high-level *n*-alkane formation by the bacterium *Vibrio furnissii* M1 (86-88).

V. furnissii strains were recognized as a distinct species in 1983 (16). Other *Vibrio* species, such as *V. cholerae* (29) and *V. parahaemolyticus* (123), have been more extensively studied because of their significant pathogenicity in humans. Both of

the latter species (54), along with *V. vulnificus* (24) and *V. fischeri* (92), have been subjected to genomic sequencing that has been completed and published. *V. furnissii* has been most extensively studied with respect to its physiological and genetic mechanisms of chitin degradation (11, 65). Marine vibrios are prominent chitinolytic organisms (72).

Thus, the recent report of a *V. furnissii* strain biosynthesizing appreciable quantities of *n*-alkanes was unusual and interesting (86-88). The organism was isolated from activated sludge of a sewage disposal plant located in the Osaka prefecture of Japan (88). It was reported to produce a copious lipid layer that floated on top of liquid cultures. The culture lipids were found to consist of 48% alkanes (88), and the alkane-producing phenotype was the subject of three papers published between 2001 and 2005. The reports were significant for several reasons. First, *n*-alkanes were produced during growth on renewable carbon sources such as sugars and polysaccharides. The latter included starch, chitin, and xylan. This is of specific interest when considering substrates for large-scale growth for production of a usable biofuel. Second, the amount of alkanes was significant, accounting for as much as 35% of the carbon consumed (87). Third, the *n*-alkane backbone was proposed to derive largely from the reduction of fatty acids through a novel mechanism, the reduction of a fatty alcohol to an alkane (86). This mechanism would be both interesting and commercially appealing, as there would be no loss of carbon in a pathway condensing acetyl coenzyme A (acetyl-CoA) units into a long-chain acyl-CoA followed by six-electron reduction to yield an alkane.

The present study was conducted to investigate the alkane-producing phenotype of *V. furnissii* M1 using a combined approach of whole-genome sequencing and biochemical studies. The major findings were that alkane-producing genes could not be identified and alkane biosynthesis could not be demonstrated *in vivo* or *in vitro*.

2.2 Materials and Methods

2.2.1 Microorganisms and cultivation

V. furnissii M1 was kindly provided by Kazuya Watanabe of the Marine Biotechnology Institute, Japan. Other strains were obtained from the American Type Culture Collection (ATCC, Manassas, VA).

V. furnissii M1 was cultivated in three different media. Marine liquid medium contained 37.4 g marine broth 2216 (Becton Dickinson and Company, Franklin Lakes, NJ) per liter distilled water. Luria-Bertani broth was adjusted to 4% (wt/vol) NaCl. Medium 3 contained 2.0 mg EDTA·2Na, 2.8 mg H_3BO_3, 0.75 mg $Na_2MoO_4{\cdot}2H_2O$, 0.24 mg $ZnSO_4{\cdot}7H_2O$, 1.6 mg $MnSO_4{\cdot}H_2O$, 0.04 mg $Cu(NO_3)_2{\cdot}3H_2O$, 0.75 mg $CaCl_2{\cdot}2H_2O$, 0.2 mg $MgSO_4{\cdot}7H_2O$, 10 mg $FeSO_4{\cdot}7H_2O$, 30 g NaCl, 1.32 g $(NH_4)_2SO_4$, 0.2 g yeast extract, 8.7 g $K_2HPO_4{\cdot}3H_2O$, and 8.4 g KH_2PO_4 per liter distilled water. Carbon sources used were 1.64 g sodium acetate, 1.92 g sodium propionate, 2.70 g disodium succinate hexahydrate and 1.87 g disodium L-malate per liter. The pH was adjusted to 7.0 with KOH. After autoclaving, the following filter-sterilized reagents (Sigma-Aldrich, St. Louis, MO) were added: 0.025 mg thiamine-HCl, 0.025 mg D-biotin, 0.025

mg nicotinamide, and 0.025 mg *p*-amino benzoic acid per liter. Plates were prepared by the addition of 1.5% (wt/vol) granulated agar (Becton Dickinson and Company) to broth media.

All strains were maintained as frozen stocks and grown at 37°C with agitation at 210 rpm, unless otherwise noted. Strains were initially transferred from frozen stocks onto marine agar plates. After 18 to 24 h, isolated colonies were used to inoculate 3 ml of medium 3. After approximately 18 h, measurements of optical density at 600 nm (OD_{600}) (Beckman DU 7400) ranged from 1.3 to 2.0. Two hundred microliters of each preculture was used to inoculate 20 ml of marine broth or medium 3 in 25-ml Erlenmeyer flasks. Cultures were grown under numerous conditions to test for alkane formation. Among the factors tested were degree of aeration, alternate carbon sources, different medium types, and harvesting media and cultures at various growth stages over 1 to 7 days. The conditions used in different experiments are provided with the relevant results below.

2.2.2 Chemicals

Chromasolv chloroform (Sigma-Aldrich, St. Louis, MO), spectrophotometric-grade methanol (Sigma-Aldrich), hexanes (Mallinckrodt, Hazelwood, MO), diethyl ether (Fisher Scientific, Hampton, NH), heptane (Sigma-Aldrich), octacosane (Acros Organics, Geel, Belgium), and hexadecane (Sigma-Aldrich) were obtained from the

sources indicated. [1-^{14}C]hexadecanol (Moravek Biochemicals Inc., Brea, CA) was 56 mCi/mmol and had a radiochemical purity of 99.3%.

2.2.3 Analytical Methods

Analytical methods largely followed those described by Park et al. (87, 88). Cultures were extracted following the Bligh and Dyer protocol (13). The concentrated extract was developed with 80:20:1 hexanes-diethyl ether-water on silica gel thin-layer chromatography plates. Spots were eluted with chloroform, concentrated to 100 µl, and analyzed by gas chromatography-mass spectrometry (GC-MS). To ensure that alkanes were not lost during initial chromatography, extracts were also directly analyzed by GC-MS. GC-MS analysis was conducted with an HP6890 gas chromatograph connected to an HP5973 mass spectrometer (Hewlett Packard, Palo Alto, CA). GC was conducted under the following conditions: helium gas, 1 ml/min; HP-5 column (5% phenylmethyl siloxane capillary; 30 m by 250 µm by 0.25 µm); temperature ramp, 100 to 300°C; 10°C/min. The mass spectrometer was run under the following conditions: electron impact at 70 eV and 35 µA.

In experiments where analytical standards were spiked into growth medium, 125 nmol hexadecane was added at the end of the growth phase to a 20-ml culture grown on medium 3 for 11 days. Octacosane (0.25 µmol) was used to spike a 50-ml culture grown on medium 3 containing 10 mM D-glucose for 24 h. The cultures were then extracted and handled as described previously.

2.2.4 Methods relating to alkane contamination reduction

The following methods were instituted to reduce alkane contamination during the course of the present study. The chloroform solvent used for extractions was switched from Chromasolv to Chromasolv Plus (both from Sigma-Aldrich, St. Louis, MO). All glassware was rinsed twice with Chromasolv Plus solvent prior to use. Stoppers were neoprene rubber, and Teflon stopcocks were used in separatory funnels. Contact of any solvents or other components with plasticware was avoided. In total, these methods greatly reduced the introduction of contaminating alkanes.

2.2.5 In vitro experiments

Cell-free enzyme fractions were obtained and assays were conducted as described previously by Park (86). Minor modifications were as follows: dispersal was accomplished by shaking rather than sonication to prevent generating radioactive aerosols; 10-mg protein aliquots were used in the assay for greater sensitivity.

2.2.6 REP-PCR and pulsed-field gel electrophoresis (PFGE)

Genomic relatedness of *Vibrio* strains was investigated utilizing repetitive extragenic palindromic PCR (REP-PCR) DNA fingerprinting with the primers ERIC 1R (3'-CACTTAGGGGTCCTCGAATGTA-5') and ERIC 2 (5'-AAGTAAGTGA

CTGGGGTCAGCG-3') (39). PCR amplification was performed using the following protocol: 95°C for 2 min; 30 cycles of 94°C for 3 min, 92°C for 30 s, 50°C for 1 min, and 65°C for 8 min; final extension at 65°C for 8 min. Samples were separated on a 1.5% SeaKem LE agarose (Cambrex Bioscience, Rockland, ME) gel in 1x Tris-acetate-EDTA (TAE) at 4°C for 16 h at 68 mV and stained for 20 min with a solution containing 0.5 µg of ethidium bromide per ml. Gel images were analyzed by BioNumerics v.2.5 software (Applied-Maths, Sint-Martens-Latem, Belgium) and normalized to an external 1-kb reference ladder. DNA fragments less than 300 bp long were not used in analyses. DNA fingerprint similarities were calculated using Pearson's product-moment correlation coefficient with 1% optimization. Dendrograms were generated using the unweighted pair-group method using arithmetic averages (39, 57, 60).

DNA for PFGE was prepared from cells lysed in plugs (44). Cells were grown on marine agar plates overnight at 37°C, washed once in 1 ml resuspension buffer (100 mM Tris [pH 8.0], 100 µM EDTA), and then resuspended to an absorbance at 600 nm of 2.1. One part cell suspension was mixed with 20 µl of 20-mg/ml protein kinase A (VWR International) and 1 part 2% SeaPlaque GTG agarose (Cambrex BioScience, Rockland, ME) in 1x TAE and molded into plugs. Plugs were lysed at 55°C in 10 ml lysing solution (50 mM Tris [pH 8.0], 50 µM EDTA, 1% sodium dodecyl sulfate, 1% *N*-laurylsarcosine, 0.1 mg/ml protein kinase A) for 2 h. Plugs were washed four times in Tris-EDTA and twice in 0.5x TAE. Plugs were immediately used in PFGE.

PFGE was performed with a CHEF DRII system (Bio-Rad, Richmond, CA). Agarose-imbedded DNA was run on 0.8% Megabase agarose (Bio-Rad) in 0.5x TAE at 14°C. Conditions were 72 h, initial pulse time of 1,200 s, final pulse time of 1,800 s, at 2 mV/cm (44). The gel was stained with a solution containing 0.5 µg of ethidium bromide per ml. Chromosome sizes were estimated based on the mobility of unknowns in comparison with DNA samples of known base pair composition (Bio-Rad).

2.2.7 Genome sequencing and annotation

DNA for sequencing was collected as previously described (94). *V. furnissii* M1 was grown overnight in marine broth. Cells were centrifuged at 10,000 rpm for 10 min at 4°C, washed once in TEN buffer (50 mM Tris, 20 µM disodium EDTA, and 50 mM NaCl, pH 8.0), and resuspended in 8 ml TEN buffer. Cells were lysed by adding 1 ml of a 5-mg/ml lysozyme solution (Promega, St. Louis, MO) at 37°C for 30 min followed by 1 ml of 5 mg/ml predigested pronase solution (Promega) at 37°C for 30 min, and then 1 ml 20% *N*-laurylsarcosine (Promega) at 37°C for 1 h. Eleven grams of cesium chloride was added, followed by 1 ml of 10-mg/ml ethidium bromide. The solution was centrifuged at 40,000 rpm at 20°C for 40 h. DNA bands were isolated using a syringe, ethidium bromide was extracted using salt-saturated butanol, and the DNA was dialyzed overnight against four washes of 0.01 M Tris-HCl, pH 8.0, containing 0.1 mM EDTA. The DNA was confirmed to be from *V. furnissii* by 16S rRNA sequencing. 16S rRNA was amplified by PCR using the primers 27f (5'AGAGTTTGATCMTGGCT CAG-3') and 785r (5'-GGACTACCIGGGTATCTAATCC-3'), 530f (5'-GTGCCAGCM

GCCGCGG-3') and 1100r (5'-GGGTTGCGCTCGTTG-3'), and 926f (5'-AAACTYAA AKGAATTGACGG-3') and 1492r (5'-TACGGYTACCTTGTTACGACTT-3'), using the following conditions: 95°C for 5 min; 35 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 90 s; 72°C for 5 min. PCRs were performed in duplicate and sequenced at the Biomedical Genomics Center at the University of Minnesota. The DNA fragments used for genomic sequencing were determined to be longer than 15 kb via electrophoresis on a 1% agarose gel.

Genome sequencing and contig assembly were performed by the Center for Genomic Sciences at the Allegheny-Singer Research Institute, using a 454 sequencer (454 Life Sciences, Branford, CT). The Newbler assembly program was used to order the sequences into 121 contigs. In order to annotate the contigs, a single pseudochromosome was constructed using a linker sequence which allows identification of partial genes at contig margins, as described by Tettelin et al. (110).

The resulting pseudochromosome was subjected to automated annotation via GenDB (76). Pfam (101) and hidden Markov models (HMM) (40) for local and global alignments were used to search the M1 genome for specific protein targets. Additional functionality was screened by searching the M1 genome using BLASTP (9).

2.2.8 Functional analysis of ORF 275

V. furnissii M1 genomic DNA and the primers GGATTATGGCATAT GATGTTAGAT and TCTTTTCGAAACTTAACGCA were used to amplify open

reading frame (ORF) 275 using the PCR. Primers contained the NdeI and HindIII restriction sites, respectively. The gene was cloned separately into either pET28b+ or pET30a+ vectors (Novagen, San Diego, CA) and transformed into *Escherichia coli* BL-21 cells. Starter cultures of the recombinant *E. coli* strain, grown at room temperature to an OD_{600} of 0.5 to 0.6, were used to inoculate 100-ml cultures. These cultures were grown at 15°C to an OD_{600} of 0.4 to 0.5 and induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside for 19.5 h before harvesting. Crude extracts were prepared from both pET vectors by sonication on a Biosonik sonicator (Bronwill Scientific, Rochester, NY) at 80% intensity. Purified His-tagged protein was prepared from the pET28b+ construct. Crude extracts were passed over a nickel column (Novagen) and eluted with a solution of 1 M imidazole, 20 mM Tris [pH 7.9], 50 mM NaCl.

Acetaldehyde dehydrogenase (CoA-acetylating) activity was measured by monitoring NADH production or consumption at 340 nm on a Beckman DU 640 spectrophotometer (Beckman Coulter, Fullerton, CA). All solutions were prepared in 20 mM Tris, pH 7.5. Conversion of acetaldehyde and CoA to acetyl-CoA was measured with both crude extracts and purified protein. The final assay buffer contained 2.5 mM NAD, 50 mM acetaldehyde, and 1 mM coenzyme A. The reverse reaction was monitored with purified protein in an assay buffer containing 100 µM acetyl-CoA and 250 µM $NADH_2$.

2.2.9 Nucleotide sequence accession number.

The *V. furnissii* M1 16S rRNA sequence obtained in this study has been deposited in GenBank under accession number EU204961.

2.3 Results

2.3.1 V. furnissii M1 general characteristics and comparison to other vibrios

The *V. furnissii* M1 culture showed a thick, floating, waxy layer as described by Park et al. (88). The previous study also reported that the 16S rRNA sequence was consistent with identification as a *V. furnissii* strain, but the sequence itself was not reported. The *V. furnissii* M1 16S rRNA sequence obtained in this study also showed the highest similarity, >99%, to two *V. furnissii* 16S rRNA sequences in GenBank.

Subsequently, a more powerful method for comparing *V. furnissii* M1 to other *Vibrio* strains was used: REP-PCR. In this technique, PCR amplification of closely spaced repetitive elements throughout the genome can be used as a genomic signature. REP-PCR was performed with primers designed to amplify repetitive elements found in diverse microbial genomes. A representative gel is shown in Fig. 2-1. A dendrogram was constructed as described in Materials and Methods. *V. furnissii* M1 was not identical to other *V. furnissii* strains tested here. However, the relatedness observed, on the order of 60 to 90%, is consistent with comparisons in the literature between

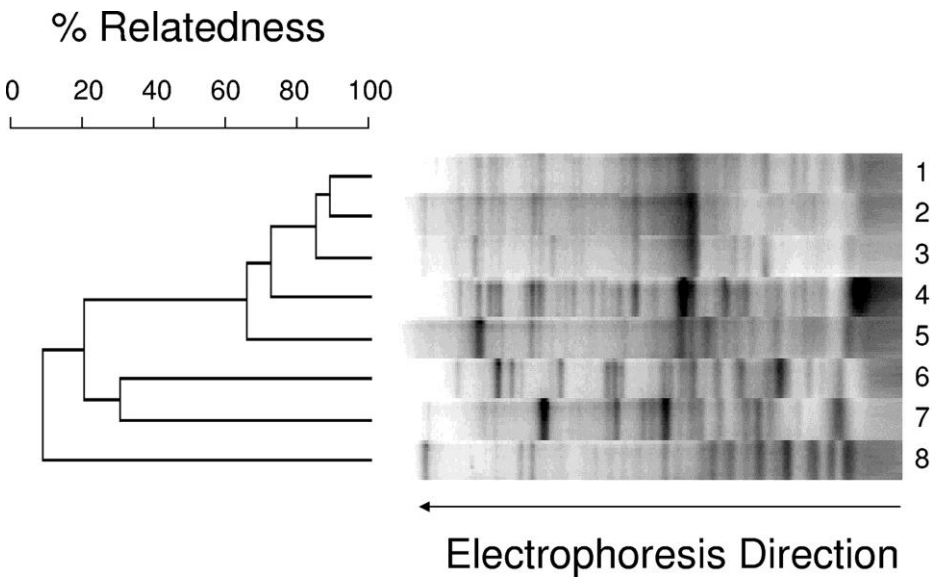


Figure 2-1. REP-PCR of genomic DNA from *V. furnissii* M1 and control strains. Percent relatedness was determined as described in Materials and Methods. The lanes contained DNA from the following strains: 1, *V. furnissii* ATCC 33841; 2, *V. furnissii* ATCC 35627; 3, *V. furnissii* ATCC 35016; 4, *V. furnissii* M1; 5, *V. furnissii* ATCC 35628; 6, *V. parahaemolyticus* LM5312; 7, *V. harveyi* B392; 8, *E. coli* K-12.

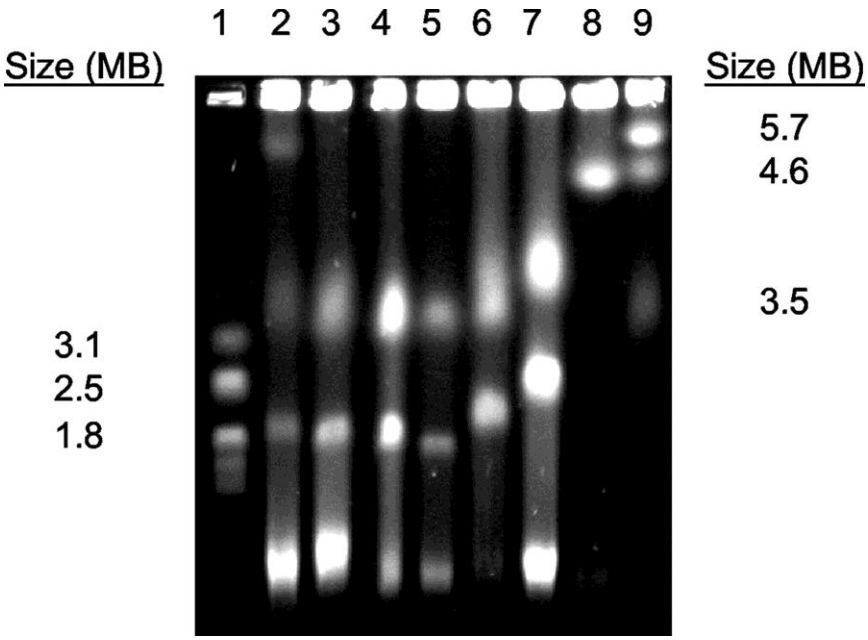


Figure 2-2. Comparative PFGE of genomic DNA from *V. furnissii* M1 (lane 2), *V. furnissii* 35016 (lane 3), *V. furnissii* 35627 (lane 4), *V. furnissii* 35628 (lane 5), *V. parahaemolyticus* LM5312 (lane 6), *V. harveyi* B392 (lane 7), and *E. coli* DH5α (lane 8). *Hansenula wingei* (lane 1) and *Schizosaccharomyces pombe* (lane 9) chromosomes were used as size markers.

organisms of the same species, confirming that *V. furnissii* M1 is indeed a *V. furnissii* strain.

2.3.2 PFGE

Genomic DNA of *V. furnissii* M1, other *V. furnissii* strains, and standard strains were analyzed by PFGE (Fig. 2-2) to determine the size of the genome and its organization. The data suggested that *V. furnissii* M1 has a bipartite genome consisting of two chromosomal elements with a total size of approximately 5 Mb. This is similar to characteristics of other *Vibrio* species, as reported in the literature and shown in Fig. 2-2 (24, 82).

2.3.3 Genome annotation

The genome of *V. furnissii* M1 was sequenced, computationally assembled and the genes annotated. The raw sequencing data consisted of 106 Mb, which represented an estimated 21-fold coverage of the total genome. The base reads were assembled into 121 contigs that converged at 4.95 Mb of genome data, consistent with the genome size estimated by PFGE. The contigs were randomly ordered and assembled into a pseudochromosome for annotation purposes.

A major goal of the genome annotation was to identify putative genes that might be involved in alkane biosynthesis. Among the candidate alkane-biosynthetic genes examined specifically were those encoding acyl-CoA reductase, aldehyde

reductase, and alcohol reductase and genes that might encode an enzyme catalyzing the reduction of an alcohol to an alkane. HMMs for the local and global Pfam alignments for acyl-CoA reductase (PF05893) were used to search the M1 genome. No sequence was found with an expectation value (e-value) less than 1.0. Dozens of putative aldehyde and alcohol reductases were identified. The genome was examined for aldehyde and alcohol dehydrogenase genes that were clustered with genes encoding enzymes resembling ribonucleotide reductase or other radical-dependent oxidoreductases, an anticipated gene constellation that might encode a carboxylic acid-to-alkane biosynthetic pathway, as proposed by Park (86).

In this context, one interesting gene cluster was annotated as consisting of genes for an aldehyde dehydrogenase (ORF 275), an iron-dependent alcohol dehydrogenase (ORF 277), a pyruvate formate lyase homolog (ORF 278), and an accompanying pyruvate formate lyase activator (ORF 279) (Fig. 2-3). These functional genes were flanked by genes that were annotated as metabolosome (carboxysome) shell proteins (Fig. 2-3). The flanking gene structure is consistent with the function of *pduJ* and *ccmO* genes in generating a carboxysome or metabolosome structure (14, 63), which is a bacterial intracellular protein shell involved in compartmentalizing metabolites or a set of reactions.

Further bioinformatics analysis of other genomes revealed a very similar cluster of genes in *E. coli* strain F11 (accession number NZ_AAJU00000000), a pathogen not

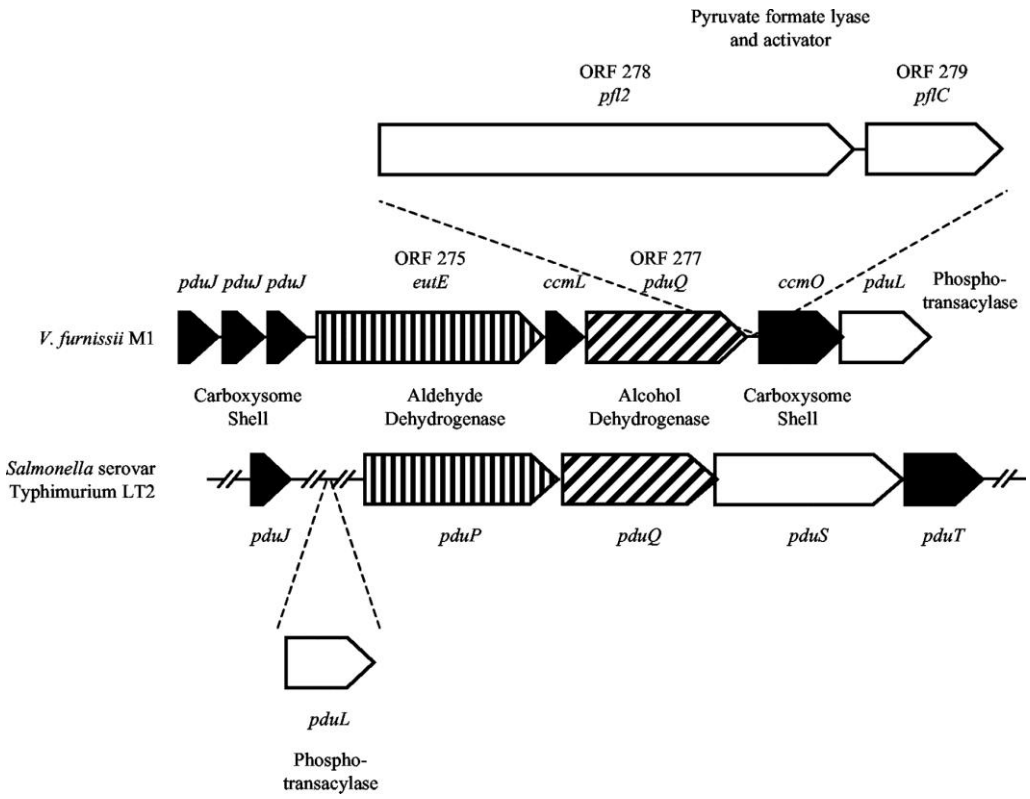


Figure 2-3. Genome region from *V. furnissii* M1 showing similarities to a gene region in *Salmonella*. Identified protein types are highlighted: carboxysome shell proteins (black), aldehyde dehydrogenases (vertical lines), and alcohol dehydrogenases (diagonal lines). The ORF numbers (275, 277, 278, and 279) are given for proteins discussed in the text.

expected to produce alkanes. Moreover, the metabolosome-like structure may replace the multidomain protein particle catalyzing the conversion of pyruvate to ethanol in *E. coli* K-12 (64). This was tested by cloning and expressing ORF 275 in *E. coli*. The recombinant *E. coli* strain was assayed for different activities as described in Materials and Methods. ORF 275 was shown with purified enzyme to encode a bidirectional acetaldehyde dehydrogenase (CoA-acetylating) enzyme. Formation of acetaldehyde occurred at a rate of 1.5 µmol/min/mg, while the reverse reaction had a rate of 2.7 µmol/min/mg. It was thus functional as an acetyl-CoA reductase, an activity complementary with a pyruvate formate lyase and an aldehyde-oxidizing alcohol dehydrogenase. These activities in total are consistent with a metabolosome that compartmentalizes a fermentative pathway metabolizing pyruvate to ethanol.

2.3.4 Whole-cell studies attempting to detect alkanes

V. furnissii M1 cells and media were extracted by methods described by Park et al. (87, 88). In preliminary experiments, alkanes were detected by GC and confirmed by MS (Fig. 2-4A). Subsequent analysis of media, solvents, and glassware revealed that they were contaminated with alkanes and other materials. The solvents used in extractions were the most significant source of contamination (Fig. 2-4B). A series of methodological alterations were made to eliminate contamination, as described in Materials and Methods, which led to greatly diminished peaks via GC.

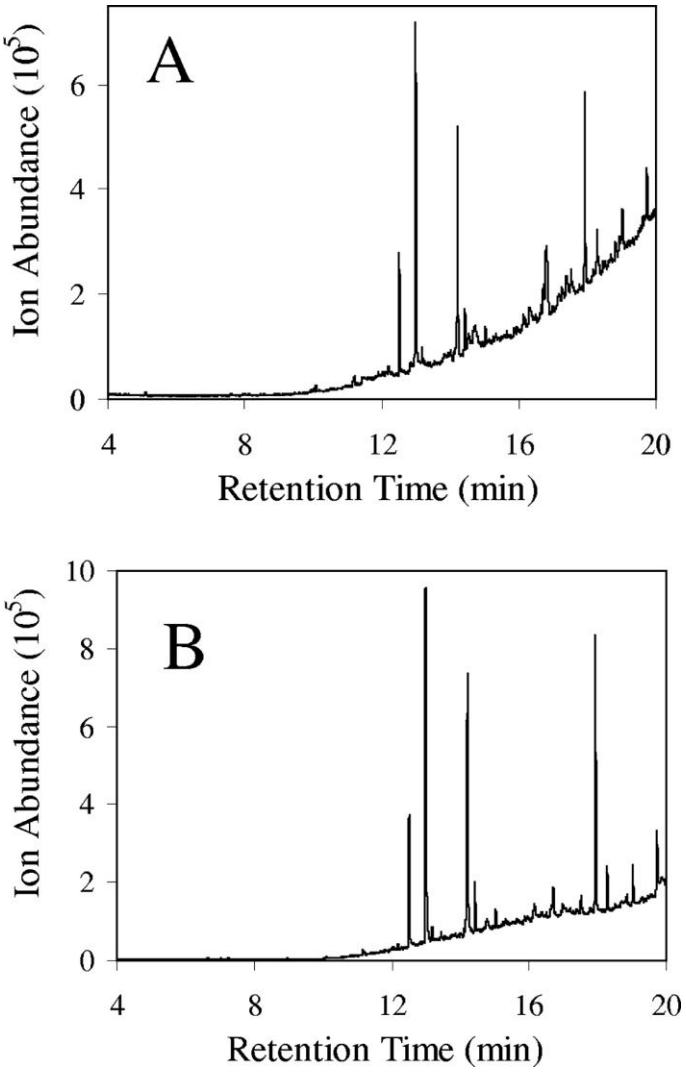


Figure 2-4. Initial gas chromatograms of *V. furnissii* M1 (A) and a chloroform blank showing the extent of contamination (B). The prominent peaks in both were identified by MS as methyl palmitate (12.5 min), dibutyl phthalate (13.0), octadecenoic acid methyl ester (14.2 min), diethylhexyl phthalate (17.9 min), and octacosane (19.7 min).

To determine whether the extraction and new workup procedures were appropriate for isolating and concentrating alkanes for detection, an internal standard was added to cultures of *V. furnissii* M1. The internal standards hexadecane and octacosane were used in independent experiments. The choice of using a C_{16} and C_{28} alkane, respectively, was made to largely bracket the entire range of *n*-alkane chain lengths previously reported to be produced by *V. furnissii* M1 (87, 88). Solvent extracts from these spiked cultures were processed and subjected to GC-MS using procedures that minimized alkane contamination. A parallel experiment was conducted using an *E. coli* culture containing the same internal standard alkanes. Figure 2-5 shows a representative chromatogram of the resultant extracts analyzed by GC-MS with octacosane-spiked medium. The large octacosane peak was clearly identifiable by both retention time and the characteristic mass spectrum. The level of octacosane added (0.25 µmol) matched the level of individual alkanes reported to be present in GC analyses by Park et al. (88). Clearly, no peaks comparable to the added standard were discernible. Very minor peaks were observed above the baseline, but the same minor peaks were found in *V. furnissii* M1 (Fig. 2-5A) and *E. coli* (Fig. 2-5B) extracts, suggesting that they are derived from a common source and are not made biosynthetically. *E. coli* and *V. furnissii* M1 cultures were grown in the same growth medium and were extracted in parallel. Results with hexadecane-spiked cultures produced similar results.

Since alkane formation could be dependent on growth conditions, analysis for alkanes was conducted with *V. furnissii* M1 cultures grown in different media, under

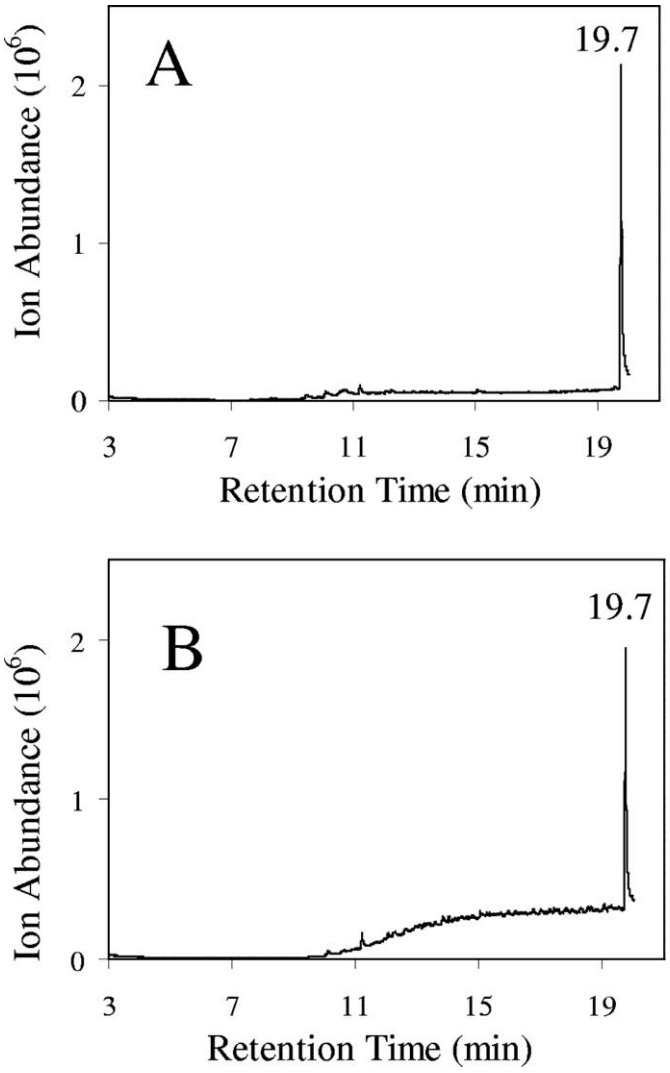


Figure 2-5. Gas chromatograms of extracts of *V. furnissii* M1 (A) and *E. coli* K-12 (B) using cleaner solvents and methods. Cultures were spiked with octacosane prior to extraction and workup. The 19.7-min peak was confirmed by MS to be octacosane.

aerobic and microaerophilic (nonshaking) conditions, and over a period of 1 to 7 days. Alkanes were not detected above background levels under any of these conditions. The media were quite different. Medium 3 is completely defined, and marine broth is a standard commercial culture medium for *V. furnissii*. Under microaerophilic conditions, the culture was observed to have a floating pellicle, and birefringence was seen in marine broth cultures on the top of the culture. While these observations were initially thought to be potential indicators of hydrocarbon formation, no alkanes attributable to *V. furnissii* M1 were obtained from extracts of these cultures.

2.3.5 Cell-free enzyme assays for alkane formation

Cell-free enzyme preparations from *V. furnissii* M1, prepared as described by Park (86), were tested for hexadecanol reductase activity (Table 2-1). No significant radioactivity was detected in the spot on a thin-layer chromatography plate corresponding to the R_f value of authentic hexadecane. The percentage of the starting radioactivity in the hexadecane fraction in all cases was less than 0.1%, a level significantly below that of the radiochemical impurities of the starting material, or 0.9%. Total recovery of radioactivity in substrate (hexadecanol) and putative product (hexadecane) fractions was 57% in the no-enzyme control and 25 to 47% in the enzyme treatments. This is similar to the total recovery reported by Park (86). A level of activity several percent of that reported by Park (86) would have been detected in this

Table 2-1. *In vitro* assay for hexadecanol reduction to hexadecane using [^{14}C]hexadecanol

Enzyme fraction	Radioactivity (dpm)	
	Hexadecanol	Hexadecane
No enzyme	1,267,002 ± 65,067	563 ± 38
V. furnissii M1, soluble	674,895 ± 23,202	303 ± 32
V. furnissii M1 membranes	799,030 ± 143,600	812 ± 162
E. coli, soluble	797,954 ± 246,564	642 ± 185
E. coli membranes	795,810 ± 127,375	474 ± 152

experiment. In separate experiments, *V. furnissii* ATCC 35028 was tested for reductase activity with hexadecanol, but no activity was detected.

2.3.6 No evidence for hydrocarbon oxidation

It was considered that cells producing alkanes may also have the capability to oxidize hydrocarbons, thus recapturing carbon and energy. Experiments were conducted to determine the potential growth of *V. furnissii* M1 in the presence of alkanes as the sole carbon source or in admixture with limiting alternative carbon sources such as glucose. No evidence for growth was observed using dodecane (C_{12}), hexadecane (C_{16}), octadecane (C_{18}), eicosane (C_{20}), docosane (C_{22}), or tetracosane (C_{24}), alkanes that Park reported to be produced by *V. furnissii* M1. In addition, bioinformatics tools were used to search for genes encoding proteins homologous to AlkA, AlkB, and cytochrome P450 monooxygenases. These genes, established to encode alkane-oxidizing enzymes in other bacteria, could not be discerned in the genome of *V. furnissii* M1. Using the HMMER 2 tool (40) in conjunction with the Pfam HMMs for cytochrome P450 (accession number PF00067), in local and global alignments against the *V. furnissii* M1 genome sequences, no match with an e-value lower than 4.0 was found. For AlkB, no HMMs were available, so the *Pseudomonas oleovorans* AlkB sequence (gi 113639) was used with the BLAST tool. The best match found in the *V. furnissii* M1 genome was 0.02. While this could indicate weak homology, there was no characteristic clustering of genes, as is found in alkane

degraders. Specifically, we could not find evidence for the presence of the gene cluster *alkFGHJKL* or the regulatory elements *alkST*.

2.3.7 Other Vibrio strains

A patent filed on *V. furnissii* M1 in Japan claimed that other *Vibrio* strains also produce alkanes, albeit in smaller amounts than *V. furnissii* M1 (77). In the present study, *V. furnissii* ATCC 35628 was tested for alkane formation *in vivo*, and no levels above background were detected. Additionally, membrane and soluble enzyme fractions were prepared from *V. furnissii* ATCC 35628 and tested *in vitro* with [^{14}C]hexadecanol. The radioactivity (in dpm) in the region of a hexadecane standard was on the order of 0.1% of the initial radioactivity, a level consistent with background radiation in negative controls. In other experiments, a strain reported in the patent to make alkanes, *V. furnissii* ATCC 35016, was obtained from the ATCC and tested. No alkanes were detected.

2.4 Discussion

The *V. furnissii* M1 strain used in this study strongly resembles the strain described previously (86-88), except that no alkane formation was observed here. It is not possible to use DNA sequence data to rigorously ascertain the relationship to the previously described *V. furnissii* M1 strain, because no DNA sequences had been

reported in the literature or deposited in GenBank. However, in this study, 16S rRNA sequence data and REP-PCR data support the idea that the organism used here was a *V. furnissii* strain and that it differed from *V. furnissii* ATCC cultures that were tested.

In this study, *V. furnissii* M1 did not make alkanes under any *in vivo* growth condition tested, and protein extracts did not catalyze alkane formation *in vitro*. Some *in vivo* studies used standards carried throughout the extraction and purification protocols to show that the methods employed would have detected alkanes, with significant sensitivity, had they been present. The conditions of growth and analysis used here followed the procedures of Park et al. (86-88) closely. Both cells and media were extracted to ensure that any alkane present would not be missed. *In vitro* assays were also very sensitive. Picomole levels of alkane would have been detectable, but nothing above background could be discerned. Levels that were orders of magnitude lower than those reported by Park (86) could have been detected in the assays conducted here.

The lack of alkane biosynthetic activity in strain M1 is consistent with the lack of activity in *V. furnissii* strains ATCC 35627, 35628, and 33841 (H. R. Beller, personal communication), which were assayed under a range of conditions comparable to those described by Park and coworkers; these *in vivo* assays involving GC-MS entailed high extraction efficiency (typically >99% based on recoveries of the surrogate compound decane-d_{22}) and would have been able to detect 0.001% of the alkane concentrations that were reported by Park and coworkers (H. R. Beller, personal communication).

In the present study, *V. furnissii* ATCC 35016 was shown not to produce alkanes under the conditions tested. A patent filed in Japan by Miyamato (77) reported that *V. furnissii* ATCC 35016 produced alkanes, albeit at lower levels than *V. furnissii* M1. This could not be reproduced in the present study.

Several observations reported by Park et al. were unexpected and unexplained. Different papers reported different hydrocarbons being produced that would derive from divergent mechanisms: even-chain alkanes, odd-chain alkenes, branched-chain alkanes, and alkenes. Differences were reported with different growth substrates, but some differences were surprising. For example, growth with acetate resulted in only a C_{18} alkene being formed, but butyl acetate, which would almost surely be metabolized via ester hydrolysis to yield acetate, gave rise to C_{18}, C_{21}, C_{24}, and C_{27} branched-chain alkanes. Butyric acid, another likely metabolite from butyl acetate, was reported to give rise to linear C_{16} to C_{18} alkanes.

No obvious genes that might be related to alkane biosynthesis were identified in this study. It must be acknowledged that alkane biosynthesis is currently poorly understood, and hence the genes may not be obvious. However, nothing resembling putative plant decarbonylases was detected. Fatty acid aldehydes and alcohols derive from acyl-CoA reductases. Only one acyl-CoA reductase homolog was identified that clustered with other genes that might be involved in alkane production. That gene was cloned, expressed, and found likely to carry out a different function (see below).

The genome sequence was also annotated to search for alkane degradation genes. The logic behind this was that bacteria producing other energy-rich, carbon-rich molecules (polyhydroxyalkanoates, triacylglycerides, and glycogen) typically oxidize these carbon storage molecules (10, 104). Thus, we looked for the readily identifiable enzymes involved in alkane oxidation: cytochrome P450 monooxygenases and Alk proteins. None of these enzyme systems were identified. In BLAST searches, expectation values for homologs were generally greater than 1.0. Moreover, these systems are multicomponent and thus encoded by gene clusters. These gene clusters should be readily identifiable, if present, even if the sequences were fairly divergent.

Most of the genes examined have homologs in other *Vibrio* species that are not known to produce alkanes. One gene region that differed from other *Vibrio* species sequenced to date contained structural genes with highly significant sequence identity to metabolosomes, or carboxysomes. Metabolosomes are intracellular, multiprotein structures consisting of shell proteins harboring metabolic proteins (63). They were initially known as carboxysomes because carbon dioxide-fixing enzymes were found associated with the first metabolosomes identified. More recently, other types of metabolism have been found to be harbored by shell proteins homologous to carboxysome shell proteins. The carboxysome gene region in *V. furnissii* M1 was initially considered intriguing. It included genes encoding oxidoreductases and a pyruvate formate lyase homolog; the latter activity could conceivably be involved in an alcohol-to-alkane reduction reaction. However, bioinformatics analysis and comparison to a similar region in *E. coli* F11 led us to the tentative conclusion that the gene cluster

and metabolosome in *V. furnissii* M1 likely function in the fermentation of ethanol. This hypothesis was tested by cloning *V. furnissii* M1 ORF 275 in *E. coli* and assaying the protein extract from recombinant cells. The data indicated that ORF 275 encoded a bidirectional acetaldehyde dehydrogenase (CoA-acetylating) enzyme, consistent with its hypothetical role in an ethanol fermentation. While *E. coli* strain F11 has a homologous metabolosome-like structure, *E. coli* K-12 produces a multifunctional, spiral-shaped polypeptide that is thought to channel pyruvate to ethanol (64). The channeling multidomain protein and a metabolosome may represent different biological mechanisms for channeling metabolic flux through a potentially toxic aldehyde intermediate.

2.5 Conclusions

V. furnissii strains, including strain M1, were observed to have chromosomes of approximately 3.2 and 1.8 Mb. No apparent alkane-producing genes or phenotypes were observed. The latter was checked *in vivo* and *in vitro* with *V. furnissii* M1 and ATCC 35628 and *in vivo* with *V. furnissii* ATCC 35016.

2. 6 Acknowledgements

We acknowledge the help of Satoshi Ishii with the REP-PCR and Tao Yan with the PFGE. We thank Jack Richman for helpful advice on chemistry and Harry Beller of Lawrence Livermore National Laboratory for providing unpublished data.

This research was partly funded by a grant from the Institute for Renewable Energy and the Environment, grant LG-B13-IREE.

2.7 Publication of thesis work

This work was the efforts of many individuals (listed as published): Wackett, Lawrence P, Janice A Frias, Jennifer L Seffernick, Dave J Sukovich, and Stephan M Cameron. My contribution focused on alkane production verification *in vivo* and *in vitro*. Citation: 73 (22) Applied & Environmental Microbiology. 2007.

CHAPTER 3

C_{29} Olefinic Hydrocarbons Biosynthesized by *Arthrobacter* Species

3.1 Introduction

Bacterial hydrocarbon biosynthesis has garnered renewed interest in the context of generating new biofuels that are superior to ethanol (56, 71). A number of bacteria make long-chain, nonisoprenoid hydrocarbons that are being explored for biofuel and specialty chemical applications (118). An unusual class of unsaturated C_{22} to C_{31} alkenes produced by *Micrococcus* and *Stenotrophomonas* species was first described more than 40 years ago, but the biological function and mechanism of formation of these alkenes have not yet been elucidated (6, 111, 114) . In addition, the assignment of the structure for specific compounds has differed in different reports (74, 112). The precise structures of the compounds are relevant to understanding the biosynthetic mechanism and biological utility of these compounds. *Micrococcus* and

Stenotrophomonas make a complex mixture of alkenes; identifying new organisms with a simpler product profile should facilitate mechanistic and biological experiments. Moreover, discovering alkenes in new bacteria that have been subjected to complete genome sequencing should advance efforts to identify the relevant biosynthetic genes and enzymes.

Arthrobacter spp. are high-G+C-content gram-positive bacteria (62) for which several genome sequences are currently available (NCBI sequence accession number NC_008541 and NCBI sequence accession number ABKU00000000 (78)). In the present study, we screened cultures of divergent *Arthrobacter* species for the presence of hydrocarbons. Long-chain alkenes, reminiscent of a subset of those previously found in *Micrococcus*, were observed in several *Arthrobacter* species. *Arthrobacter* strains tested here were observed to produce a more uniform alkene chain length, predominantly C_{29}. To positively identify the products, C_{29} alkenes with different methyl branching patterns and with *cis* or *trans* stereochemistry were prepared by chemical synthesis (see Fig. 3-2). The corresponding alkanes were also synthesized. This set of 11 chemical standards allowed the identification of the products as specific dimethyl-13-heptacosenes with an unambiguous demonstration of a *cis* relative stereochemistry at the double bonds.

3.2 Demonstration of alkenes in cultures of *Arthrobacter* spp.

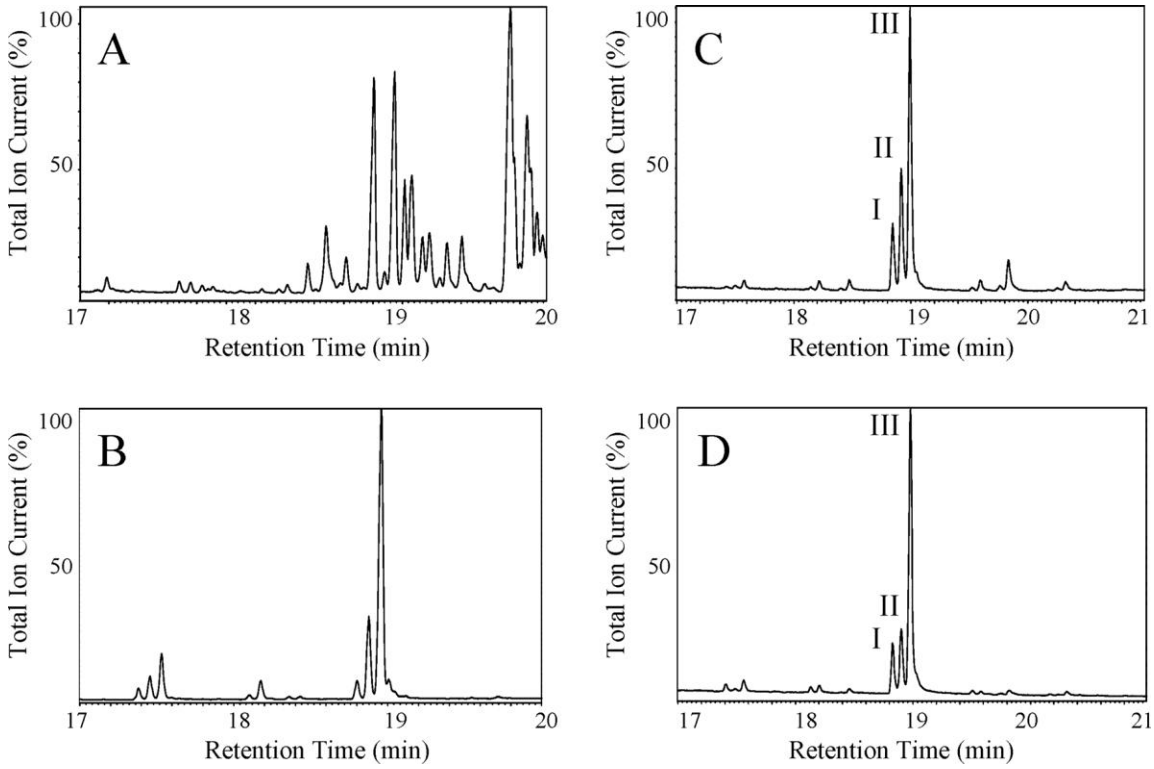


Figure 3-1. Gas chromatograms of nonpolar extracts from *Stenotrophomonas*, *Micrococcus*, and *Arthrobacter* strains showing the variety of alkenes produced by each strain. (A) *Stenotrophomonas maltophilia* ATCC 17674. (B) *Micrococcus luteus* ISU. (C) *Arthrobacter chlorophenolicus* A6. (D) *Arthrobacter crystallopoietes* ATCC 15481. The numbers I, II, and III refer to the major C_{29} alkene products, and details of their structures are described in the text.

Nonpolar material extracted from *Arthrobacter* strains was compared to the hydrocarbons produced by *Micrococcus luteus* ISU and *Stenotrophomonas* (formerly *Pseudomonas*) *maltophilia* ATCC 17674 (106) (Fig. 3-1). Four-day-old cultures were extracted by using a modified Bligh and Dyer protocol (13) as described previously (119). Evaporated extracts were dissolved in 1 ml chloroform and applied to a 3.5-g silica gel column, eluted with 30 ml hexanes, concentrated, and subjected to gas chromatography-mass spectrometry (GC-MS) analysis using an HP6890 gas chromatograph connected to an HP5973 mass spectrometer (Hewlett Packard, Palo Alto, CA). The GC conditions were helium gas, 1 ml/min; HP-1ms column (100% dimethylpolysiloxane capillary; 30 m by 250 μm by 0.25 μm); temperature ramp, 100 to 320°C; and 10°C/min, with a 5-min hold at 320°C. The MS was run in electron impact mode at 70 eV and 35 μA.

Unlike the more-complex *Stenotrophomonas* (Fig. 3-1A) and *Micrococcus* (Fig. 3-1B) hydrocarbon profiles, the large majority of the *Arthrobacter* hydrocarbons were apparent in three readily resolvable peaks that eluted in the narrow range of 18.8 to 19.1 min. The major hydrocarbons extracted from *Arthrobacter* strains, represented in Fig. 3-1C and D, were identified by MS as C_{29} monoalkenes (*m/z* = 406; $C_{29}H_{58}$). Treatment of the alkenes with hydrogen gas over a palladium catalyst produced saturated alkanes, and subsequent GC-MS gave slightly longer retention times and mass spectra that were consistent with a gain of two mass units for each (*m/z* = 408; $C_{29}H_{60}$).

Table 3-1. Distribution of alkene chain lengths detected by GC-MS for a variety of *Arthrobacter* strains and comparison to their distribution in two previously studied bacteria, *S. maltophilia* and *M. luteus*

Organism	**% of indicated chain length in strain**[a]						
	C_{25}	**C_{26}**	**C_{27}**	**C_{28}**	**C_{29}**	**C_{30}**	**C_{31}**
Stenotrophomonas maltophilia ATCC 17674	–	–	1.15	5.15	36.3	39.4	18.0
Micrococcus luteus ISU ATCC 27141	4.0	1.2	14.0	5.2	75.5	–	–
Arthrobacter aurescens TC1	–	–	–	–	79.7	12.7	7.6
Arthrobacter chlorophenolicus A6	–	–	1.5	4.5	80.3	11.6	2.1
Arthrobacter crystallopoietes ATCC 15481	–	–	3.9	2.9	93.2	–	–
Arthrobacter oxydans ATCC 14358	–	–	–	–	100	–	–
Arthrobacter nicotianae ATCC 15236	–	–	–	–	–	–	–
Arthrobacter sp. strain FB24	–	–	–	–	–	–	–
Arthrobacter sp. strain 1-NP	–	–	–	–	–	–	–
Arthrobacter globiformis ATCC 35698	–	–	–	–	–	–	–

[a]–, not detected.

Table 3-1 shows the relative distribution of different alkene chain lengths in *Stenotrophomonas*, *Micrococcus*, and eight *Arthrobacter* species. The *Arthrobacter* strains were obtained from other researchers (59, 80) or the American Type Culture Collection (ATCC) or were isolated in this laboratory (38, 105). The *Arthrobacter* strains generally showed a narrower distribution of chain lengths (Table 3-1). This trend was most pronounced with *A. oxydans* ATCC 14358, which produced only C_{29} alkenes. Four other *Arthrobacter* strains did not yield any detectable alkenes. One of those not producing alkenes, *Arthrobacter* sp. strain FB24, has been subjected to genome sequencing (NCBI sequence accession number CP000454). Two *Arthrobacter* strains producing alkenes have had their genomes sequenced (NCBI sequence accession numbers ABKU00000000 and NC_008711 (78)).

3.3 Rigorous assignment of structures to resolved C_{29} isomers.

The mass spectra of the major alkenes produced by *Arthrobacter* bacteria were consistent with their assignment as dimethylheptacosenes (*m/z* = 406; $C_{29}H_{58}$). *Arthrobacter* strains produce predominantly C_{15} methyl-branched fatty acids (21, 116), and thus, if a head-to-head fatty acid condensation mechanism were operative as previously proposed (2, 4), dimethylheptacosenes ($C_{29}H_{58}$) would be the anticipated products. *Arthrobacter* bacteria produce both iso and anteiso methyl-branched C_{15} fatty acids, and so the alkenes could be iso-iso, iso-anteiso, anteiso-iso, anteiso-anteiso, or some mixture of the different isomers (see Fig. 3-2). Hydrogenation of the alkenes provided preliminary evidence from fragmentation patterns (M-15 and M-43) that the

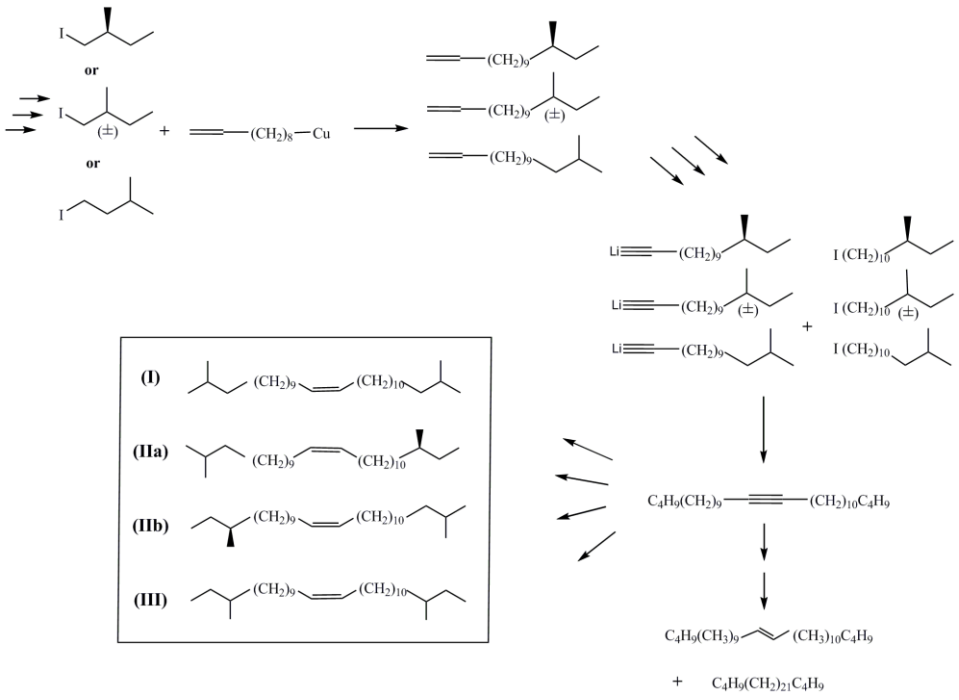


Figure 3-2. Synthetic scheme for preparing iso- and anteiso-*cis*-dimethyl-13-heptacosenes (I-III, in box) and corresponding alkyne, *trans*-alkene and alkanes (lower right).

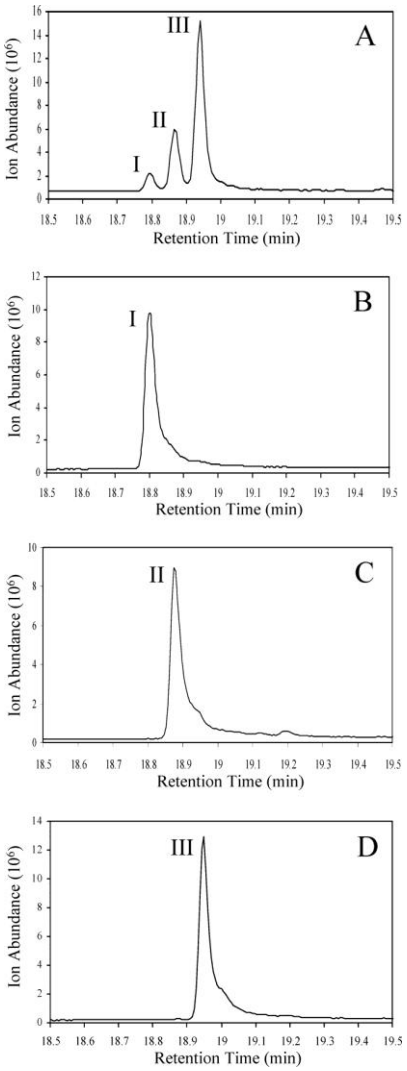


Figure 3-3. Gas chromatograms of the C_{29} alkene region from an *Arthrobacter chlorophenolicus* A6 extract shown in comparison with synthetic standards. (A) *Arthrobacter chlorophenolicus* A6 extract. (B to D) Synthetic standards: *cis*-2,26-dimethyl-13-heptacosene (B), a mixture of *cis*-2,25-dimethyl-13-heptacosene and *cis*-3,26-dimethyl-13-heptacosene (C), and *cis*-3,25-dimethyl-13-heptacosene (D).

18.8-min peak (I) was iso-iso branched, or 2,26-dimethylheptacosene (data not shown). The 19.0-min (major) peak (III) showed fragmentation (M-29 and M-59) consistent with an anteiso-anteiso branching that identifies the compound as 3,25-dimethylheptacosene (data not shown). The middle (18.9 min) peak (II) showed characteristics of iso-anteiso branching and thus could be identified as 2,25-dimethylheptacosene, 3,26-dimethylheptacosene, or a mixture of the two.

To help resolve these issues, and to determine the precise structures of the products, synthetic standards were prepared. Previous reports of alkenes biosynthesized by *Micrococcus luteus* indicated that the double bond of the alkenes was near the middle of the chain, based on their chemical degradation to fatty acids (3). The stereochemistry of the double bond was proposed to be *cis* based on their retarded migration on silica gel impregnated with silver (3). However, no synthetic chemical standards were available in those previous studies to compare the properties of *cis*- and *trans*-alkenes. Moreover, the same peak was assigned different structures in different studies (74, 112). In that context, *cis*- and *trans*-13-dimethylheptacosenes with all combinations of iso and anteiso branching patterns were synthesized (detailed synthetic conditions will be described elsewhere). The selective synthesis of *cis* or *trans* isomeric standards could be controlled by the synthetic methods used and was confirmed by nuclear magnetic resonance spectroscopy. Hydrogenation of different dimethylheptacosynes to the corresponding alkenes by using a Lindlar catalyst produced predominantly *cis*-olefins with trace amounts of the *trans* isomer, as shown in Fig. 3-3B, C, and D. The *trans* isomers eluted as a shoulder on the tail end of each *cis*-

isomer peak. A more-aggressive hydrogenation reaction using 5% Pd/C catalyst produced predominately *trans* isomers. Long-term hydrogenation with 5% Pd/C led to complete reduction, yielding the corresponding alkanes. In this manner, eight different dimethyl-13-heptacosene standards and three dimethylheptacosane standards were synthesized. All of these were used as GC standards to determine retention times and mass spectra by GC-MS. This allowed the identification of the three major peaks as (Fig. 3-3A, left to right) *cis*-2,26-dimethyl-13-heptacosene (I), either *cis*-2,25-dimethyl-13-heptacosene or *cis*-3,26-dimethyl-13-heptacosene or a mixture of the two (II), and *cis*-3,25-dimethyl-13-heptacosene (III).

The structural identifications made via separate injections were further confirmed by coinjection of standards in admixture with biological material on a GC. Coinjection of the respective standards gave uniform peaks, thus confirming the identity of the biological material eluting in peaks I and III as described above. Synthetic *cis*-2,25-dimethyl-13-heptacosene and *cis*-3,26-dimethyl-13-heptacosene had identical retention times and similar mass spectra, consistent with the conclusion that peak II could be either one of the compounds by itself or a mixture of the two.

3.4 Growth studies

Micrococcus bacteria are spherical cells at all growth stages, whereas *Arthrobacter* species grow as rod-shaped cells during the exponential phase and become spherical during stationary phase (62). Thus, it was considered that alkenes

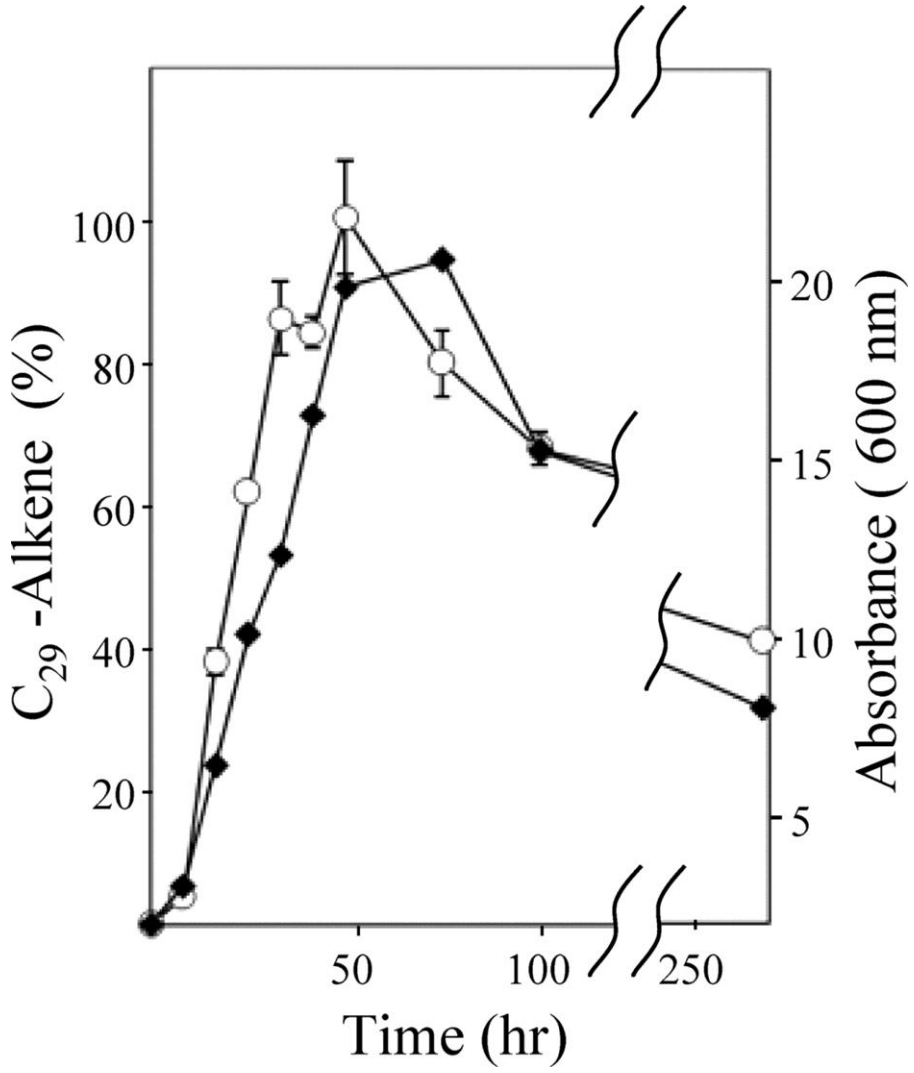


Figure 3-4. Time course of culture optical density and C_{29} alkene accumulation in *Arthrobacter chlorophenolicus* A6. Alkene accumulation was determined by quantifying *cis*-3,25-dimethyl-13-heptacosene and normalizing the maximum concentration, found at 48 h, to 100%. The open circles denote alkene percentage. Closed diamonds represent absorbance. Error bars indicate standard deviations divided by *n*.

might be preferentially formed by *Arthrobacter* bacteria during stationary phase. To test this hypothesis, a 50-µl preculture was used to inoculate 50 ml tryptic soy broth in a 125-ml Erlenmeyer flask. Cultures were grown at 28°C with shaking at 225 rpm. Duplicate cultures for growth studies were extracted at the time points of 0 h, 8 h, 16 h, 24 h, 32 h, 48 h, 96 h, and 288 h. The cells underwent a characteristic rod-to-coccus transition at 48 h. Prior to extraction, 62.5 µmol of *cis*-9-tricosene was added as an internal standard to quantitatively determine the levels of 3,25-dimethylheptacosene. The data obtained from these extractions showed that the alkene levels closely paralleled culture optical density (Fig. 3-4). Thus, both rod and spherical cell forms contain similar levels of 3,25-dimethylheptacosene.

Arthrobacter chlorophenolicus A6 was tested for growth on *S,S-cis*-3,25-dimethyl-13-heptacosene, the major isomer thought to be formed biosynthetically by that strain. *A. chlorophenolicus* A6 was grown in M9 minimal medium [95] containing either glucose or *S,S-cis*-3,25-dimethyl-13-heptacosene, added to a concentration equivalent to 130 mM of carbon. Phenol was also used as a positive control, at 2 mM, 10 mM, and 20 mM in different cultures (80). Growth studies with each carbon source were set up in triplicate. For the inoculum, cells were grown overnight in tryptic soy broth and washed twice with M9 minimal medium without a carbon source. The medium was inoculated to an optical density of 0.07 at 600 nm and grown in capped test tubes at 28°C with shaking at 255 rpm. Optical density measurements were taken at 600 nm (Beckman DU 7400) after 17 days of growth.

No discernible cell growth was supported by *cis*-3,25-dimethyl-13-heptacosene. The average optical density at 600 nm in the test cultures was 0.049 ± 0.001 (mean ± standard deviation). That compared to an average of 0.042 ± 0.001 in a control without alkene. With phenol substituted for the alkene, the optical density at the same time point was 1.04 ± 0.03. With D-glucose as the carbon source, the optical density was 1.70 ± 0.18. These data suggested that the long-chain alkenes are not produced for the function of storing carbon or energy. The observation in this study that some *Arthrobacter* strains produce long-chain alkenes and others do not (Table 3-1) indicated that these compounds are not essential under the laboratory growth conditions used.

3.5 Conclusions

This study has shown that some *Arthrobacter* strains produce C_{29} olefinic hydrocarbons, the structures of which were rigorously established by comparison with synthetic standards. In one previous study of *Micrococcus* bacteria, it was noted that one *Arthrobacter* strain, now identified as *Arthrobacter citreus*, also produced alkenes and that other *Arthrobacter* strains did not (66). No data were shown. However, that report, coupled with the present study, suggests that alkene formation is not ubiquitous amongst *Arthrobacter* species, in contrast to *Micrococcus* strains, which appear to uniformly produce olefinic hydrocarbons. Complete genome sequences are available for *Arthrobacter aurescens* TC1 (78) and *Arthrobacter chlorophenolicus* A6 (NCBI sequence accession numbers NC_00871 and ABXU00000000), which produced

alkenes, and for *Arthrobacter* sp. strain FB24 (NCBI sequence accession number NC_008541), which did not. These observations pave the way to use comparative genomic analysis to identify alkene-biosynthetic genes in these microorganisms.

3.6 Acknowledgements

We thank Cindy Nakatsu and Janet Jansson for providing *Arthrobacter* sp. strain FB24 and *Arthrobacter chlorophenolicus* A6, respectively.

This work was supported by NIH Training Grant T32 GM08347 and NIH Training Grant 5 T32 GM008700 (to J.A.F.) and by grant LG-B13 from the Institute for Renewable Energy and the Environment and a Discovery Grant from the Institute of the Environment (to L.P.W.).

3.7 Publication of thesis work

This work was published: Frias, Janice A., Jack E. Richman, and Lawrence P. Wackett. C_{29} Olefinic Hydrocarbons Biosynthesized by *Arthrobacter* Species. 75 (6) Applied & Environmental Microbiology. 2009. Jack E. Richman designed and completed the synthesis of the standards; his synthetic scheme is depicted in Figure 3-2.

CHAPTER 4

Studies on expressing the *Arthrobacter oleA* gene

4.1 Introduction

Alkene biosynthesis by bacteria is intrinsically interesting and has application in biofuels research. The finding that *Arthrobacter* strains synthesized hydrocarbons (47) stimulated the search for the relevant genes in *Arthrobacter* strains for which genome sequences are available. The complete genome sequence of *Arthrobacter aurescens*, studied because of its ability to degrade atrazine, has recently been reported (105). While those studies were in progress, a U.S. patent application was published (49). This facilitated identification of the *ole* gene cluster. Some bacteria contain four genes, *oleABCD*. *Arthrobacter* sp.have a 3 gene cluster, in which *oleB* and *oleC* are a fusion gene. The *oleA* gene encodes for a thiolase protein with 63% and 59% similarity to the *Micrococcus* and *Stenotrophomonas* genes, respectively. The goal of this research was to clone and express one or more of the *ole* genes in an effort to study individual reactions and the olefin biosynthetic pathway.

4.2 Methods

4.2.1 Cloning thiolase

DNA for cloning was obtained as a plasmid from the sequencing of *Arthrobacter aurescens* TC1 by The Institute for Genomic Reseacrh (TIGR) (78), in which the clones could be identified by TIGR as having the thiolase gene sequence (AAur_1998) on the plasmid. Plasmid DNA was prepared using the Qiagen Miniprep Kit. PCR was performed using the following primers to add an NdeI and HindIII site: 5'-CATATGGCAGGGAATGC GACCTTCCGGCACAGC-3'and 5'-CCGACGTG AAGCTTACGCGGAGTGT CCGGGAC-3'. The PCR program was as follows: initial denaturation at 98°C for 5 min, followed by 30 cycles of denturation at 98°C for 30 sec, annealing at 58°C for 30 sec, and extension at 72°C for 3 min, and then a final extension at 72°C for 10 min after the last round of cycling. PCR reaction included: 10 µl Herculase II polymerase buffer, 4 µlDMSO, 1 µl forward primer, 1 µl reverse primer, 0.5 µl dNTPS (25 mM stock), 30.5 µl water, 2 µl plasmid prep from Tiger sequencing plasmid containing the thiolase gene, 1 µl Herculase II polymerase. The Qiagen Gel Extraction kit was used to purify the excised bands running at approximately 1100 bp by agarose gel electrophoresis.

The thiolase was subcloned using the Strataclone cloning protocol. Poly A tails were added and then the fragment was ligated into the Strataclone vector with poly T sticky ends. Plasmids were then transformed into StrataClone SoloPack competent cells and colonies screened by blue/white colony formation due to the addition of Xgal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) on the LB ampicillan (150 µg/ml) agar plates. Once the gene was inserted and verified in the Strataclone vector, the thiolase was digested out using the restriction enzymes NdeI and HindIII. The

expression vector pET28b+ (Novagen, Madison, WI) was cut with NdeI and HindIII and dephosphorylated in order to ligate the thiolase into the vector. After ligation of the thiolase into pET28b+, vectors were transformed into *E. coli* Max Efficiency DH5α and then into *E. coli* BL21 (DE3) pLysE for expression.

4.2.2 Expression studies and purification

E. coli BL21 (DE3) pLysE cells with pET28b AAur_1998 were induced in 50 ml LB kanamycin (50 µg/ml) and chloramphenicol (34 µg/ml) shake flasks grown to an OD_{600} of 0.5-1.0 at 37°C. Cells were induced with 0.1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) for 3, 4, or 5 hours and then soluble and insoluble fractions were analyzed by SDS-PAGE to assess expression of the thiolase of MW = 36,158 Da.

E. coli BL21 (pAAur_1998) cells were cultured in 500 ml LB medium containing kanamycin (50 µg ml^{-1}) and chloramphenicol (34 µg ml^{-1}) at 37°C. Cultures were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM when the OD_{600} of the culture reached 0.55-0.70. After 24 h at 37°C, the induced cells were harvested by centrifugation at 3000×*g* for 25 min and resuspended in 20 mM Tris, 500 mM NaCl, 1 mM β-mercaptoethanol, pH 7.4 buffer. The cells were disrupted by three passes through a chilled French pressure cell at 1200 psi and centrifuged at 27 000×*g* for 90 min to remove cell debris and insoluble protein. The soluble fraction was filtered through a 0.45 µm membrane prior to loading onto a

Pharmacia Biotech LCC 501 FPLC fitted with a 5 ml HisTrap HP (Amersham Biosciences) column complexed with Ni^{2+} and equilibrated with 20 mM Tris, 500 mM NaCl, and 1 mM β-mercaptoethanol, pH 7.4. The his-tagged thiolase protein eluted at 150 mM imidazole. The purity of the protein was assessed by SDS-PAGE, with a dark band running approximately 45-50 kDa and a second higher band running at approximately 70 kDa. Protein was dialyzed at 4°C against buffer without imidazole prior to assessing activity. To check that the higher molecular weight band was not due to resistant disulfide dimers, protein samples from this purification were also analyzed by SDS-PAGE in which the samples were subjected to higher levels of DTT, 25 mM and 100 mM, in order to ensure that disulfide bonds leading to dimerization would be reduced. The 2 bands were cut from the gel and processed for mass spectrometry identification by Todd Markowski in the Center for Mass Spectrometry and Proteomics (University of Minnesota).

Varying induction temperature was also assessed using 50 ml shake flask culture of *E. coli* BL21 (pAAur_1998) in order to prevent inclusion body formation over soluble protein. Cultures were grown at 37°C and induced as described in the previous section, then flasks were moved to 16°C or 30°C and harvested at 2, 4, and 24 hours. IPTG levels were also varied from 0.1 mM to 0.3, 0.4, and 0.5 mM. FPLC purification was evaluated on cultures induced at 30°C for 2 hrs with 0.3 mM IPTG using a different buffer system: 20 mM sodium phosphate, 500 mM NaCl, pH 7.4.

4.2.3 Enzyme Assays

Cultures were initially grown in shake flasks and extracted (47) for GC-MS analysis to probe for evidence of condensation of fatty acids by the thiolase gene, resulting in ketones as reported by LS9 (49). Cultures were induced with 0.1 mM IPTG at 37°C for 6 hrs and 24 hrs and peak differences compared to a no insert control were evaluated.

DTNB assays were used to assess the activity of the purified thiolase protein to release CoA on the condensation of substrate. Assays were performed with palmitoyl-CoA and the addition of a purification fraction from the 24 hr induction at 37°C that contains the thiolase and a fraction that does not contain thiolase as a control. Another variation included *Micrococcus luteus* ISU cell extract that had been put through a PD-10 Desalting Column (GE Healthcare) to remove $M_r < 5000$.

A radiolabelled assay was developed to detect if the acyl-CoA substrate with a 1-$^{14}C_1$ label was bound in the active site of the protein. The assay was incubated at room temperature with the ^{14}C substrate and then spun through the PD-10 column, to remove any compounds $M_r < 5000$. The eluant radioactivity was determined in a Beckman scintillation counter and compared to controls with boiled protein to determine if any label remained bound to the protein. The ^{14}C counts that eluted in samples with protein and labeled myristoyl-CoA was below background, therefore showing no binding.

4.2.4 Re-cloning thiolase with C terminal his tag

In attempts to improve folding, the AAur_1998 was alternatively cloned into pET28b+ with a C terminal his tag rather than an N terminal his tag. In order to remove the N terminal tag added by the vector, the gene needed to be cloned in at the NcoI sight. The C terminus restriction sight was HindIII with the stop codon removed. Overnight 5 and 10 ml cultures were prepped from a freshly streaked *Arthrobacter aurescens* LB plate grown overnight at 30°C. The Qiagen Blood and Tissue kit was used following the protocol for preparing DNA from gram positive bacteria. Recently prepared lysing buffer was used in this protocol: 20 mM Tris, 1.2% Triton X-100, and 2 mM NaEDTA. Approximately 2×10^9 cells were used, calculated from OD_{600} as described in the protocol.

Primers used to add NcoI and HindIIi sites, as well as removing stop codon: 5'-CCATGGCAGGAATGCGACCTTCCGG-3' (forward) and 5'-GGGCCAGTTGG AAGCTTCCAAACGATTTCC-3' (reverse). The PCR program was as follows: initial denaturation at 98°C for 2 min, followed by 30 cycles of denturation at 98°C for 30 sec, annealing at 65°C for 30 sec, and extension at 72°C for 2.5 min, and then a final extension at 72°C for 10 min after the last round of cycling. PCR reaction included: 10 µl GC Phusion polymerase buffer, 1.5 µlDMSO, 0.5 µl forward primer, 0.5 µl reverse primer, 2 µl dNTPS (10 mM stock), 33 µl water, 2 µl genomic DNA, 0.5 µl Phusion polymerase. A 40 ml TAE and 1% agarose gel was run at 90V for 1 hr to cut the band for gel extraction. The Qiagen kit was used to clean the DNA from the gel and the Poly A tails were added per the Strataclone protocol. Reaction included: 20 µl PCR product, 23.5 µl water, 5 µl Taq/PCR buffer, 1 µl dNTPs (10 mM), 0.5 µl MgCl2 (50

mM) and 1 µl Taq. Reaction was held at 72°C for 15 min. PCR product was ligated into Strataclone vector using 2µl of the PolyA tail reaction and transformed per protocol. Cells were plated on LB ampicillan (150µg/ml) and incubated at 37°C overnight. Proceeded to identify potential clones by blue/white screen and set up overnights. Digested plasmid preps as follows: 25 µl prep, 3 µl Invitrogen React2 buffer, 0.5 µl NcoI, and 0.5 µl HindIII. Vector was prepared with pET28b+ plasmid prep, digested with NcoI and HindIII, and dephosphorylated. Thiolase gene was ligated into the prepped vector and transformed into *E. coli* Max Efficiency DH5α and then into *E. coli* BL21 Rosetta 2 cells with the pRARE plasmid.

4.2.5 Expression with pRARE plasmid

AAur_1998 gene was assessed by Graphical Codon Usage Analyser 2.0 (GCUA 2.0) for expression in *E. coli* and 6 codons were found to be codons that are rare to *E. coli* and could potentially cause issues for soluble expression. Of the 6 codons, 5 of them were arginine, using AGG or AGA, and the final codon was the stop codon, TAG. TAG was suggested to be used only rarely by *E. coli* due to selection upon the termination rate or translational read through (Sharp and Bulmer 1988). Using pRARE addressed the rare Arg codon and the stop codon was removed by the primer for the C terminal his tag.

Expression of the C terminal his tagged protein was evaluated at induction temperatures of 16, 30 and 37°C at 2, 4 and 24 hour induction times. The vector

prepped with the N terminal tag was also transformed into the *E. coli* BL21 Rosetta 2 plasmid as well and evaluated for expression at 16°C. The N terminal his tagged thiolase was purified after a 4 hr 16°C induction with 0.05 mM IPTG on the FPLC as described in section 4.2.2, using 20 mM sodium phosphate, 500 mM NaCl pH 7.4 buffer.

4.3 Results

Initial studies in shake flasks resulted in visible expression of a protein running on SDS-PAGE between 36 kDa and 53 kDa, near the thiolase molecular weight of 36 kDa (Figure 4-1). Figure 1A shows that a significant percentage of the overexpressed protein is in the insoluble fraction of the cell lysate. The soluble fraction does contain significant expression and was subjected to FPLC purification.

The FPLC purification resulted in no significant peaks in the 50 mM imidazole step increase program, but fractions eluting at 150 mM and 200 mM were run on an SDS-PAGE gel and showed elution of the band of interest (Figure 4-2A). Another highly overexpressed band appeared to be eluting similarly to the band of interest, between 53 and 78 kDa. In order to be certain this was not an unreduced dimer formed through reduction-resistant disulfide bonds, protein fractions were prepared with higher levels of reducing agent, DTT (Figure 4-2B). When this resulted in the same pattern of bands, the bands were cut from the gels for mass spectrometry analysis to identify these proteins.

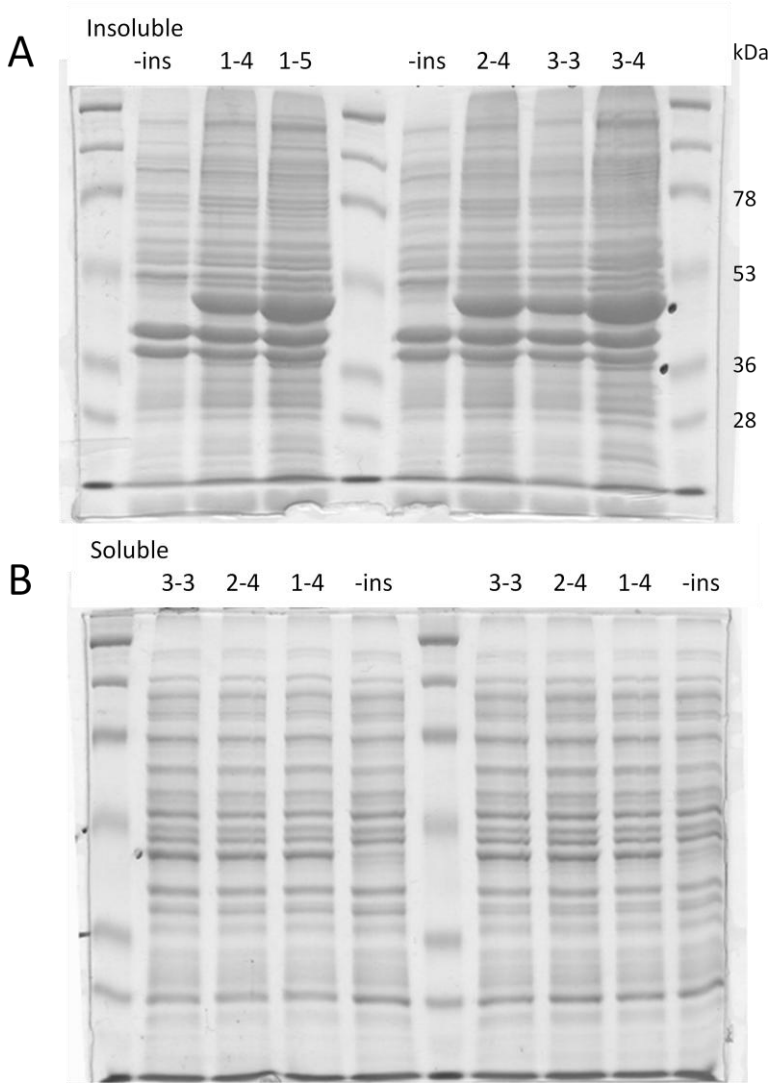


Figure 4-1. 10% Tris-HCl SDS-PAGE of (A) insoluble and (B) soluble *E. coli* fractions of crude lysate containing thiolase gene from *Arthrobacter*. An overexpressed band can be seen between 36 kDa and 53 kDa molecular weight markers.

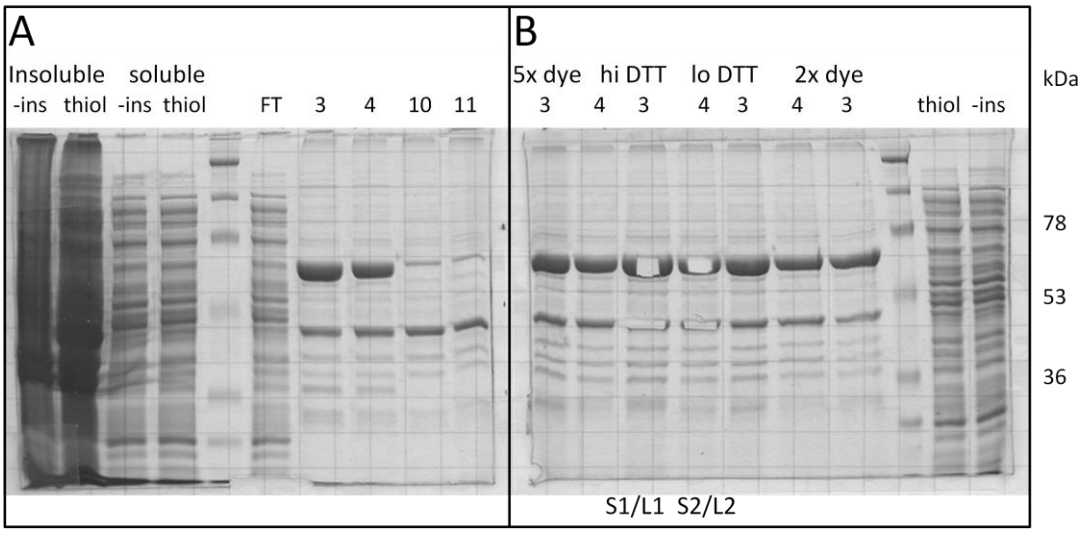


Figure 4-2. 10% Tris-HCl SDS-PAGE gel of FPLC fractions (A) and fractions 3 and 4 with varying levels of reducing agent, DTT (B). The bands in lanes 3 and 4 were cut for mass spectrometer anaylsis to determine protein identity.

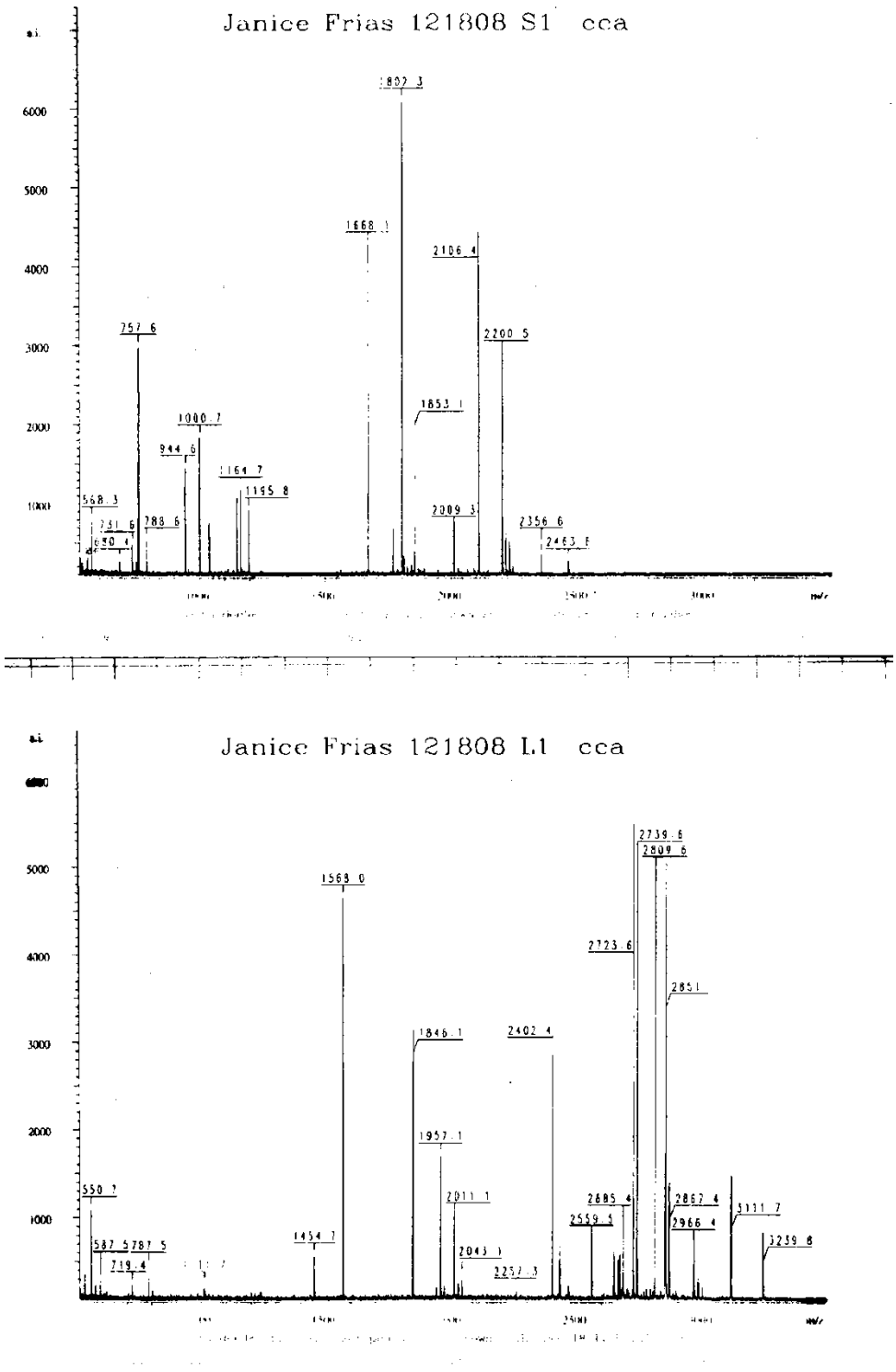


Figure 4-3. Mass Spectrometry results of the processed gel slices.

Mascot Search Results

Protein View

Match to: **gi|119962129** Score: **221** Expect: **3e-16**
3-oxoacyl-(acyl carrier protein) synthase III [Arthrobacter aurescens TC1]

Nominal mass (M_r): **36437**; Calculated pI value: **5.27**
NCBI BLAST search of gi|119962129 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Arthrobacter aurescens TC1
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|119948988 from Arthrobacter aurescens TC1

Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: **28**
Number of mass values matched: **19**
Sequence Coverage: **49%**

Matched peptides shown in **Bold Red**

```
    1 MAGNATFRHS NTALLSVSSV EAPRIVSSTD FDRRLASTLQ RLKFPPRLLE
   51 RVAGITHRRW WAAGTSFDDA AVEAGAKALA EAGVEASEVG LLINTSVTRR
  101 NLEPSVAVKI HHELGLPSSA MNFDLANACL GFVNGLILAA NMIDSGQIRY
  151 AVIVNGEDAQ GTQEATLARL QRPETTREDF NREFATLTLG SGAAAAVLGP
  201 RDEYPGAHRL VGGVMRAGTE HHELCVGGID GMSTDTKGLL DGGLQLVVDA
  251 WHEAQPEWDW ASMDRYVTHQ VSNAYTQAII NAIDLDPDKV PITFPHWGNV
  301 GPASLPMTLA AQAQSLTSGD RVLCMGVGSG LNAGMVEIVW
```

[redacted] ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Sequence
9 - 24	1668.1000	1667.0927	1666.8638	137	0	R.HSNTALLSVSSVEAPR.I
25 - 33	1039.7000	1038.6927	1038.4982	187	0	R.IVSSTDFDR.R
25 - 34	1195.8000	1194.7927	1194.5993	162	1	R.IVSSTDFDRR.L
34 - 41	944.6000	943.5927	943.5563	39	1	R.RLASTLQR.L
35 - 41	788.6000	787.5927	787.4552	175	0	R.LASTLQR.L
42 - 47	757.6000	756.5927	756.4646	169	1	R.LKFPPR.L
52 - 58	753.6000	752.5927	752.4293	217	0	R.VAGITHR.R
59 - 77	2009.3000	2008.2927	2007.9439	174	1	R.RWWAAGTSFDDAAVEAGAK.A
60 - 77	1853.1000	1852.0927	1851.8428	135	0	R.WWAAGTSFDDAAVEAGAK.A
78 - 99	2200.5000	2199.4927	2199.1747	145	0	K.ALAEAGVEASEVGLLINTSVTR.R
78 - 100	2356.6000	2355.5927	2355.2758	135	1	K.ALAEAGVEASEVGLLINTSVTRR.N
150 - 169	2106.4000	2105.3927	2105.0389	168	0	R.YAVIVNGEDAQGTQEATLAR.L
170 - 177	1000.7000	999.6927	999.5461	147	0	R.LQRPETTR.E
178 - 182	680.4000	679.3927	679.2925	147	0	R.EDFNR.E
178 - 201	2463.6000	2462.5927	2462.2554	137	1	R.EDFNREFATLTLGSGAAAAVLGPR.D
183 - 201	1802.3000	1801.2927	1800.9734	177	0	R.EFATLTLGSGAAAAVLGPR.D
210 - 216	731.6000	730.5927	730.4160	242	0	R.LVGGVMR.A
210 - 216	747.6000	746.5927	746.4109	244	0	R.LVGGVMR.A Oxidation (M)
217 - 237	2215.3000	2214.2927	2213.9682	147	0	R.AGTEHHELCVGGIDGMSTDTK.G

No match to: 568.3000, 1148.7000, 1164.7000, 1769.1000, 1811.1000, 1839.2000, 2229.3000, 2230.3000, 2245.3000

Figure 4-4. Mascot search results of mass spectrometry fragments of L1 protein band running between 36 and 53 kDa by SDS-PAGE.

Mascot Search Results

Protein View

Match to: **gi|15804735** Score: **172** Expect: **2.4e-11**
chaperonin GroEL [Escherichia coli O157:H7 EDL933]

Nominal mass (M_r): **57447**; Calculated pI value: **4.81**
NCBI BLAST search of gi|15804735 against nr
Unformatted sequence string for pasting into other applications

Taxonomy: Escherichia coli O157:H7 EDL933
Links to retrieve other entries containing this sequence from NCBI Entrez:
gi|25300357 from Escherichia coli
gi|12519125 from Escherichia coli O157:H7 EDL933

Fixed modifications: Carbamidomethyl (C)
Variable modifications: Oxidation (M)
Cleavage by Trypsin: cuts C-term side of KR unless next residue is P
Number of mass values searched: **36**
Number of mass values matched: **27**
Sequence Coverage: **53%**

Matched peptides shown in **Bold Red**

```
     1 MAAKDVKFGN DARVKMLRGV NVLADAVKVT LGPKGRNVVL DKSFGAPTIT
    51 KDGVSVAREI ELEDKFENMG AQMVKEVASK ANDAAGDGTT TATVLAQAII
   101 TEGLKAVAAG MNPMDLKRGI DKAVTAAVEE LKALSVPCSD SKAIAQVGTI
   151 SANSDETVGK LIAEAMDKVG KEGVITVEDG TGLQDELDVV EGMQFDRGYL
   201 SPYFINKPET GAVELESPFI LLADKKISNI REMLPVLEAV AKAGKPLLII
   251 AEDVEGEALA TLVVNTMRGI VKVAAVKAPG FGDRRKAMLQ DIATLTGGTV
   301 ISEEIGMELE KATLEDLGQA KRVVINKDTT TIIDGVGEEA AIQGRVAQIR
   351 QQIEEATSDY DREKLQERVA KLAGGVAVIK VGAATEVEMK EKKARVEDAL
   401 HATRAAVEEG VVAGGGVALI RVASKLADLR GQNEDQNVGI KVALRAMEAP
   451 LRQIVLNCGE EPSVVANTVK GGDGNYGYNA ATEEYGNMID MGILDPTKVT
   501 RSALQYAASV AGLMITTECM VTDLPKNDAA DLGAAGGMGG MGGMGGMM
```

[redacted]

[redacted] ◉ Residue Number ○ Increasing Mass ○ Decreasing Mass

Start - End	Observed	Mr(expt)	Mr(calc)	ppm	Miss	Sequence
59 - 75	2011.1000	2010.0927	2009.9438	74	1	R.EIELEDKFENMGAQMVK.E
59 - 75	2027.1000	2026.0927	2025.9387	76	1	R.EIELEDKFENMGAQMVK.E Oxidation (M)
59 - 75	2043.1000	2042.0927	2041.9336	78	1	R.EIELEDKFENMGAQMVK.E 2 Oxidation (M)
81 - 105	2402.4000	2401.3927	2401.2336	66	0	K.ANDAAGDGTTTATVLAQAIITEGLK.A
172 - 197	2851.4000	2850.3927	2850.3230	24	0	K.EGVITVEDGTGLQDELDVVEGMQFDR.G
172 - 197	2867.4000	2866.3927	2866.3179	26	0	K.EGVITVEDGTGLQDELDVVEGMQFDR.G Oxidation (M)
198 - 225	3111.7000	3110.6927	3110.6216	23	0	R.GYLSPYFINKPETGAVELESPFILLADK.K
198 - 226	3239.8000	3238.7927	3238.7165	24	1	R.GYLSPYFINKPETGAVELESPFILLADKK.I
243 - 268	2723.6000	2722.5927	2722.4939	36	0	K.AGKPLLIIAEDVEGEALATLVVNTMR.G
243 - 268	2739.6000	2738.5927	2738.4888	38	0	K.AGKPLLIIAEDVEGEALATLVVNTMR.G Oxidation (M)
278 - 284	719.4000	718.3927	718.3398	74	0	K.APGFGDR.R
286 - 311	2777.5000	2776.4927	2776.4238	25	1	R.KAMLQDIATLTGGTVISEEIGMELEK.A
286 - 311	2793.5000	2792.4927	2792.4187	27	1	R.KAMLQDIATLTGGTVISEEIGMELEK.A Oxidation (M)
286 - 311	2809.6000	2808.5927	2808.4136	64	1	R.KAMLQDIATLTGGTVISEEIGMELEK.A 2 Oxidation (M)
287 - 311	2649.4000	2648.3927	2648.3288	24	0	K.AMLQDIATLTGGTVISEEIGMELEK.A
287 - 311	2665.4000	2664.3927	2664.3238	26	0	K.AMLQDIATLTGGTVISEEIGMELEK.A Oxidation (M)
328 - 345	1846.1000	1845.0927	1844.9116	98	0	K.DTTTIIDGVGEEAAIQGR.V
351 - 362	1454.7000	1453.6927	1453.6321	42	0	R.QQIEEATSDYDR.E
396 - 404	1011.7000	1010.6927	1010.5145	176	0	R.VEDALHATR.A
405 - 421	1568.0000	1566.9927	1565.8730	76	0	R.AAVEEGVVAGGGVALIR.V
426 - 430	587.5000	586.4927	586.3438	254	0	K.LADLR.G
446 - 452	787.5000	786.4927	786.4058	111	0	R.AMEAPLR.Q
453 - 470	1957.1000	1956.0927	1955.9986	48	0	R.QIVLNCGEEPSVVANTVK.G
471 - 498	2966.4000	2965.3927	2965.2746	40	0	K.GGDGNYGYNAATEEYGNMIDMGILDPTK.V
471 - 498	2982.4000	2981.3927	2981.2695	41	0	K.GGDGNYGYNAATEEYGNMIDMGILDPTK.V Oxidation (M)
471 - 498	2998.5000	2997.4927	2997.2644	76	0	K.GGDGNYGYNAATEEYGNMIDMGILDPTK.V 2 Oxidation (M)
502 - 526	2670.4000	2669.3927	2669.3114	30	0	R.SALQYAASVAGLMITTECMVTDLPK.N

No match to: 522.7000, 550.7000, 568.3000, 2257.3000, 2431.4000, 2467.4000, 2559.5000, 2685.4000, 2894.4000

Figure 4-5. Mascot search results of mass spectrometry fragments of S1 protein band running between 53 and 78 kDa by SDS-PAGE.

The mass spectrometry results for both the larger and smaller sized proteins, (Figure 4-3) were used to search MASCOT for proteins in the database that matched sequence fragmentation patterns. The top results for the band running between 36 and 53 kDa, were the *Arthrobacter* thiolase, AAur_1998, from which the gene had been cloned (Figure 4-4). Although the band appeared to be running higher than one would expect for a 36 kDa protein, it was confirmed to be the thiolase of interest. The top hits for the band running between 53 and 78 kDa, were the *E. coli* GroEL protein (Figure 4-5). This protein is found natively in *E. coli* and functions as a chaperone protein to help protein folding and is often upregulated in overexpression. Fraction 16 in Figure 4-6 is the fraction eluting at 125 mM imidazole and a large percentage of the GroEL appears to elute with the protein of interest. The increase to 150 mM imidazole still has considerable contamination of the GroEL, but it is greatly reduced. The final protein yield was quite low, resulting in 1-2 ml of 1 mg/ml from fractions eluting at 150 mM imidazole. The % thiolase protein in the insoluble fraction is high, ultimately resulting in significant losses in protein yield.

Expressing the thiolase with pRARE did seem to have some positive effect on expression. Soluble expression with the N terminal his tagged protein in 50 ml shake flasks appeared to increase. Potentially also indicative of increased solubility, the co-elution with GroEL was reduced (Fig 4-7). The C terminal his tag, regardless of the pRARE addition, appeared to be reducing the levels of soluble expression in comparison to the N terminal his tag, and the approach was quickly abandoned. Activity assays were performed along the route to improving the purity of the thiolase,

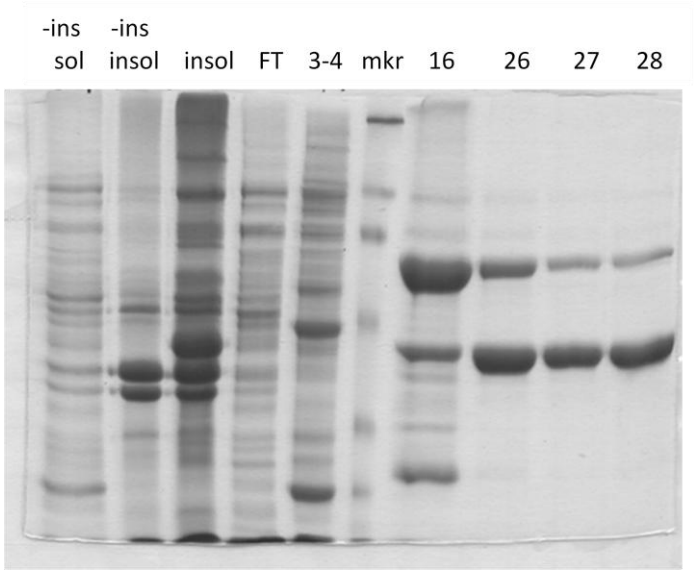
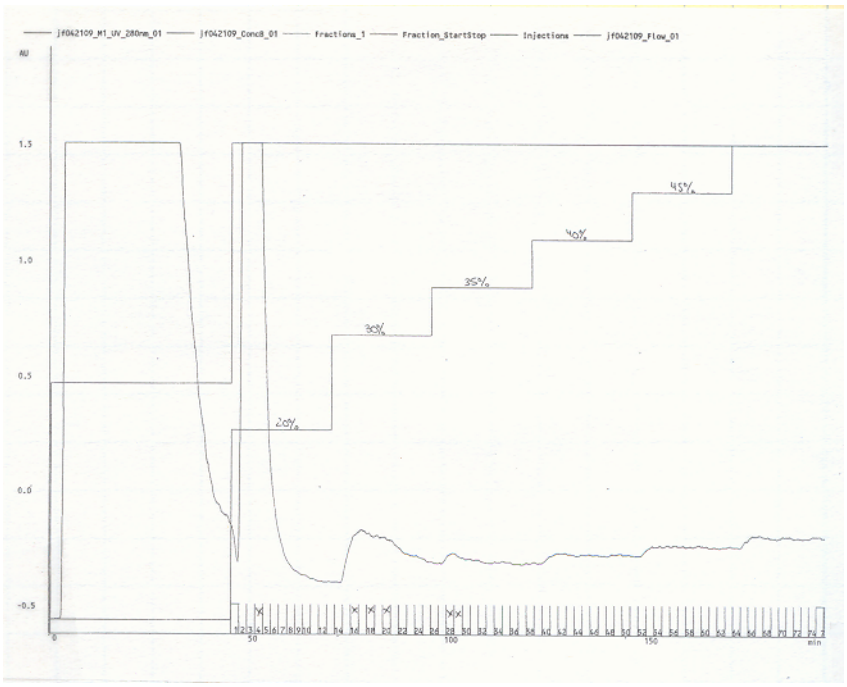


Figure 4-6. FPLC fractions with modified elution program, induction at 30°C for 2 hr.

A



B

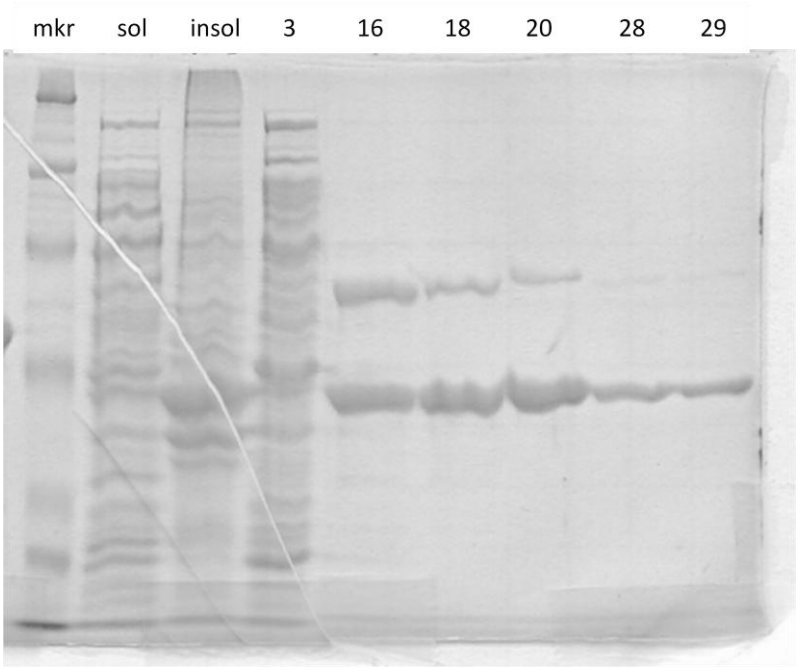


Figure 4-7. Purification results with BL21 Rosetta 2 cells, 16°C induction for 4 hr. (A) Chromatogram labeling the % in each step of Buffer B containing 500 mM imidazole. (B) Gel results of select fractions.

and no significant CoA release was observed above background. In addition to no observable activity, purified protein yield was still quite low.

4.4 Discussion

Concern in the ability to produce active protein was two-fold. The protein appeared to be running closer to 45 kDa than its true weight 36 kDa and did not appear to be active in releasing CoA. The definitive identity of the thiolase protein from *Arthrobacter* was unconfirmed. The second concern was that the protein did not appear to be folding properly.

The question of the protein's identity was quickly addressed with the mass spectrometry analysis of the overexpressed band running identified by SDS-PAGE. The mass spectrometry data confirmed the identity of the protein was indeed the thiolase cloned from *Arthrobacter* and expressed with a his-tag for purification ease. Another upregulated protein band was analyzed by MS and found to be the *E. coli* chaperone protein, GroEL. This protein was not his-tagged or overexpressed, so this highlighted that the thiolase folding was problematic and the GroEL could even be associated with the thiolase when the cells are lysed.

Improving the solubility of the *Arthrobacter* thiolase proved to be significantly more challenging. Due to the fact that the protein appeared to be expressing well, but not folding properly, the issue was first addressed by looking at lower temperature inductions to slow down the expression and increase proper folding. Figure 4-6 was

the result of purification from a 30°C induction. More problematic than the low soluble protein yield was the inability to see any activity associated with this protein and the co-elution with the GroEL, which could be related to the inactivity. If the hypothesis is that GroEL is still potentially associated with the thiolase, than this complexation might be preventing the thiolase from functioning in a normal capacity.

A final effort to improve the soluble expression of the *Arthrobacter* thiolase protein, involved looking at the codon usage in *E. coli* and simultaneously looking at whether the his-tag could be the reason for improper folding. Although the addition of the pRARE plasmid to improve the translation of rare codons in *E. coli* appeared to have promise initially (fig 4-7), no activity was observed by the DTNB assay.

4.5 Conclusions

The lack of activity with acyl-CoAs or with *Micrococcus* cell extracts coupled with poor expression and purity of the *Arthrobacter aurescens* thiolase resulted in a system that was unable to be studied. Other expression systems, such as in *Pseudomonas*, could be tried to counter the poor expression, but this could still not alleviate the problem of lack of activity. The question of substrate specificity was unanswered, and there was the possibility that acyl-ACPs and not acyl-CoAs were required for thiolase activity, as in fatty acid synthase type thiolases. The other possibility was that *Arthrobacter* thiolases did have tight substrate specificity such that the chains needed to be 15 carbons long or terminally branched, as described in

Chapter 3. Looking to other organisms known to produce long chain alkenes led to the study of *Stenotrophomonas* genes, which appeared to have much wider substrate specificity as was evident by the wide range of products produced *in vivo* (47).

CHAPTER 5

Expressing *Stenotrophomonas maltophilia* genes, *oleABCD*

5.1 Introduction

The gene cluster necessary for alkene formation had been identified as *oleA, oleB, oleC,* and *oleD* (12, 49, 108). From bioinformatic analysis (108), the protein superfamilies for each Ole protein were identified. The superfamilies were identified as follows (108): OleA-thiolase (condensing enzymes), OleB- α/β hydrolase, OleC- AMP dependent ligase/synthase or LuxE, OleD-short chain dehydrogenase/ reductase.

The thiolase superfamily enzymes catalyze the condensation or thiolytic cleavage of ACP or CoA activated substrates. Thiolases are responsible for forming the carbon-carbon bond in fatty acid synthesis, polyketide synthesis, polyisoprenoid synthesis, and polyhydroxy butyrate synthesis (51). One well studied thiolase is from *Zoogloea* and condenses two acetyl-CoA molecules in the PHB biosynthesis pathway (33). The active site cysteine of this thiolase attacks the C-1 carbon of the acetyl-CoA, forming a covalently bound acyl group with the release of CoA. A proton is abstracted from the second acetyl-CoA to create a carbanion which attacks the covalently bound substrate and forms a new carbon-carbon bond (51). The HMG CoA Synthase is a

variation on this reaction in which the carbanion is formed on the tethered substrate and then attacks the second substrate (51). Homology of this superfamily is relatively low, but the catalytic triad of the active site is completely conserved (51).

The α/β hydrolase superfamily enzymes share the same fold, but large insertions enable the accommodation of vastly different substrates and members act in unrelated reactions and pathways (55). The catalytic triad of this family includes a nucleophile (Ser, Asp, or Cys), a general acid, and a histidine to activate water (108). Generally a nucleophilic attack occurs on the substrate, followed by an ester hydrolysis (18). Common enzymes in this family include lipases, esterases, and haloalkane dehalogenases (18). OleB in the α/β hydrolase superfamily is found to not be required for alkene biosynthesis, but has been observed to increase alkene formation yield (48).

AMP dependent ligase/synthase or LuxE superfamily contains enzymes that utilize ATP to activate an acyl compound for the ligation of CoA to the carboxyl. Many members of this family, including acetyl-CoA synthetase, malonyl-CoA synthetase, acyl-CoA synthetase, propionyl-CoA synthetase, and 4-chlorobenzoate:CoA ligase (102, 121), catalyze reactions in an ordered bi uni uni bi ping pong mechanism in which the acyl substrate and the AMP bind, releasing the PO_4 after reaction and then the Coenzyme A is bound and ligated to the acyl group with the release of AMP and product (102). The mobility of the N and C terminal domains is thought to be the key to accommodating the reactions, allowing very different conformations that enable the enzyme to catalyze the two steps of the reaction (102, 121).

The short chain dehydrogenase/reductase superfamily covers a wide range of diverse enzyme reactions. Enzyme families in this superfamily include oxidoreductases, lyases, and isomerases (83). Enzymes of this family are known to act on steroids, prostaglandins, aliphatic alcohols, and xenobiotics (83). The conserved N terminal motif T-G-X_3-G-X-G is essential for the nucleotide cofactor, NAD^+ or $NADP^+$ (83). The catalytic residues were found to be Ser-Tyr-Lys, in which the Tyr is the catalytic base, Ser stabilizes the substrate, and Lys interacts with the cofactor to reduce the pKa of the Tyr (45, 83).

Given the new developments in the pathway and the difficulty of expressing pure and active OleA protein from *Arthrobacter* another investigative approach is described here. It was already known that *Stenotrophomonas maltophilia* produces a wide variety of hydrocarbons and this was also verified upon receipt of the strain (Fig 3-1A). Because there was concern that the *Arthrobacter aurescens* TC1 might not have a wide enough substrate range to exhibit activity with the substrates that were used, experimenting with *S. maltophilia* would seem to alleviate that concern. *S. maltophilia*, based on culture extracts, appears to accept anything from 14 carbon chain length to 16 carbon chain length for condensation. It could also be determined from the sequenced genome that unlike *Arthrobacter, S. maltophilia* has separate *oleA, oleB, oleC, oleD* genes, rather than the fused *oleBC* gene. Transcribing the genes separately possibly allows for better protein expression of the individual proteins and therefore, ease of *in vitro* studies.

5.2 Methods

5.2.1 Cloning

Synthesized genes (*oleA, oleB, oleC,* and *oleD*) were ordered from GenScript Corporation (Piscataway, New Jersey, USA) and arrived in pUC like vectors (Appendix II) (49). The synthesized genes had added restriction sites, NdeI and HindIII, for cloning into pET30B+ (Novagen, Madison, WI), as described in (46). Removal of the stop codon from the genes allowed for a C terminal his tag to be added to the protein. Plasmids were transformed into BL21 (DE3) pLysE One Shot cells (Invitrogen).

5.2.2 Expression studies

Cultures of 50 ml LB with 50 ug/L kanamycin and 34 ug/L of chloramphenicol were inoculated from a starter culture and grown to an OD_{600} of between 0.5 and 1.0, at which time they were induced with isopropyl-β-D-1-thiogalactopyranoside (IPTG) at 0.1 mM. Cultures continued to shake at 37°C for 4 hours and were then harvested to assess protein expression. Cultures were centrifuged at 3000 × g for 15 min and then pellets resuspended in 20 mM sodium phosphate, 500 mM NaCl, pH 7.4. Cells were sonicated at 40% power setting once for 30 sec and three times for 20 sec with approximately 60 sec on ice between sonication periods. Lysate was centrifuged for 60 min at 14,000 × g. Soluble and insoluble fractions were analyzed by SDS-PAGE on a 10% polyacrylamide gel. Soluble extracts were also subjected to the Qiagen kit,

Nickel Affinity Spin Column, to assess a quick affinity purification of all four proteins with their appended C-terminal poly histidine tails.

5.2.3 Purification of OleA, B, C, and D

Cultures were grown in 500 ml LB with 50 µg/ml kanamycin and 34 µg/ml chloramphenicol. Cultures were inoculated with 10 ml overnight cultures and grown at 37°C to an OD_{600} of 0.5 – 1.0. Cells were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 mM when the OD_{600} of the culture reached 0.5 and 1.0. After induction (temperature and time varied as specified in subsequent sections), the induced cells were harvested by centrifugation at 3000 ×*g* for 25 min and resuspended in 20 mM sodium phosphate, 500 mM NaCl, pH 7.4 buffer. The cells were disrupted by three passes through a chilled French pressure cell at 1200 psi and centrifuged at 27,000 × *g* for 90 min to remove cell debris and insoluble protein. The soluble fraction was filtered through a 0.45 µm membrane or centrifuged for an additional 30 min at 27,000 × *g* prior to loading onto a Pharmacia Biotech LCC 501 FPLC fitted with a 5 ml HisTrap HP (Amersham Biosciences) column complexed with Ni^{2+} and equilibrated with 20 mM sodium phosphate, 500 mM NaCl, pH 7.4 buffer. Purification was assessed by SDS-PAGE, using Simply Blue Safestain (Invitrogen).

5.2.4 Improving expression of OleA

Due to the absence of a detectable OleA band by SDS-PAGE analysis, additional expression work was done to increase expression of OleA. Expression studies were carried out for 24 hours at 16°C, 30°C, and increased IPTG to 0.3 mM for 4 hr induction at 37°C in order to increase detectable expressed protein.

The penultimate amino acid was also mutated from leucine to alanine due to the N-end rule described in the pET expression manual (Novagen), which states that protein half-lives decrease to only 2 min when certain amino acids are in the penultimate position (Arg, Lys, Phe, Leu, Trp, or Tyr) (81). The Strataclone Quick Change Protocol was used to mutate the residue. Primers used were 5'-CATATGGC GTTCAAAAATGTATCTATCGCTCCT-3' and 5'-GAGCCAGACCAGCGATAGAT ACATTTTTGAACG-3'. Plasmids were also transformed into BL21 (DE3) One Shot (Invitrogen) and expression studied in case the pLysE strain was suppressing expression too much.

An expression study was performed on the L2A mutant in BL21 (DE3) and BL21 (DE3) pLysE at 4 hr and 24 hr 37°C induction with 0.1 mM IPTG. A band was potentially detected in the BL21 (DE3) cells in the 24 hr induction samples. A protocol from Jennifer Seffernick for the inoculation of the cultures was employed to potentially help increase expression. The protocol was as follows for the starter culture preparation: added one colony to 3 -5 ml media, grew to an OD_{600} of 0.5-1 and moved to 4°C overnight. The next day, the cells were centrifuged and resuspended in new medium that included kanamycin and chloramphenicol. The resuspension was used as start-culture inoculum for expression studies.

Using the new plasmid, cells, and inoculation method, time and temperature studies were performed, but it became increasingly difficult to detect the band of interest. In order to determine if the potential band was indeed the thiolase, the BL21 (DE3) L2A mutant was purified. Four shake flasks with 500 ml LB media and 50 µg/ml kanamycin were induced for 6 hr with 0.2 mM IPTG at 37°C. This was purified using the protocol described in the previous section with 50 mM imidazole step increases.

5.2.5 Improving soluble expression of OleB

Expression of OleB was not optimal in initial studies, so various attempts to improve expression were evaluated. Temperature studies on OleB were done in BL21 pLysE. Inductions of 24 hr at 30 and 16°C was compared to the 4 hr induction at 37°C that was initially used for evaluation.

The codon usage was also assessed using Graphical Codon Usage Analyzer (GCUA) 2.0 and three amino acids were identified as rare tRNAs in *E. coli*: CGG (2) and ATA. Rosetta 2 BL21 (DE3) cells were purchased (EMD Chemicals) and the cells were transformed with *oleB* in pET30b. The protocol for culture inoculation of hard to express proteins (previous section) was also employed at this point.

Rosetta cells were evaluated to find optimal induction time and temperature conditions. Time and temperature studies of 0-24 hr 0.2 mM IPTG inductions at 16 and 37°C were evaluated for soluble and insoluble expression. OleB from Rosetta2

BL21 was purified on the FPLC after induction of 0.2 mM IPTG at 16°C for 4 hr. Purification was also performed after 0.1 mM IPTG induction.

5.2.6 Improving soluble expression of OleD

OleD expression was optimized to increase the soluble to insoluble ratio of expression. Induction conditions were first investigated, with 0.1 mM IPTG induction at 16 and 30°C and 0.05 mM IPTG induction at 37°C. Expression at 16°C was investigated further with 0.05 and 0.1 mM IPTG induction, sampling at 0, 3, 5, and 24 hrs. Expression at 30 and 37°C was investigated further with 0.05 and 0.1 mM IPTG induction, sampling at 0, 2.5, 3.5, and 4.5 hrs. Purified OleD from BL21 pLysE induced for 4 hrs at 37°C with 0.05 mM IPTG.

Another approach that was investigated was to co-express chaperone proteins with OleD. GroEL and GroES are 2 chaperones on the PAG plasmid with chloramphenicol resistance. The pET30b with *oleD* and the PAG plasmid were transformed into BL21 (DE3). An initial study was done with 0.005-0.5% arabinose induction of the chaperones and 0.01-0.1 mM IPTG induction of OleD. The cultures were inoculated with overnight cultures and the flasks were grown at 37°C until an OD_{600} of 0.5 and then cultures were transferred to 15°C. The cultures shook at 15°C for 30 min to equilibrate to temperature and then arabinose was added to induce expression of the chaperones. Cultures continued to shake for 90 min and then IPTG was added. After 24 hr cultures were harvested as previously described. Once an

optimal arabinose and IPTG concentration combination was identified in 50 ml cultures, the cultures were scaled up to 500 ml and then purified by FPLC as previously described in 20 mM sodium phosphate, 500 mM NaCL, pH 7.4 with EDTA free protease inhibitor tablets (Roche) added.

5.3 Results and Discussion

5.3.1 Expression studies with spin columns

The four *S. maltophilia* proteins were analyzedby SDS-PAGE after the cell extracts had been chromatographed onQiagen Nickel affinity spin columns (Fig 5-1). No band could be identified in the OleA fractions that was present in the A1 (induced) and not in the A0 (uninduced). OleB appeared to have a band of low expression present in the B1 (induced) elution that is not in the B0 (uninduced) elution. OleC had fairly high expression in the elutions, and did not appear to need much optimization. OleD D1 (induced) elutions appeared to have a potential band that is not in the D0 (uninduced) elutions. From these initial studies it can be seen that significant optimization on expression needed to be done for OleA, OleB, and OleD.

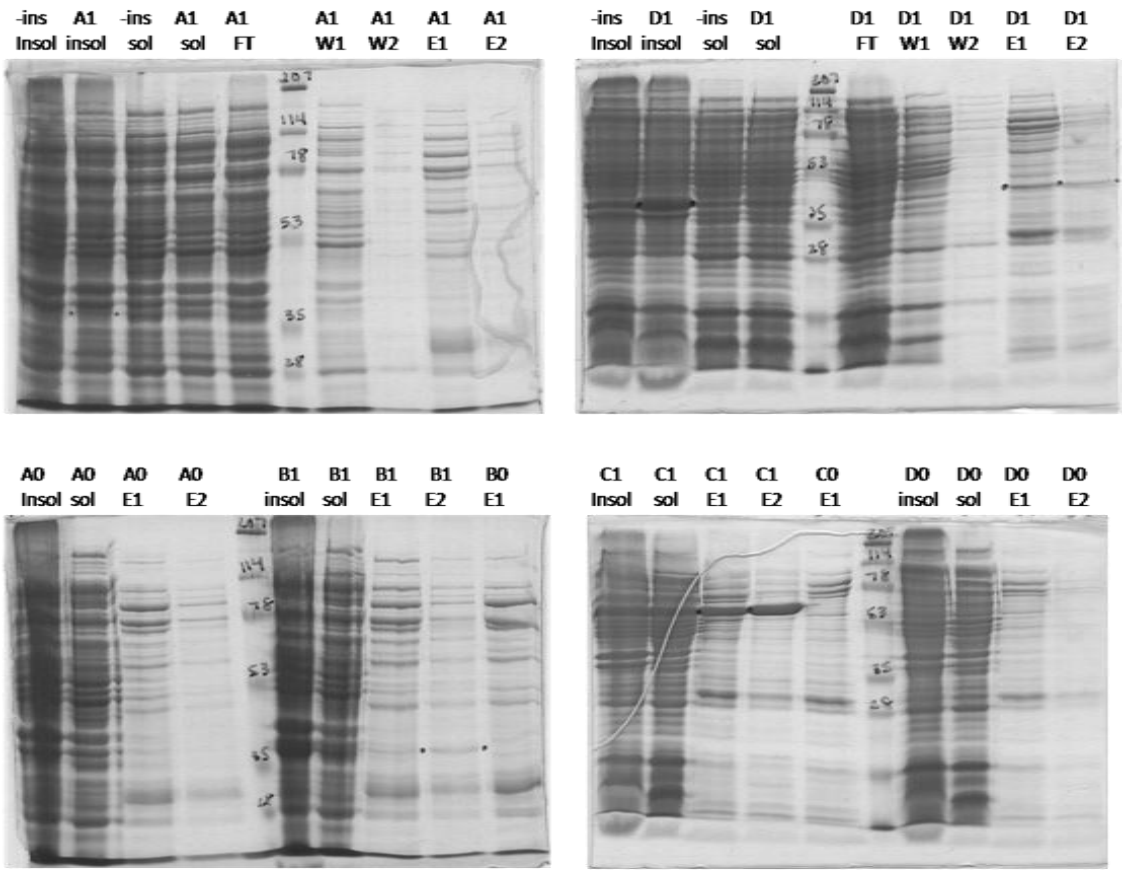
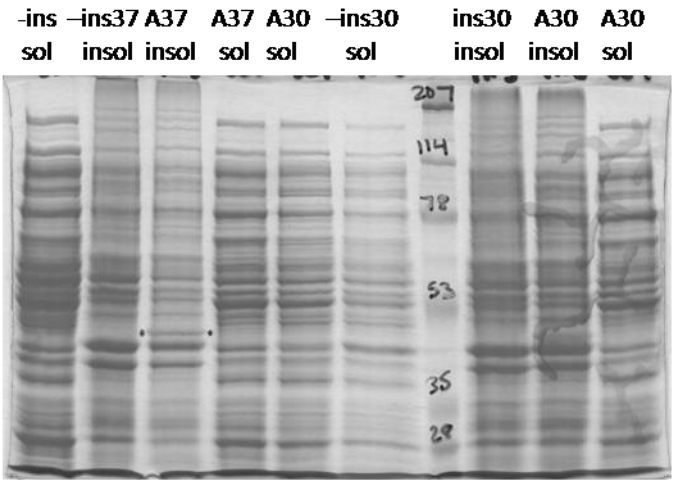
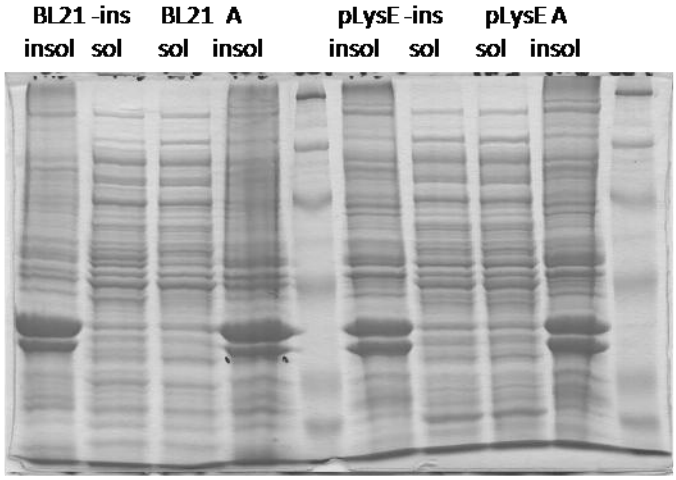


Figure 5-1. Nickel affinity spin column results. OleA, B, C, and D results as labeled above. X0 vs X1 indicates uninduced (X0) and induced (X1). Insoluble (insol) and soluble (sol) crude extracts are labeled accordingly. E1 = 1st elution, E2 = 2nd elution, W1 = 1st wash, W2 = 2nd wash. X=A, B, C, D



A) BL21 (DE3) pLysE OleA (*S. maltophilia*)



B) BL21 (DE3) OleA L2A (*S. maltophilia*)

Figure 5-2. Expression study of OleA strains.

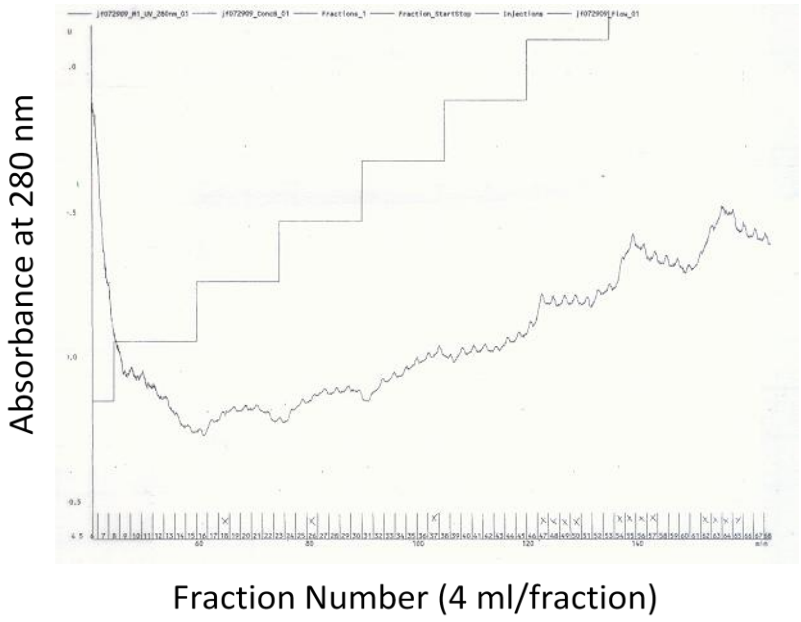


Figure 5-3. FPLC Chromatogram of BL21 (DE3) OleA purification. Protein concentration assay on representative fractions per step verified no significant protein.

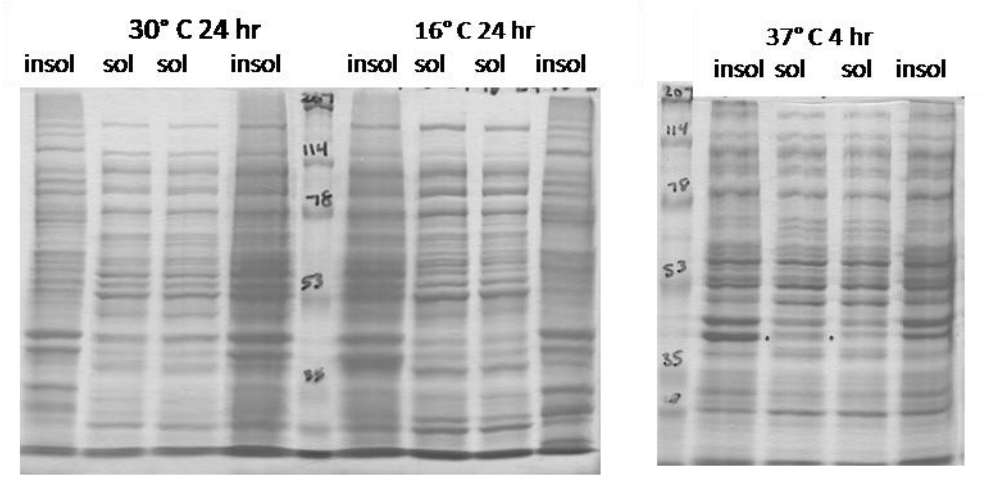
5.3.2 Purification of OleA

A very low level of expression was detected using the originally synthesized OleA at 37°C with 0.3 mM IPTG in the insoluble fraction only (Figure 5-2, A). In an attempt to increase the soluble expression, further studies were done using the L2A mutant. After confirming the mutation of the penultimate amino acid, an expression study was performed in 50 ml flasks with 0.1 mM IPTG and induced for 4 and 24 hr at 37°C. A potential band was identified in the insoluble fraction of the BL21(DE3) cells induced for 24 hr (Figure 5-2, B).

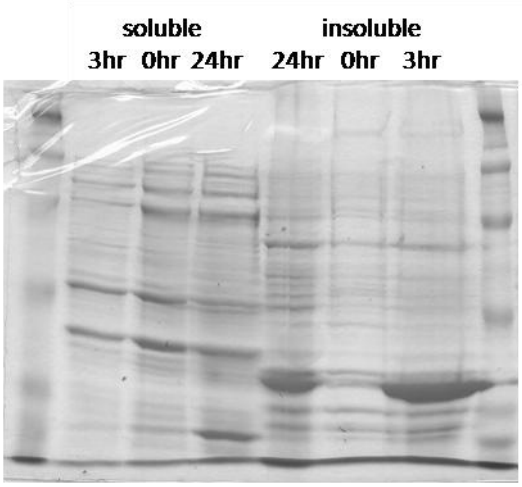
Purification results were incomplete, but still informative. The chromatogram and Biorad protein assay on all increasing imidazole steps of the purification resulted in no detectable protein (Fig. 5-3). No gel was run, therefore it was difficult to determine if the levels were just below detection or if there was no expression. In the absence of improved OleA expression that resulted in no detectable soluble protein, this meant OleA could not be studied *in vitro*.

5.3.3 OleB

Temperature studies were done at 30°C and 16°C for 24 hrs to compare to the initial study done at 37°C for 4 hrs. The initial 37°C induction didn't yield OleB overexpression. At 30°C and 16°C it was still difficult to see any highly expressed band in crude extracts (Figure 5-4, A). After analyzing the codon usage of



A) BL21 (DE3) pLysEOleB (*S. maltiphilia*)



B) BL21 (DE3) Rosetta2 OleB (*S. maltphilia*)

Figure 5-4. Expression studies of OleB before (A) and after (B) Rosetta2 plasmid.

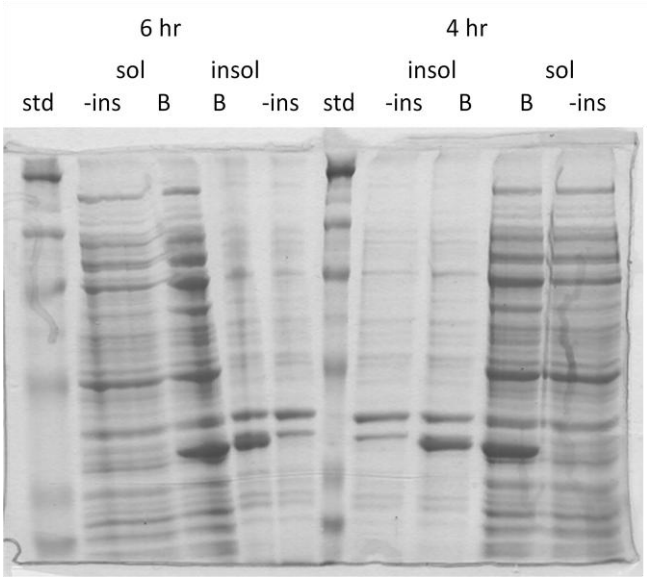
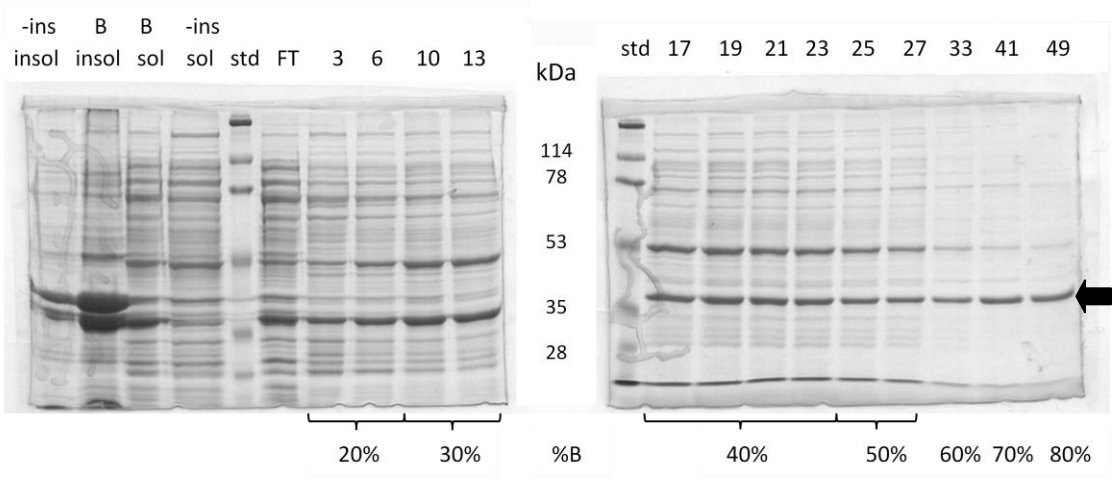


Figure 5-5. OleB expression time study at 16°C.

A) 16°C 4 hr, 0.2 mM IPTG



B) 16°C 4 hr, 0.1 mM IPTG

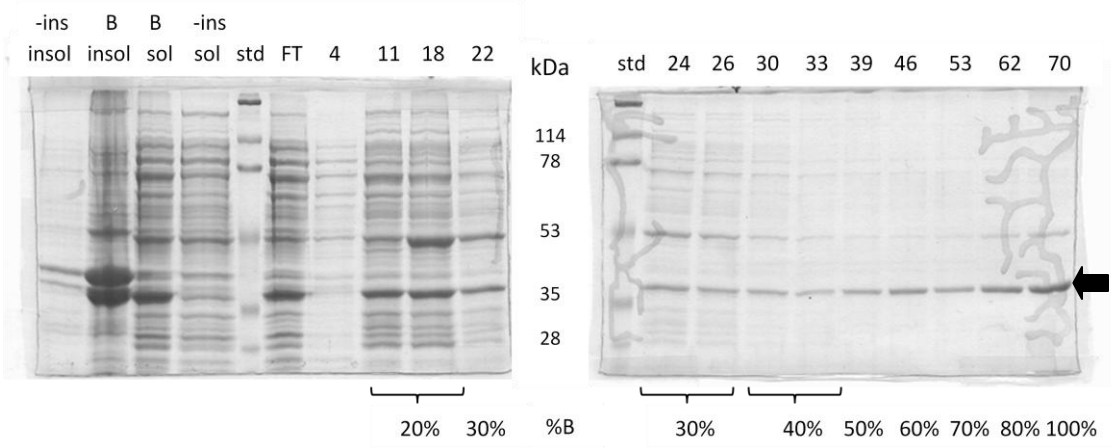


Figure 5-6. OleB purification results. OleB is highlighted with an arrow on the right.

the *oleB* gene using GCUA 2.0, three codons were found that are rare tRNAs in *E. coli*. OleB was then expressed with the BL21 (DE3) Rosetta2 plasmid and a time study completed at 37°C in 50 ml shake flasks. A clear overexpressed protein band could now be seen. Due to the fact that the overexpressed band is largely in the insoluble fraction, more optimization is required.

It became evident, that in this system slow expression was preferable. Expression of OleB was next evaluated at an induction temperature of 16°C and 0.1 mM IPTG, for 4 and 6 hrs. It was difficult to interpret which condition, 4 or 6 hrs was preferable, and so the more conservative induction of 4 hrs was chosen (Figure 5-5). The purification of OleB was next performed after a 4 hr induction at 16°C with 0.2 mM IPTG, Figure 5-6 A. The purification was not ideal, as there were contaminating proteins throughout all the fractions, eluting up to 500 mM imidazole. Significant protein was being lost in the insoluble fraction. The next purification was induced at a reduced concentration of IPTG of 0.1 mM. The results only differed in the fact that all the protein in each fraction was now reduced. Many contaminating proteins still remained.

5.3.4 OleD

The temperature study in Figure 5-7 showed high expression in the insoluble fraction after the 30°C, 6 hr induction and the 37°C, 4 hr induction. The 16°C, 6 hr induction results were difficult to interpret because there was another band that ran at

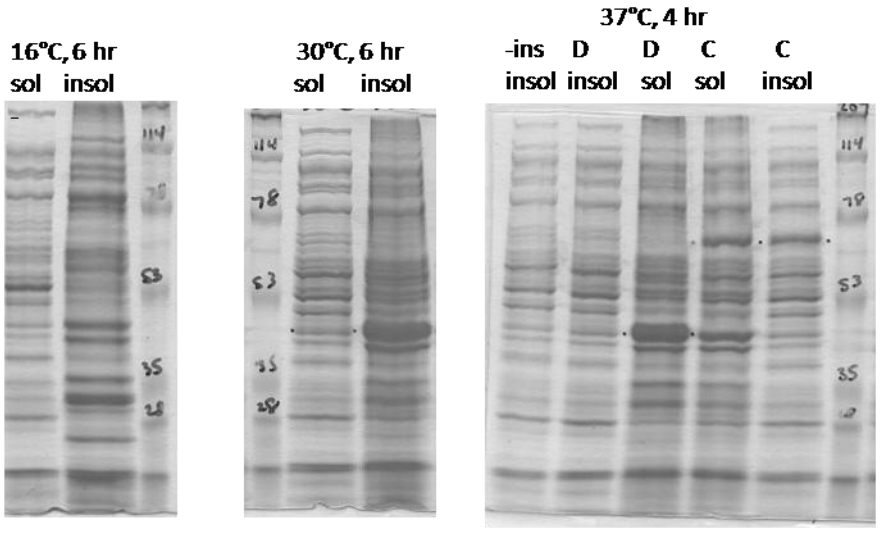
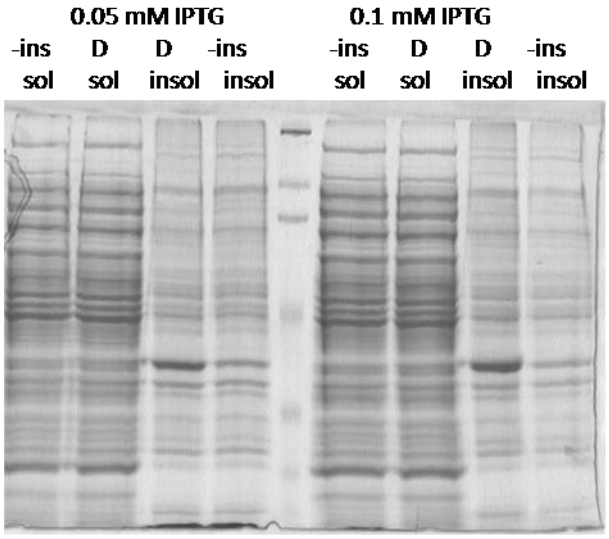


Figure 5-7. Temperature expression studies with OleD

A) 30°C, 4.5 hr induction



B) 37°C, 4.5 hr induction

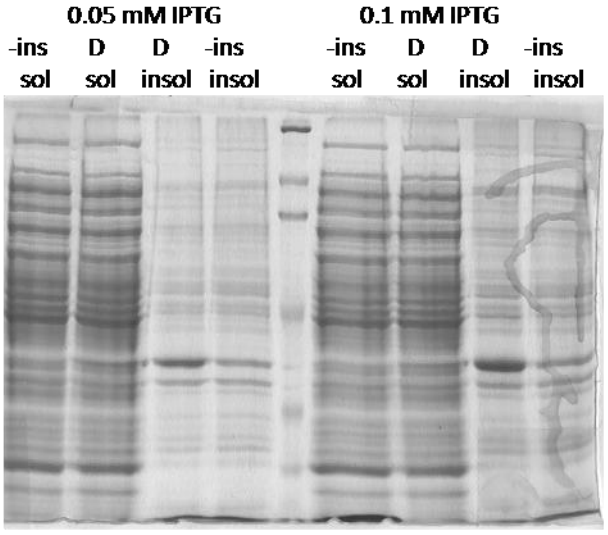


Figure 5-8. Induction study to increase OleD soluble expression

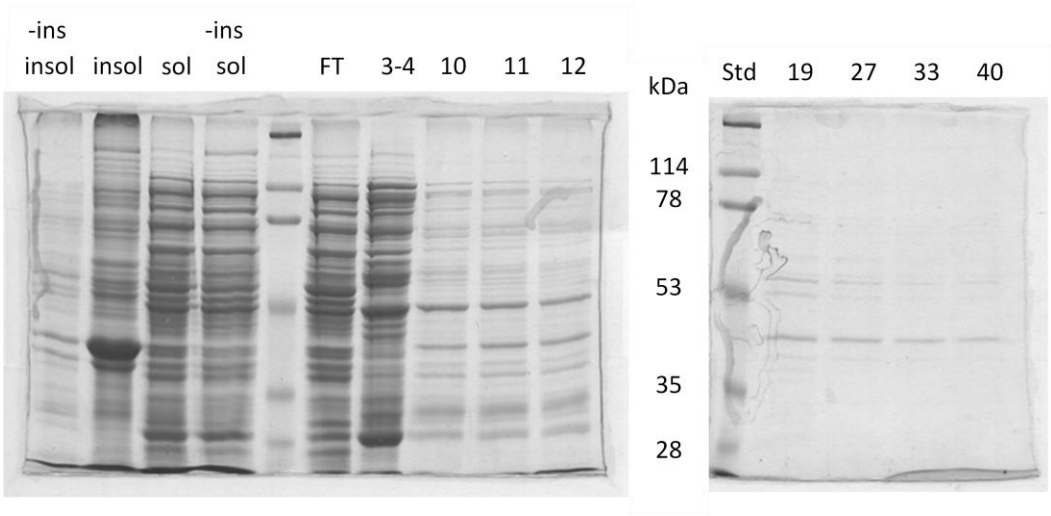


Figure 5-9. Purification of OleD

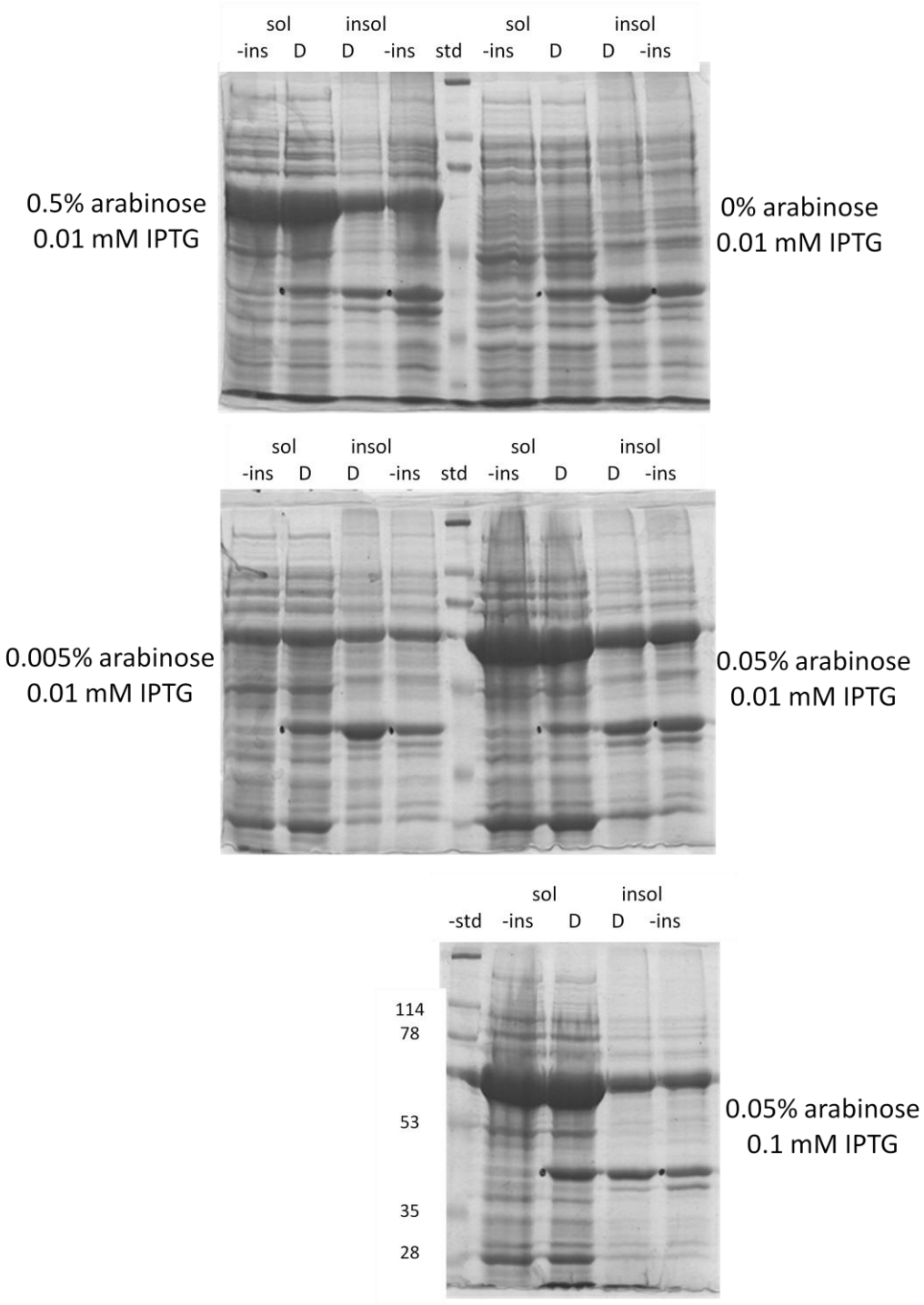


Figure 5-10. OleD soluble vs. insoluble fractions with chaperones

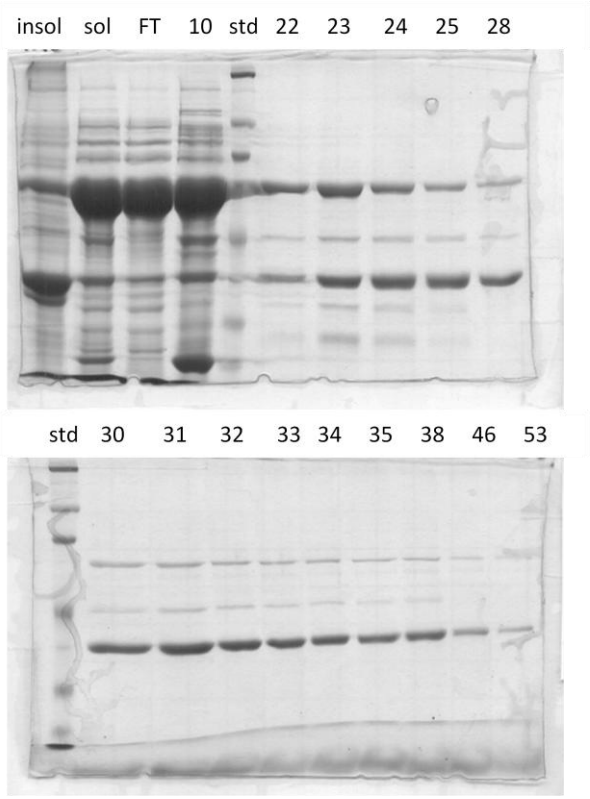


Figure 5-11. OleD purification with chaperones

the same size. A follow up study at 16°C was performed and no overexpression was detectable by SDS-PAGE. A follow up time and IPTG study was performed at 30 and 37°C, in which no OleD overexpression could be detected in the soluble fraction, and results looked very similar whether the induction was 2.5, 3.5, or 4.5 hr (4.5 hr data shown in figure 5-8). Results of the purification of OleD on the nickel column, after induction of 4 hr at 37°C showed low levels of OleD in all fractions throughout the increasing imidazole profile (Figure 5-9). The fractions also contained large amounts of contaminating proteins, leaving OleD as only a small percentage of the total protein in the fractions based on the SDS-PAGE analysis.

The chaperone experiment resulted in higher soluble OleD expression than had been detected in previous studies. The combination of 0.05% arabinose and 0.1 mM IPTG resulted in more OleD detected in the soluble crude fraction than in the insoluble crude fraction (Fig 5-10). The 0.5 and 0.05% arabinose both appeared to have a higher ratio of soluble to insoluble OleD than the 0.005% arabinose. The 0.5% arabinose seemed to have such a high level of expression of GroEL and GroES that the initial purification was done with 0.05% arabinose. IPTG levels of 0.1 mM and 0.01 mM were tested with the range of arabinose, but the 0.1 mM level IPTG seemed to have significantly more soluble expression than the 0.01 mM so that was carried forward to the purification. The purification results, Figure 5-11, showed significantly higher levels of OleD than the previous purification, Figure 5-9. Significant levels of chaperone proteins were co-eluting with the OleD, which would raise the question

again as to whether the chaperones were still associated with the protein in the crude extract.

5.4 Conclusions

Stenotrophomonas maltophilia oleA, *oleB*, and *oleD* genes did not appear to express well in soluble form to assay them in detail. OleC was purified to homogeneity and studied further (Chapter 7). Further studies of OleA, B, and D will require returning to the toolbox of organisms to obtain expressed protein in a soluble and active form.

CHAPTER 6

Purification and crystallization of OleA from *Xanthamonas campestris* and demonstration of a non-decarboxylative Claisen condensation

6.1 Introduction

Commodity hydrocarbons derive from petroleum, but nature provides a rich source of hydrocarbons for which biosynthetic pathways are being elucidated. Isoprenoid biosynthesis has been well-studied (79) and an enzymatic decarbonylation of fatty aldehydes to produce alkanes has recently been demonstrated for cyanobacteria (96). It has been known for more than forty years that some bacteria biosynthesize long (C_{23}-C_{33}) hydrocarbon chains containing a double-bond at the median carbon via a mechanism known as a "head-to-head" condensation of fatty acyl groups (2-5, 113). For example, bacteria from the genus *Arthrobacter* produce largely C_{15} fatty acids (116) and make predominantly C_{29} olefins (47). These observations are consistent with studies in 1969 showing the loss of the ^{14}C label at carbon-1 of one of the acyl groups

undergoing head-to-head condensation (2). These early *in vitro* studies were conducted with crude cell protein extracts. It was not until 2010 that the genes involved in the head-to-head biosynthetic pathway were described in the peer-reviewed literature (12, 109), providing new insights into the biosynthetic pathway based on a bioinformatics analysis of the gene and protein families.

OleA is homologous to proteins in the thiolase or condensing enzyme superfamily (12, 108). This is a very large superfamily of over 13,000 known proteins. The known thiolase superfamily proteins typically catalyze condensation reactions between acyl-thioester substrates, either with or without the loss of a carboxyl group. Approximately seventy bacteria are known to contain genes denoted as *oleABCD* and those tested produce long-chain olefinic hydrocarbons (108). The precise role of each *ole* gene product in the biosynthesis remains to be defined. When the *oleC* gene is deleted, or only the *oleA* gene is present *in vivo*, a long-chain ketone(s) is observed. These data supported the idea that OleA is involved in the initial stages of the head-to-head hydrocarbon biosynthetic reactions (12, 49, 108).

There are two alternative proposals in the literature regarding the OleA condensation reaction (Figure 6-1). Beller, et al (Fig. 6-1A) proposed that OleA catalyzes a decarboxylative condensation between a β-ketoacyl-CoA and a fatty acyl-CoA (12). Sukovich, et al (Fig. 6-1B) have proposed that OleA catalyzes a non-decarboxylative Claisen condensation between two fatty acyl-CoA substrates (108). These two types of condensation reactions are difficult to differentiate *in vivo* where both fatty acyl-CoAs and β-ketoacyl-CoAs may be present simultaneously and many

enzymes are present. The study by Beller and coworkers used a purified OleA enzyme, but their demonstration of activity required the addition of a crude soluble protein extract from *Escherichia coli* (12). The proposed β-ketoacyl-CoA substrate was suggested to have been generated from the corresponding acyl-CoA by the proteins present in the *E. coli* soluble fraction. A clear differentiation between OleA reaction A and B could be obtained using a purified OleA preparation in admixture with defined substrates *in vitro*. The two types of condensation reactions could also be differentiated by determining the reaction product. OleA reaction A produces a 1,3-diketone while OleA reaction B yields a β-ketoacid.

There are other important questions that can be answered directly using a purified OleA protein and purified single substrates. These include determining the substrate specificity of OleA with respect to chain length, determining the complete reaction stoichiometry, determining what drives the apparent Claisen condensation to completion and revealing why cloning *oleA* genes in heterologous hosts produces monoketones. These issues are addressed in the present manuscript.

The OleA protein from *Xanthomonas campestris* was cloned, overexpressed in *E. coli*, and purified to homogeneity. The putative product of the reaction was synthesized chemically to allow comparison with the biochemical product. OleA was shown to react with myristoyl-CoA to produce the corresponding β-ketoacid via a non-decarboxylative Claisen condensation reaction. This intermediate was shown to react, in the presence of OleC and OleD, to yield a long chain olefin. In the absence of OleC

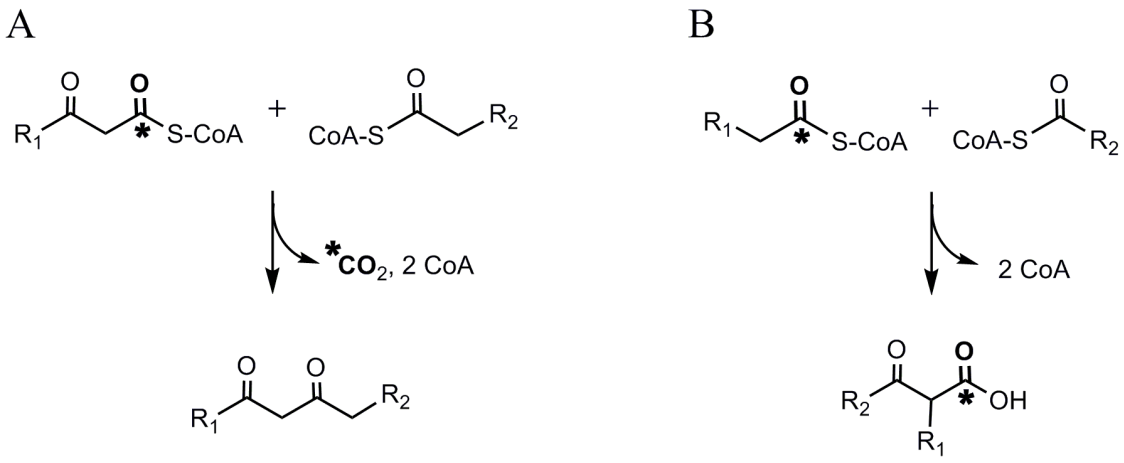


Figure 6-1. Fundamentally different condensation mechanisms have been proposed for OleA:

(A) decarboxylative condensation between a β-keto ester and an acyl thioester (Beller, et al (12)), or
(B) non-decarboxylative condensation between two acyl thioesters.

and OleD, the product of the OleA reaction was shown to undergo spontaneous chemical decarboxylation to yield a ketone. This explains previous *in vivo* observations of ketone formation with the expression of an *oleA* gene in a heterologous host (12, 108).

6.2 Experimental Procedures

6.2.1 *Chemical Synthesis and analysis*

β-Ketocarboxylic acid syntheses have been previously reported (37), but the literature does not describe the synthesis of higher benzyl esters of β-ketocarboxylic acids derived from fatty acids. The detailed procedure used for the synthesis of 2-myristoylmyristic acid (2MMA, 2-dodecyl-3-ketohexadecanoic acid) is described in the Supplemental Data.

In brief, the forced Claisen condensation method described by Briese and McElvain (1, 17) was adapted to the coupling of benzyl myristate, using a half equivalent of sodium benzyl alcoholate in benzyl alcohol as the basic promoter, removing benzyl alcohol by heating under vacuum to force the condensation nearly to completion. Benzyl 2-myristoyl-myristate (B2MM, benzyl 2-dodecyl-3-ketohexadecanoate) so prepared (70% yield and approximately 90% purity after bulb-to-bulb distillation) crystallized, mp 34.5-35.8°C. The purity of B2MM was estimated from NMR data in $CDCl_3$ solution, which show no conclusive evidence for tautomeric enolic forms being present.

Carefully monitored hydrogenolysis of B2MM (Pd/C catalyst, methyl *t*-butyl ether solvent) isolated by filtration and cooling of the filtrate to -80°C, produced a 5:2 (mol/mol) mixture of 2MMA and its decarboxylated product 14-heptacosanone as determined by GC-MS analysis after methylation of the acid (CH_2N_2). On cold storage (-80°C) the solution was enriched to an 8:1 2MMA and the derived ketone, apparently by preferential precipitation of the ketone.

6.2.2 Cloning and Expression of OleA

Synthetic *oleA* genes were designed based on *oleA* genes from *Congregibacter litoralis* KT71 (ZP_01103251.1), *Xanthomonas campestris* spv. campestris str. ATCC 33913 (NP_635607.1), *Xylella fastidiosa* 9a5c (NP_299252.1), *Plesiocystis pacifica* SIR-1(ZP_01906524.1), and γ-*proteobacterium* NOR5-3 (ZP_05127044.1), see supplementary figure S1, and purchased from DNA 2.0 (Menlo Park, CA. The genes were cut with NdeI and BamHI restriction enzymes and cloned into pET28b+ (Novagen, Madison, WI). All 5 genes were separately transformed into *E. coli* One Shot BL21 (DE3) (Invitrogen). All five recombinant strains were screened for soluble protein expression in 50 ml cultures induced for 4 h at 37°C. Two of the five constructs expressed soluble protein in *E. coli*, only *X. campestris* was found to be active *in vitro*, and that was selected for further study.

X. campestris for OleA purification was cultivated under two different conditions. Small-scale cultivations were conducted in 2 L flasks containing 500 ml LB

with 50 µg/ml kanamycin and induced at an OD_{600} of 0.7-0.85 with 0.1 M isopropyl-β-D-thiogalactopyranoside (IPTG). After 4 h, cells were harvested by centrifugation for 25 min at 3000 × g. Large-scale cell cultivation was conducted in the Biotechnology Resource Center, University of Minnesota. A 440 L culture was prepared in a 550 L DCI bioreactor (DCI-Biolafitte, St. Cloud, MN) using a Rhapsody digital controller system and induced with 0.5 mM IPTG. Cells were harvested, lyophilized, and then stored at -80°C.

6.2.3 Purification of OleA

Cells were resuspended in 20 mM sodium phosphate buffer, 500 mM NaCl pH 7.4 with EDTA-free protease inhibitor tablets (Roche). Cells were disrupted by 3 passes through a chilled French pressure cell at 1200 psi and centrifuged at 27,000 × *g* for 90 min to obtain the soluble protein fraction. The soluble fraction was centrifuged at 27,000 × *g* for 30 min to clear prior to loading onto a Pharmacia Biotech LCC 501 FPLC equipped with a 5 ml Ni(II)-loaded HisTrap HP column (Amersham Biosciences) equilibrated with 20 mM sodium phosphate, 500 mM NaCl pH 7.4 buffer. The OleA protein was eluted at 135 mM imidazole. Ten g wet weight of cell paste yielded 60 mg purified OleA. Fractions were analyzed by SDS-PAGE and Simply Blue Safestain (Invitrogen). Pooled fractions were concentrated, and imidazole removed, with 3 passes through a 50 ml pressure concentrator (Amicon) using a 10,000 MWCO membrane (Millipore). Alternatively, after concentration of fractions, the

OleA protein was dialyzed 3 times at 4°C to remove the imidazole. Protein concentrated up to 30 mg/ml remained soluble and active.

6.2.4 Identification and Expression of Active OleD

Using the NCBI Blast algorithm (9), *oleD* genes were identified. Sequences from *Chloroflexus auranticus* (Caur_3530), γ-*proteobacterium* NOR5-3 (ZP_05127041.1), *Xylella fastidiosa* Temecula1 (NP_779252.1), *Xanthomonas campestris* spv. campestris str. ATCC 33913 (NP_635614.1) were optimized for expression in *E. coli* (see supplementary figure S2) and cloned into pJexpress expression vectors with a T7 promoter by DNA 2.0 (Menlo Park, CA). Vectors were transformed into *E. coli* One Shot BL21 (DE3) (Invitrogen). Proteins were screened for activity and expression using 50 ml LB cultures with 50 µg/ml kanamycin. Cells were induced with 0.1 mM IPTG at an OD_{600} of 0.55-0.75. Soluble cell extracts were combined with OleA, OleC, and cofactors to test for the production of alkenes using the GC-MS enzyme assay described below. The OleD protein originating from *X. campestris* was the only protein found to support alkene biosynthesis. Cultures were scaled up using 2 L flasks containing 500 ml LB with 50 µg/ml kanamycin. Cultures were grown at 37°C with agitation at 225 rpm. The culture was induced at an OD_{600} of 0.7-0.8 with 0.1 – 0.4 mM IPTG and grown at 30°C shaking at 225 rpm for 20 h. Cells were harvested by centrifugation for 25 min at 3000 × *g*.

6.2.5 Purification of OleC, OleD, and Assay of OleD

The cloning, expression, and purification of OleC was previously described (46). For purification of OleD, cell pellets were resuspended in 20 mM sodium phosphate, 500 mM NaCl, pH 7.4 with EDTA-free protease inhibitor tablets (Roche) and passed through a chilled French pressure cell three times at 1200 psi. The cell lysate was centrifuged at 27,000 × *g* for 90 min and the soluble fraction was centrifuged for an additional 30 min. The soluble fraction was passed though a 0.20 µm syringe filter prior to chromatography. A 5 ml Ni(II)-loaded HisTrap HP column (Amersham Biosciences) equilibrated with 20 mM sodium phosphate, 500 mM NaCl pH 7.4 buffer was used for purification. Alternatively, 50 mM MOPS, 1% Tween 20, pH 7.0 was used for purification to improve solubility. Fractions were analyzed for purity by SDS-PAGE and Simply Blue Safestain (Invitrogen). Fractions eluting at 450 and 500 mM imidazole were pooled. The pooled protein was concentrated using a 50 ml pressure concentrator (Amicon) with a 10,000 MWCO membrane (Millipore) and dialyzed three times at 4°C. After dialysis, protein was centrifuged at 14,000 × *g* for 15 min at 4°C to remove precipitated protein.

OleD was previously suggested to be a ketone reductase (108). It was shown here to be active in a 250 µl reaction mixture consisting of 100 mM Tris, pH 7.4 containing OleA, OleD, OleC, 8 mM $MgCl_2$, 80 µM ATP, 260 µM myristoyl-CoA and 120 µM NADPH. NADPH oxidation was followed spectrophotometrically at 340nm.

6.2.6 Detecting the Release of CoASH Thiol for Assay of OleA Substrate Range

Release of the free thiol group of CoASH was detected by the addition of 5,5'-dithio-bis-(2-nitrobenzoic acid) (DTNB) measured spectrophotometrically at 412 nm (ε_{412}=13,600 $M^{-1}cm^{-1}$) (8, 42). Acyl-CoA substrates, purchased from Sigma Aldrich (Milwaukee, WI), were reacted with OleA protein in 100 mM Tris pH 7.4 and incubated at room temperature for 5 min in either 1 ml or 250 µl. DTNB was incubated with the reaction for 2 min and quantified spectrophotometrically.

6.2.7 Hydrocarbon Detection Enzyme Assay

A glass vial containing 250 µl total volume of 100 mM Tris pH 7.4 with 200-600 µg OleA, 5-25 µg OleD, 66 µg OleC, 1.4 mM NADPH, 8 mM $MgCl_2$, 3 mM ATP, and 1.2 mM myristoyl-CoA or an excess of 14-heptacosanone were incubated overnight at 30°C with gentle shaking. Products were extracted with 250 µl ethyl acetate using 16-hentriacontanone ketone (Tokyo Kasei Kogyo Co., Ltd., Japan) as an internal standard. After vortexing and 5 min of gentle centrifugation, the top solvent layer was transferred to a glass vial and analyzed using a gas chromatograph equipped with a flame ionization detector HP 7890A (Hewlett Packard, Palo Alto) and mass spectrometer HP 5975C (GC-MS-FID). GC was conducted under the following conditions: helium gas, 1.75 ml/min; HP-1ms column (100% dimethylsiloxane capillary; 30 m by 250 µm by 0.25 µm); temperature ramp, 100 to 320°C; 10°C/min, hold at 320°C for 5 min, 250°C injection port, and split at the outlet between MS and

FID. The mass spectrometer was run under the following conditions: electron impact at 70eV and 35 µA. The flame ionization detector was set at 250°C with hydrogen flow set at 30 ml/min, air set at 400 ml/min, and helium makeup gas set at 25 ml/min.

GC-MS was also used as described (47) for detecting ketones derived from spontaneous decarboxylation of the OleA β-keto acid products, using 200-600 µg OleA and 1 mM acyl-CoA substrates.

6.2.8 Radiolabeled Acyl-CoA Assay

Reactions of 200 µl included 0.2 µCi [1-^{14}C]-myristoyl-CoA, 40-60 mCi/mmol (American Radiolabelled Chemical, St. Louis, MO), 750 µM myristoyl-CoA and 1 mg OleA in 100 mM Tris pH 7.4. Samples were analyzed using high pressure liquid chromatography (HPLC) on a Shimadzu HPLC system equipped with a UV detector (Shimadzu, Columbia, MD) and a β-ram radioflow detector operated with the Laura 4 data acquisition/evaluation software (IN/US Systems, Tampa, FL).UV detection was set at 259 or 274 nm. Unfiltered samples of 50 or 100 µl volume were injected onto an analytical reverse phase Alltima HPC8 column with 5µm packing (Alltech 250 x 4.6 mm) and C8 guard column. The column was equilibrated in 50% 20 mM ammonium acetate pH 5.4 (A) and 50% 85:15 acetonitrile:methanol (B) and the following method adapted from (22). Linear gradients were as follows: 50% A:50% B 0-10 min, ramp to 70% B 10-15 min, 70% B 15-30 min, ramp to 100% B 30-35 min, 100% B 35-50 min, return to 50% A : 50% B 50-55 min and equilibrate 55-70 min. The flow rate was 1

ml/min. The scintillant (Monoflow X; National Diagnostics, Atlanta, GA) flow rate was 3 ml/min.

6.2.9 Mass Spectrometry Analysis

Mass spectrometry on enzymatically-produced and synthetic β-keto acid was performed using an LCQ-classic (Thermo Fisher Scientific) ion trap mass spectrometer with electrospray ionization mode (ESI). Samples were introduced by loop injection of 5 µl. Product ion spectra for 2-myristoylmyristic acid *m/z* 437 (M - H) was identified in negative ion mode, as well two additional ions *m/z* 393 and *m/z* 473/475. The fatty acid HPLC peak was analyzed by direct infusion into a Quantum Discovery Max (Thermo Finnigan) mass spectrometer operated in negative ion mode. ESI^--MS spectra for myristic acid was *m/z* 227 (M - H). Electron impact mass spectrometry in conjunction with GC (GC-MS) was performed to obtain the spectra of the β-keto acid derivatized with diazomethane. The solvent was evaporated with N_2, and then taken up in methyl-*t*-butyl-ether (mtbe) to run on GC-MS as previously described (47). The ketone molecular ion, *m/z* 394, was also identified by this method.

6.2.10 Analytical Gel Filtration

A Superdex 75 10/100 GL (Amersham Biosciences) size exclusion column was used on an AKTA (General Electric) FPLC with elution at 0.5 ml/min. The column was equilibrated with 20 mM sodium phosphate, 500 mM NaCl, pH 7.4. Molecular

weight standards (Biorad, Hercules, CA) were used with a range of 1,350 – 670,000 to create a standard curve. Three additional standards were used in a closer MW range to where OleA eluted, chymotrypsin (M_r = 25 kDa), albumin (M_r = 67 kDa) and conalbumin (M_r= 77 kDa).

6.3 Results

6.3.1 Cloning and Expression of oleA Genes, and Purification of OleA Protein

The *oleA* genes from *Congregibacter litoralis* KT71, *Xanthomonas campestris* spv. campestris str. ATCC 33913, *Xylella fastidiosa* 9a5c, *Plesiocystis pacifica* SIR-1,and γ-proteobacterium NOR5-3 were each cloned into *E. coli* and tested for the production of soluble OleA protein. The recombinant *E. coli* containing the *X. campestris* gene showed a high amount of soluble OleA protein as determined by SDS-PAGE and was found to be active in purified form, and that strain was therefore selected for further studies. *E. coli* cells expressing a His-tagged *X. campestris* OleA protein were grown in a 550 L bioreactor vessel, harvested, and lysed. Following chromatography on a Ni-column, the protein was shown to be homogenous as indicated by SDS-PAGE (Fig. 6-2B).

6.3.2 General Characteristics of OleA

The subunit molecular weight of the native OleA protein is 36,629 but, as engineered here with the His-tag, it is 38,792. The protein migrates somewhat higher than this on SDS-PAGE (Fig. 6-2B). The OleA protein subunit MW is near the middle of the range found in homologous proteins from the thiolase superfamily (Table 6-1). The *Mycobacterium* Pks13 protein is a large multidomain protein with the condensing enzyme domain defined as 44,122 in the annotation on the NCBI server. The native molecular weight of OleA was estimated to be 62,000 by gel filtration chromatography. This is suggestive of a subunit stoichiometry of two for the native enzyme. Hydroxymethylglutaryl-CoA (HMG-CoA) reductase, FabH and the *Mycobacterium* Pks13 are all dimers and the *Zoogloea* thiolase is a tetramer (27, 32, 33, 50).

The OleA protein shows a low sequence relatedness with homologous proteins in the thiolase superfamily (Table 6-1). With such divergence, it is not surprising that the cellular functions of the proteins are quite different. However, the general biochemical reaction catalyzed by all of the enzymes shown in Table 6-1 involves the condensation of acyl substrates. These condensation reactions occur by either a decarboxylative or non-decarboxylative (53) mechanism and the results described below are consistent with a non-decarboxylative mechanism for OleA. In both cases, thiolase superfamily proteins use a conserved active site cysteine and generate an acyl enzyme intermediate (51). OleA shares this conserved active site cysteine residue (Table 6-1). In the vicinity of the conserved cysteine, the amino acids in the OleA from *X. campestris* and *M.luteus* (12) are highly conserved (Table 6-1).

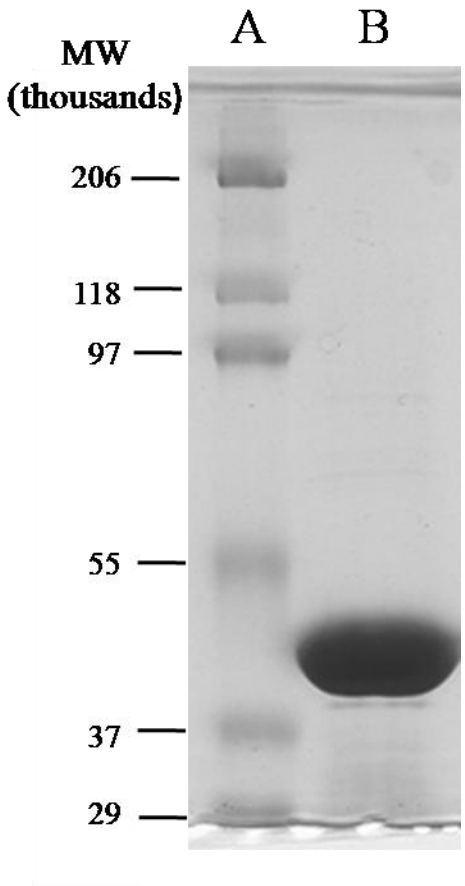


Figure 6-2. SDS-PAGE gel showing: (A) standard molecular weight markers, and (B) purified OleA.

Table 6-1: Properties of OleA compared with homologous proteins in the thiolase superfamily.

Protein (organism)	Accession #	% Seq ID*	Calc MW	Calc pI	Cellular Function	Claisen Mechanism	Sequence Signature[#]
OleA (*X. campestris*)	NP_635607	100	36,629	5.6	Alkene biosynthesis	Proposed Non-Decarboxylative	NA**C**LAFING
OleA (*M. luteus*)[12@]	YP_002957382.1	38	36,653	4.8	Alkene biosynthesis	Proposed Decarboxylative	NA**C**LGFVNG
Thiolase (*Z. ramigera*)[33]	AAA27706.1	19	40,416	5.9	PHB biosynthesis	Non-Decarboxylative	QL**C**GSGLRA
HMG-CoA synthase (*H. sapiens*)[27]	1XPL_A	16	43,204	5.0	Mevalonate pathway	Non-Decarboxylative	EA**C**YAATPA
FabH (*E. coli*)[32]	1EBL_A	24	33,523	5.1	Fatty Acid biosynthesis	Decarboxylative	AA**C**AGFTYA
Mycobacterium Pks13[+ 50]	CAA17864	19	44,122	5.2	Mycolic Acid biosynthesis	Decarboxylative	TA**C**SSSLVA

*Via Needleman-Wunch and BLAST algorithms and comparison to OleA from *X. campestris*

[+]Alignment to keto-acyl synthase domain only, as defined by NCBI

[#]Amino acid sequence surrounding active site cysteine conserved in thiolase superfamily proteins

[@]Number of the reference from which the data represented in the table was obtained

6.3.3 Initial Defining of Substrate Specificity and Reaction Products of OleA

Enzymes catalyzing condensation or hydrolysis reactions with acyl-CoA substrates release coenzyme A that can be assayed colorimetrically using DTNB (8, 99, 100, 122). Both proposed mechanisms (Fig. 6-1 A&B) showed OleA-catalyzed coenzyme A release and this assay was used to monitor enzyme activity during purification. With purified OleA, DTNB was used to determine the stoichiometry of coenzyme A formation and to begin to discern the substrate specificity of OleA.

First, the coenzyme A product stoichiometry was determined using either myristoyl-CoA or palmitoyl-CoA and allowing the substrate to completely react. With either substrate, the reaction stoichiometry was 1.0 mole of coenzyme A released for each mole of acyl-CoA consumed. In this manner, acyl-CoA chains of different lengths were tested with OleA using a time of incubation in which palmitoyl-CoA reacts completely as described in the Methods section. Under those conditions (Table 6-2) palmitoyl-CoA reacted more completely than myristoyl-CoA. Octanoyl-, decanoyl-, lauroyl-, palmitoleoyl-, and stearoyl-CoA were also found to undergo reaction to release coenzyme A (Table 6-2), but acetyl-CoA did not.

Subsequent experiments were conducted to examine if coenzyme A was formed as a consequence of acyl-group condensation, thioester bond hydrolysis, or some mixture of the two reactions. Based on previous observations (12, 49, 108), it was known that long-chain ketones were the observed condensation products. In subsequent

Table 6-2. Substrate specificity of OleA as determined by CoA release. Values shown are the average of triplicate determinations with standard error.

Substrate		CoA	% of
Common Name	Carbon Chain Length	Product (μM)[1]	Theoretical[2]
Palmitoyl-CoA	16	65.0 +/- 0.9	100
Myristoyl-CoA	14	63.2 +/- 0.4	97
Lauroyl-CoA	12	51.4 +/- 1.9	79
Palmitoleoyl-CoA	16	36.9 +/- 0.9	57
Decanoyl-CoA	10	27.2 +/- 1.6	42
Stearoyl-CoA	18	18.7 +/- 1.8	29
Octanoyl-CoA	8	8.0 +/- 2.2	12
Acetyl-CoA	2	ND[3]	ND

[1]Free coenzyme A detected as described in the methods
[2]Starting substrate was 65 μM; 65 μM of product is 100% of theoretical yield
[3]ND = No detectable activity

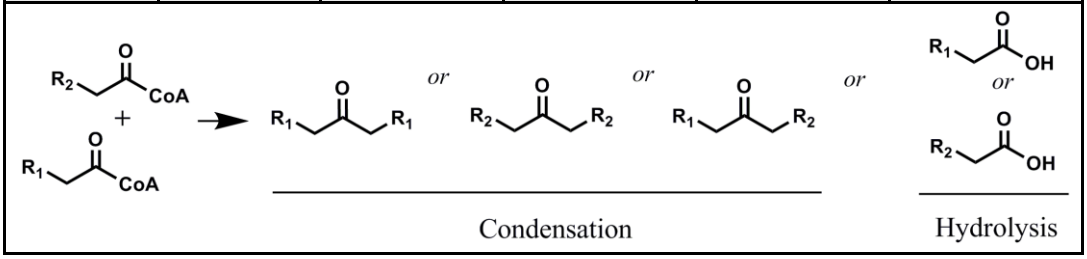
experiments in this study, it was shown that β-keto acids are the initial products and those decarboxylate quantitatively to the corresponding ketone. In this context, reaction mixtures were solvent extracted and subjected to GC-MS to identify ketones derived from condensation and/or fatty acids derived from acyl chain hydrolysis.

Previous *in vivo* experiments identified asymmetric ketones, indicating that fatty acyl chains of different chain lengths could be condensed (108). In this context, experiments were conducted with mixtures of fatty acyl-CoA substrates. All pairwise combinations of C_{10}, C_{12}, C_{14}, C_{16}, saturated and C16 monounsaturated ($C_{16:1}$) acyl-CoA substrates were incubated, extracted, and analyzed for products by GC-MS and GC-FID. In all, 15 product mixtures were analyzed. The results are shown in Table 6-3. It was found that acyl-CoA hydrolysis to the corresponding fatty acid was a major reaction in most cases. Only with C_{12} acyl condensation and C_{14} plus $C_{16:1}$ condensation were the major products derived from a condensation of fatty acyl chains. In the case of C_{14} condensations (myristoyl-CoA), the ketone 14-heptacosanone was produced at only slightly lower levels than the hydrolysis product myristic acid.

The identification of ketones as the condensation products by use of GC-MS led to the question of whether the decarboxylation was enzymatic or whether the decarboxylation occurs due to the labile nature of the reaction intermediate preceding the formation of the ketone. Further investigations were conducted using a C_1-labelled acyl-CoA substrate to track the carboxyl carbon.

Table 6-3. Product ratios determined by GC-MS for reactions of OleA with acyl-CoA substrates of different carbon chain lengths as indicated by the left-hand column and the top-row. The products, ketones (C_x) and fatty acids (FA_x), are indicated in order of decreasing abundance as determined by peak area integration as described in the Methods section. The observed partitioning between condensation of similar or different chains, or acyl-CoA hydrolysis, is illustrated at the bottom.

Fatty acyl-CoA chains	**C_{10}**	**C_{12}**	**C_{14}**	**C_{16}**	**$C_{16:1}$**
C_{10}	FA_{10} > **C_{19}**	FA_{12} > FA_{10} > **C_{23}** > **C_{21}** > **C_{19}**	FA_{10} > **C_{19}** > **C_{23}** > **C_{27}** > FA_{14}	FA_{16} > FA_{10} > **C_{19}** > **C_{25}** > **C_{31}**	FA_{10} > $FA_{16:1}$ > **$C_{25:1}$** > **C_{19}** > **$C_{31:2}$**
C_{12}	-	**C_{23}**	FA_{14} > **C_{27}** > **C_{25}** > FA_{12} > **C_{23}**	FA_{16} > **C_{23}** > FA_{12} > **C_{27}** > **C_{31}**	$FA_{16:1}$ > FA_{12} > **$C_{27:1}$** > **$C_{31:2}$**
C_{14}	-	-	FA_{14} > **C_{27}**	FA_{16} > **C_{27}** > FA_{14} > **C_{29}** > **C_{31}**	**C_{27}** > **$C_{29:1}$** > FA_{14} > **$C_{31:2}$**
C_{16}	-	-	-	FA_{16} > **C_{31}**	$FA_{16:1}$ > FA_{16} > **$C_{31:2}$** > **$C_{31:1}$**
$C_{16:1}$	-	-	-	-	$FA_{16:1}$ > **$C_{31:2}$**



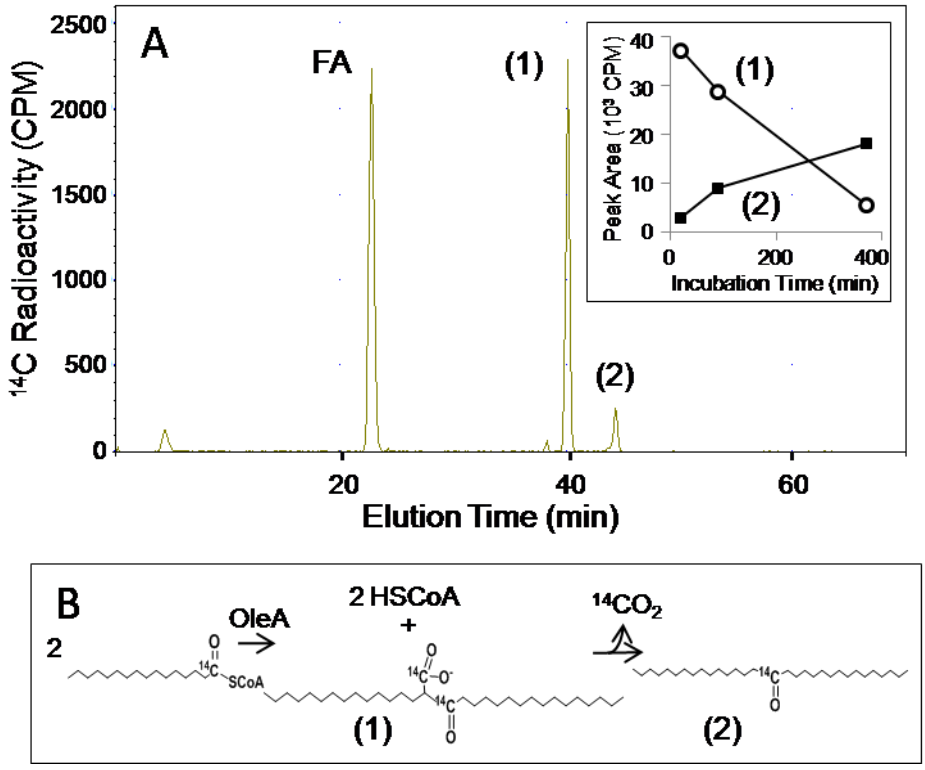


Figure 6-3. OleA reaction products with [^{14}C]-myristoyl-CoA as the substrate. (A) HPLC profile showing radioactive peaks. (Inset) Plot of the radioactivity detected in product 1 and product 2 over the course of 6 hours when a reaction mixture was incubated at room temperature. (B) Schematic of the reactions leading to the formation of product 1 and product 2.

6.3.4 Identification of Initial Condensation Product

OleA reactions with [1-^{14}C]-myristoyl-CoA were analyzed using a high pressure liquid chromatograph (HPLC) fitted with a radioflow detector. A major peak eluting at 22.4 minutes was identified as myristic acid (Fig. 6-3). The HPLC peak eluting at 44.3 (compound 2) min was analyzed by GC-MS and found to be 14-heptacosanone, but more of the radioactivity co- migrated with a more polar product eluting at 40.0 min (Fig. 6-3). The major peak eluting at 40.0 min (compound 1) showed very little absorbance at 259 nm consistent with the absence of a coenzyme A moiety. Over time, the peak at 40 min diminished with a concomitant increase in the peak at 44.3 min (Fig. 6-3, inset). This observation was consistent with a decarboxylation of the compound at 40.0 min giving rise to increasing concentrations of 14-heptacosanone over the course of 6.5 hours. This was also indicated because the compound eluting at 44.3 min contained only half of the ^{14}C as did the compound at 40 min, consistent with a loss of a carbon atom as carbon dioxide. Moreover, β-keto acids are known to be labile and the decarboxylation of 2-myristoylmyristic acid would be expected to produce 14-heptacosanone

To investigate this further, the benzyl ester of 2-myristoylmyristic acid was synthesized. It was hydrogenolyzed with palladium and hydrogen to produce 2-myristoylmyristic acid. This latter compound was observed to undergo rapid decarboxylation to produce 14-heptacosanone.

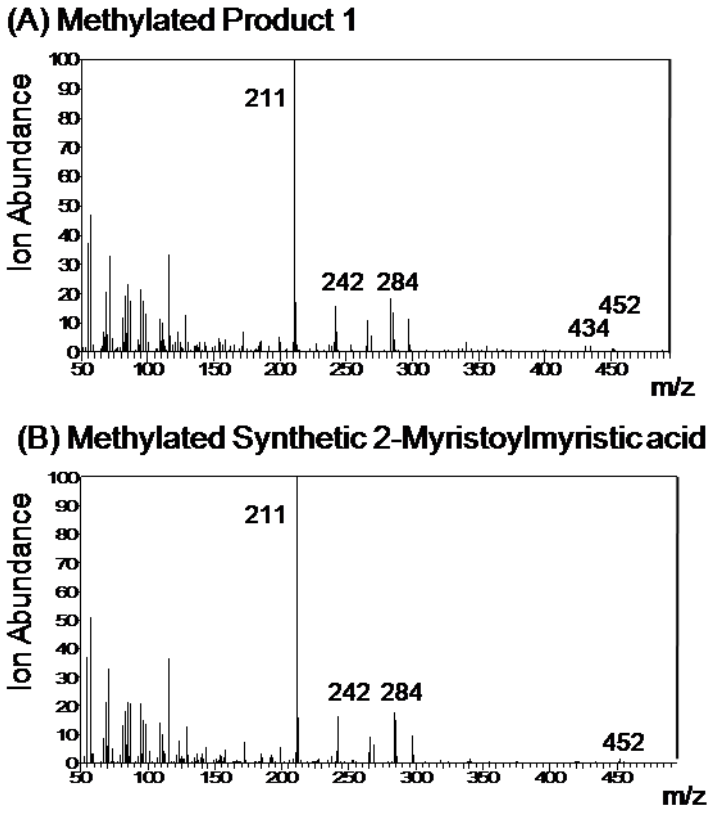


Figure 6-4. Mass spectra for products from the reaction of diazomethane with: (A) product 1 (40 min retention time) from the reaction of OleA with 2-myristoylmyristic acid, and (B) synthetic 2-myristylmyristic acid.

Table 6-4. Electrospray-ionization (ESI) mass spectrometry of product 1 from reaction of OleA with myristoyl-CoA as shown in Figure 3 and synthetic 2-myristoylmyristic acid.

Sample	Molecular Mass	ESI, negative ion mode *m/z* (intensity)
Product 1	-	393 (52), 437 (29), 473 (19)
2-Myristoyl-myristic acid	438	393 (45), 437 (36), 473 (19)

Treatment of synthetic 2-myristoyl-myristic acid with diazomethane yielded the methyl ester. The synthetic methyl ester was compared to the enzyme-produced compound collected at 40.0 min that had been immediately reacted with diazomethane. Both methylated compounds showed a GC retention time of 20.6 min and essentially identical mass spectra (Figure 6-4). The parent ion at *m/z* 452 is present in both but it is a minor ion. In this context, electrospray ionization mass spectrometry was conducted on the free acid product 1 from the OleA reaction with myristoyl-CoA (Table 6-4, top) and the synthetic standard 2-myristoylmyristic acid (Table 6-4, bottom). In this case, a major negative ion was observed (*m/z* 437) with a mass of one less than the molecular mass of 2-myristoylmyristic acid in both the biological product and the standard. A second major fragment of *m/z* 393 found in both is consistent with the loss of carbon dioxide in the mass spectrometer. Another ion fragment was detected at *m/z* 473/475, suggested to be [M – H + HCl].

6.3.5 Role of OleA in olefin biosynthesis

OleA has been proposed to function with other Ole proteins to produce olefins (12, 49, 108). Other Ole proteins were purified as described in the methods section and tested in admixture with OleA and with myristoyl-CoA as the substrate. Gas chromatography-mass spectrometry was used as it can detect both the OleA product following its decarboxylation to 14-heptacosanone and the expected olefin 14-heptacosene if the entire biosynthetic pathway were functional. Figure 5 shows that OleA and OleC in admixture produced only 14-heptacosanone (elution time 21.8), the

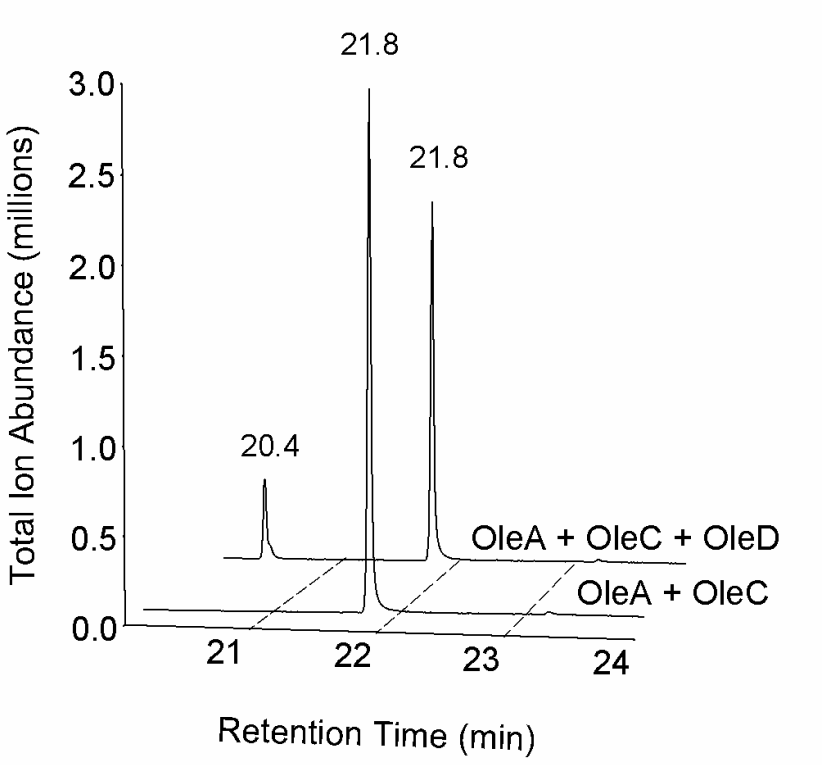


Figure 6-5. Gas chromatogram with mass detector showing products observed on co-incubations with OleA and OleC (foreground trace) and OleA, OleC and OleD (background trace).

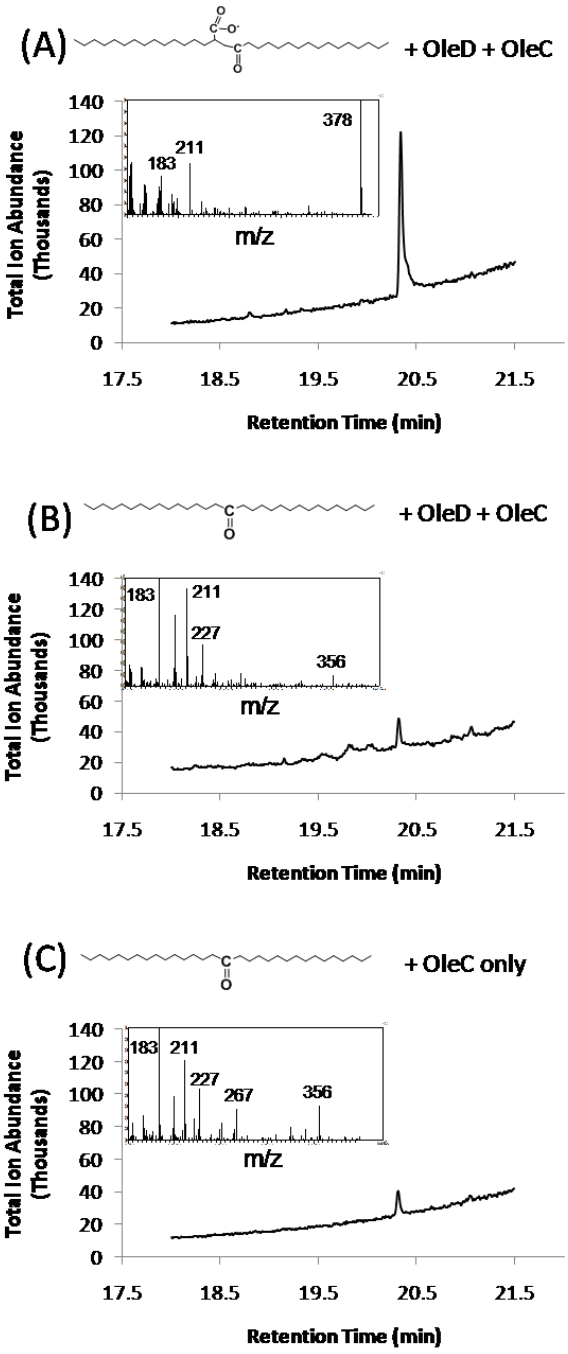


Figure 6-6. Gas chromatogram and accompanying mass spectra of peaks eluting between 20.3 and 20.4 minutes from:
(A) extract of reaction mixture containing OleC and OleD incubated with 2-myristoyl myristic acid;
(B) extract of reaction mixture containing OleC and OleD incubated with 14-heptacosanone.
(C) extract of reaction mixture containing OleC incubated with 14-heptacosanone.

product observed with OleA alone. However, when OleA was incubated with myristoyl-CoA, OleC and OleD, the olefin 14-heptacosene (20.4 min) was observed in addition to the peak corresponding to 14-heptacosanone. The identities of the products were confirmed by mass spectrometry.

To directly demonstrate that 2-myristoylmyristic acid was the intermediate giving rise to the olefin, we incubated synthetic 2-myristoylmyristic acid with OleC and OleD. The experiment yielded 14-heptacosanone and the expected olefinic product from head-to-head condensation, 14-heptacosene (Fig. 6-6, peak A, 20.4 min). The identity of the compound was confirmed by the mass spectrum shown above the GC chromatogram in Figure 6-6A, with the major ion, *m/z* 378, representing the molecular ion.

Next, it was tested if the ketone 14-heptacosanone could also give rise to 14-heptacosene or any other discernible product. In this experiment (Fig. 6-6B), OleC and OleD were incubated with 14-heptacosanone under the same conditions as described above and the olefinic product 14-heptacosene was not detected. A minor peak was observed at 20.35 min that had a mass spectrum different than that of 14-heptacosene (*m/z* 356).The small 20.35 min peak was also present in incomplete reaction mixtures that lacked OleD (Fig. 6-6C), further demonstrating it is a contaminant and not relevant to olefin biosynthesis. It was found to be a minor impurity in the synthetic ketone preparation.

Overall, these data suggest that *X. campestris* OleA produced β-ketoacid intermediates from acyl-CoAs (C_8-C_{18}) and that the ketone is a non-physiological product arising from spontaneous decarboxylation. Note that ketones have been observed *in vivo* in recombinant bacteria containing heterologous OleA genes (12, 108). Additionally, Albro and Dittmer incubated ketones with crude protein fractions and failed to observe olefins (7). However, when a full suite of *ole* genes are present in native hosts, olefinic products, and not ketones, are typically observed (109). Thus, previous *in vivo* results are fully consistent with the *in vitro* data obtained in the present study.

6.4 Discussion

In this study, the OleA protein from *X. campestris* was purified to homogeneity and shown to condense fatty acyl-CoA substrates to produce a condensed β-ketoacid with the release of two moles of CoA. The β-ketoacid, synthesized chemically or enzymatically, was shown to undergo further metabolism to yield a long-chain olefin in the presence of OleC and OleD. These studies confirmed that OleA catalyzes the first reaction in alkene biosynthesis with acyl-CoA substrates and carries out a non-decarboxylative Claisen condensation reaction.

An OleA protein was previously purified from *Micrococcus luteus* and it was proposed to catalyze a different reaction (12) than the one demonstrated here with the OleA protein from *Xanthomonas campestris*. The *Xanthomonas* and *Micrococcus*

OleA proteins showed 38% sequence identity (Table 6-1) in a pairwise alignment of their amino acid sequences (9) so they could conceivably catalyze different reactions. The *oleA* genes from both organisms cluster with *oleBCD* genes. In the *Micrococcus* genome, the *oleB* and *oleC* genes are fused and likely produce a multi-domain protein. However, the OleA, OleB, OleC and OleD domains are present in both organisms. It was shown in the present study that OleC and OleD proteins act on the β-ketoacid product generated by *X. campestris* OleA to produce a long chain olefin. When the *Micrococcus oleA* gene was cloned and expressed in *E. coli*, long chain ketones were observed (12). In the present study, the recombinant *E. coli* strain expressing the *X. campestris* OleA protein alone was also observed to produce long-chain ketones that were not observed in the wild-type *E. coli* (data not shown). The *in vitro* data in this study showed that the ketones readily arise from the decarboxylation of a corresponding β-ketoacid intermediate. These observations are all consistent with a non-decarboxylative Claisen condensation as shown in Figure 6-1B and difficult to reconcile with the proposed decarboxylative reaction shown in Figure 6-1A.

The reaction catalyzed by OleA is somewhat reminiscent of the *Zoogloea* thiolase reaction that catalyzes the first step in the biosynthesis of polyhydroxybutyrate (33). In the latter reaction however, the condensed product is a β-ketoacetyl-CoA, acetoacetyl-CoA, and with OleA, the product is a β-keto acid. Several lines of evidence strongly suggested that OleA does not produce a β-ketoacetyl-CoA that is hydrolyzed to the acid by another enzyme. First, the *oleA* gene was cloned as a single open reading frame (ORF) from synthetic DNA and expressed in *E. coli*, a bacterium that does not

natively synthesize hydrocarbons. Enzymes capable of hydrolyzing 2-myristoylmyristoyl-CoA are not likely to be present in *E. coli*. Second, OleA was highly purified as shown by SDS-PAGE (Fig. 6-2), so even the unlikely *E. coli* hydrolytic enzyme would have been removed. Lastly, our HPLC conditions would have detected 2-myrisotylmyristoyl-CoA and this was never detected.

Based on the data obtained, and the known role of the conserved cysteine found in other members of the thiolase superfamily, a working reaction mechanism can be presented for the OleA catalyzed reaction (Fig. 6-7). We propose that initially an active site cysteine in the resting enzyme (Fig. 6-7A) is acylated and coenzyme A is liberated (Fig. 6-7B). Subsequently, the tethered substrate is likely activated by an active site base to yield a carbanion on the tethered substrate (Fig. 6-7C). The carbanion then can react at the active site with the carbonyl carbon of a non-covalently bound acyl-CoA (Fig. 6-7D). That reaction forms a carbon-to-carbon bond with the condensed product still tethered to the enzyme cysteine and producing the second molecule of coenzyme A formed in the reaction cycle (Fig. 6-7E). The covalently-bound condensation product can then undergo hydrolysis to yield the final β-ketoacid product and regenerate the free cysteine residue of the resting enzyme state (Fig. 6-7A). While several features of this proposed mechanism are not yet demonstrated directly, there are multiple data that support this proposal. First, this mechanism explains the observed stoichiometry in which two moles of coenzyme A are observed per mole of condensed product. Secondly, the observed high rate of hydrolysis of acyl-CoAs to produce fatty acids is not unexpected if the enzyme has a mechanism to hydrolyze thioester-linked

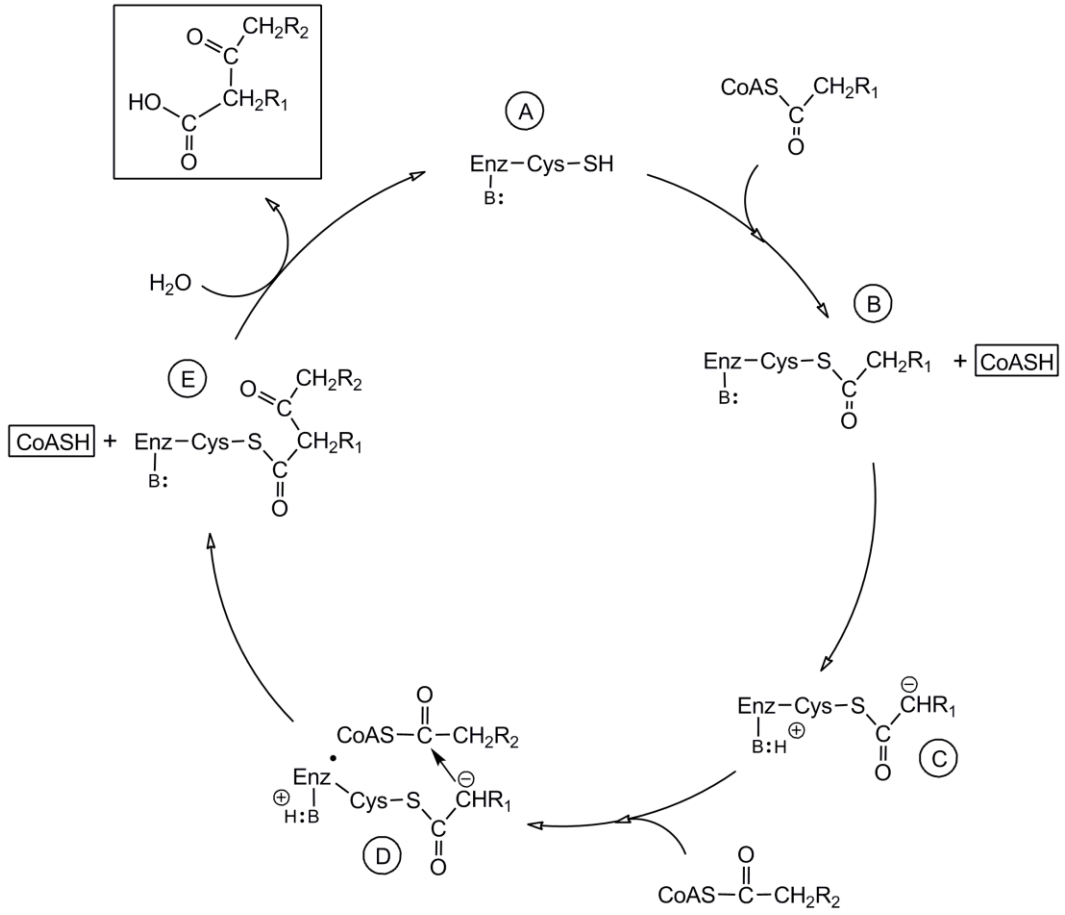


Figure 6-7. Proposed reaction cycle for OleA. The top of the cycle (A) shows the resting enzyme that reacts with an acyl-CoA to start the reaction cycle. CoASC(O)CH_2R_1 and CoASC(O)CH_2R_2 represent the first and second reacting acyl-Coenzyme A, respectively. B: attached by a line to Enz represents an enzyme base. The products of the reaction, 2 molecules of Coenzyme A (CoASH) and a b-keto acid, are highlighted by boxes.

intermediates during its normal reaction cycle. Thus, there could be a kinetic competition between hydrolysis of the initially-bound acyl-group (Fig. 6-7B) and the tethered condensation product (Fig. 6-7E). Depending upon the binding affinity for the different length acyl-CoA used in the experiment described in Table 6-3, hydrolysis of intermediate 7B or 7E would occur preferentially. Lastly, proteins in the thiolase superfamily typically use an active site cyteine to acquire an acyl chain to initiate catalysis (51, 53) and the region around the cysteine residue shown in Table 6-1 is the most highly conserved region of OleA with other members of the superfamily.

There are significant questions that remain to be addressed regarding this proposed mechanism (Fig. 6-7). First, the identity of the proposed cysteine nucleophile has not been directly demonstrated here. Second, the suggested generation of a carbanion (Fig. 6-7B) requires a general base that remains to be identified. Additionally, this mechanism would be supported by the identification of the binding sites for the acyl chains and showing that the chains are covalently and non-covalently bound, respectively.

This study identified the product of the OleA-catalyzed reaction to be a β-keto acid. The production of olefins required the presence of OleC and OleD, in addition to OleA. These data indicated that OleC and OleD catalyze further reactions with the β-ketoacid intermediate generated by OleA. This was supported by experiments in which 2-myristoyl myristic acid was transformed to an olefin by OleC and OleD. The corresponding ketone was not transformed to an olefin, consistent with the idea that the ketone is not a physiologically-relevant intermediate. There is also the issue that C-2 in

2-myristoyl myristic acid is a chiral center. The synthetic 2-myristoyl myristic acid is racemic and it is plausible that only one enantiomer will react with OleD. The chirality of the reaction is currently under investigation.

6.5 Ole enzyme pathway

The identification of the OleA reaction intermediate allowed for the reaction pathway to be more completely hypothesized, Figure 6-8. The OleD reaction was investigated in conjunction with OleA as part of this chapter and preliminary OleC data is described in Chapter 7.

6.6 Publication of thesis work

With the exception of Figure 6-8, this research was originally published in the Journal of Biological Chemistry. Frias, Janice A., Jack E. Richman, Jasmine S. Erickson, and Lawrence P. Wackett. Purification and characterization of OleA from *Xanthamonas campestris* and demonstration of a non-decarboxylative Claisen condensation. *The Journal of Biological Chemistry*. Published online 1/25/11. © the American Society for Biochemistry and Molecular Biology. Dr. Jack Richman was responsible for the chemical synthesis and analysis of the beta keto acid standard, as well as helpful advice and discussions.

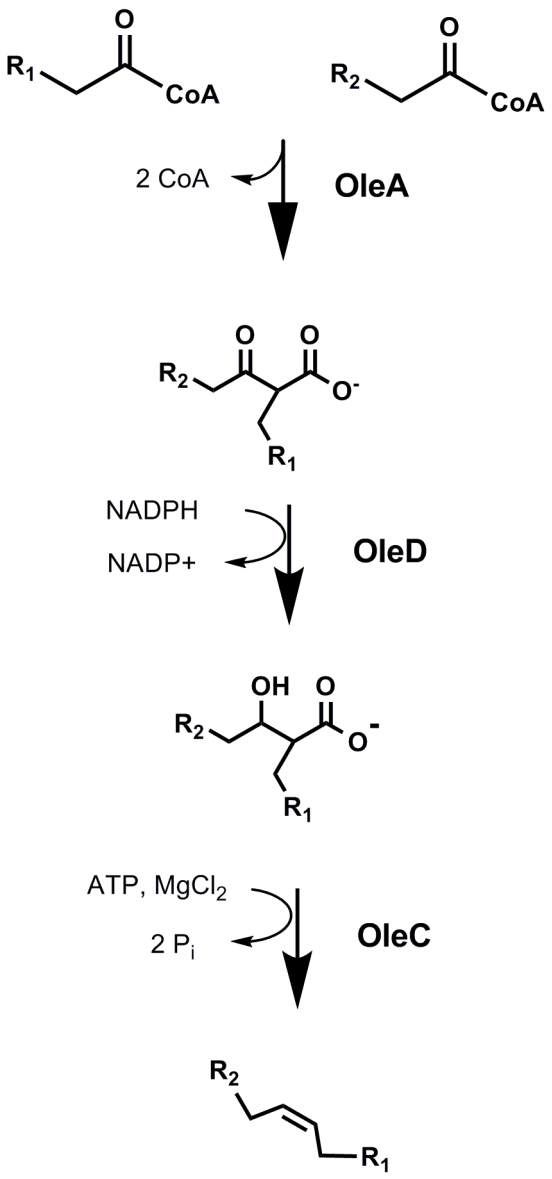


Figure 6-8 Ole Enzyme Pathway

6.7 Acknowledgements

As submitted in the manuscript: We thank David Sukovich for providing a selection of *oleA* genes used in these studies. We acknowledge Fred Schendel and Mary Pruss for fermentation and cell harvesting. Special thanks to Drs. Sharon Murphy and Linda von Weymarn for assistance with the HPLC radioflow equipment. We thank Brandon Goblirsch and Carrie Wilmot for insightful discussions regarding the OleA mechanism described here. Thanks to Dr. Burckhard Seelig and lab members J. Haugen and L. Hagman for the use and instruction of the FPLC and size exclusion column. Mass spectrometry (ESI only) was conducted in the Center for Mass Spectrometry and Proteomics and Mass Spectrometry Services in the Masonic Cancer Center, with the assistance of Tom Krick, Brock Matter, and Peter Villalta.The research was supported by the University of Minnesota Doctoral Dissertation Fellowship to JAF and the Department of Energy ARPA-E under Award Number DE-AR0000007 to LPW. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability of responsibility for the accuracy, completeness, or usefulness of any information, appartus, product, or process disclosed, or represents that its use would infringe privately controlled rights. References herein to any specific commercial product, process, or service by tradename, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and

opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

CHAPTER 7

Cloning, purification, crystallization and preliminary X-ray diffraction of the OleC protein from *Stenotrophomonas maltophilia* involved in head-to-head hydrocarbon biosynthesis

7.1 Introduction

The biological mechanisms of hydrocarbon biosynthesis have recently attracted attention as a means of producing important commercial chemicals from renewable resources (93, 103). Plants, animals and microbes have evolved several different biosynthetic pathways for generating hydrocarbons, but the biochemical details are only now beginning to be revealed. A head-to-head condensation of fatty acids that generates long-chain olefins has been known for more than 40 years (2), but only in the last year have the olefin (*ole*) biosynthetic genes been revealed (12). The head-to-head condensation reaction requires a minimum of three gene products (OleACD). A genetic

knockout of the *oleC* gene led to the loss of hydrocarbon biosynthesis in *Shewanella oneidensis* MR-1(108). In a recent study, 69 divergent bacteria were indicated to generate olefins *via* an OleC-dependent biosynthetic pathway (109). The genes have yet to be demonstrated in plants and animals, although marine eukaryotic algae make similar compounds (91).

OleC is a member of the LuxE acyl-protein synthetase superfamily based on a conserved-domain search at the National Center for Biotechnology Information (NCBI). This family includes LuxE, which is involved in bioluminescence, and fatty acyl-CoA synthase, which is involved in the ligation of fatty acids to a coenzyme A moiety with an AMP-activated acyl group as an intermediate. As of 8 April 2010, 63 crystal structures in the Protein Data Bank belong to this superfamily, with the most closely related being only 26% identical to OleC in amino-acid sequence. Only OleC is known to be involved in olefin biosynthesis. Previous studies suggested that the Ole proteins from *Stenotrophomonas maltophilia* would have a relatively broad specificity for different fatty-acid chain lengths and degrees of unsaturation (107). Thus, the *S. maltophilia oleC* gene was selected for cloning and expression studies. The purification and crystallization of an OleC protein has not previously been described.

7.2 Experimental

7.2.1 Cloning of the oleC gene

DNA consisting of the *S. maltophilia* ATCC 17679 *oleC* gene sequence (49) and flanking *Nde*I and *Hin*dIII restriction sites was synthesized by the GenScript Corporation (Piscataway, New Jersey, USA). The DNA was cloned into a pET30b vector (Novagen, Madison, Wisconsin, USA) containing a C-terminal His tag. The recombinant plasmid was transformed into *Escherichia coli* BL21 (DE3) pLysE One Shot cells (Invitrogen) for expression.

7.2.2 Expression and purification of OleC

E. coli BL21(pOleC) cells were cultured in 500 ml LB medium containing kanamycin (50 µg ml^{-1}) and chloramphenicol (34 µg ml^{-1}) at 310 K. Cultures were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.1 m*M* when the OD_{600} of the culture reached 0.65-0.70. After 4 h at 310 K, the induced cells were harvested by centrifugation at 3000*g* for 25 min and resuspended in 20 m*M* sodium phosphate, 500 m*M* NaCl pH 7.4 buffer with EDTA-free protease inhibitors (Roche). The cells were disrupted by three passes through a chilled French pressure cell at 8.3 MPa and centrifuged at 27 000*g* for 90 min to remove cell debris and insoluble protein. The soluble fraction was either filtered through a 0.45 µm filter or centrifuged for 30 min prior to loading onto a Pharmacia Biotech LCC 501 FPLC fitted with a 5 ml HisTrap HP (Amersham Biosciences) column complexed with Ni^{2+} and equilibrated with 20 m*M* sodium phosphate and 500 m*M* NaCl pH 7.4. The His-tagged OleC protein eluted at 200 m*M* imidazole. The purity of the protein was confirmed by SDS-PAGE, with a single band running at 60 kDa (Fig. 7-1). The protein

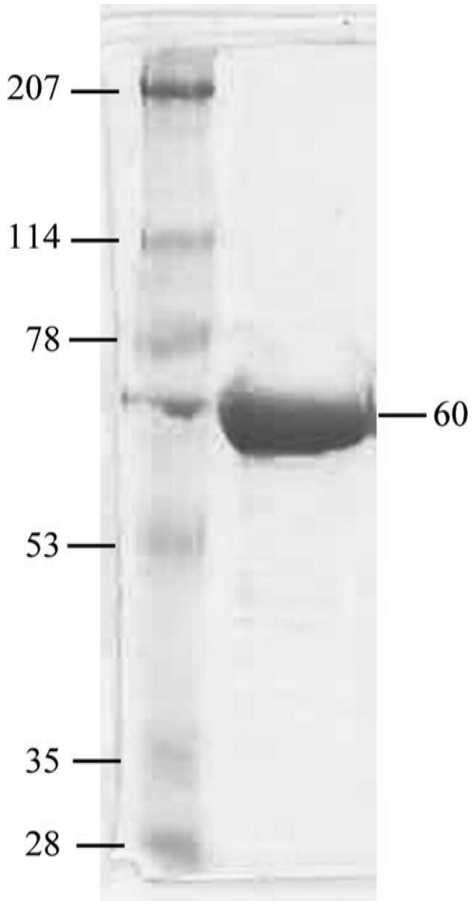


Figure 7-1. SDS-PAGE analysis of OleC.

Proteins were analyzed on a 10% SDS-PAGE gel and stained with SimplyBlue Safe Coomassie stain. The left lane contains standard molecular-weight markers (kDa); the right lane contains purified OleC.

was concentrated to 8-13 mg ml^{-1} and the buffer was exchanged for 20 m*M* HEPES, 500 m*M* NaCl pH 7.4 using a 50 ml Amicon pressure concentrator with a YM-10 membrane (Millipore). After centrifugation at 27 000*g* for 20 min to remove precipitated protein, 2.2 m*M* adenosine 5'-monophosphate (5'-AMP) was added. The protein was rocked gently on ice for 1.5 h prior to flash-freezing in liquid nitrogen for storage.

7.2.3 Crystallization

Initial crystallization trials of OleC were carried out by the Hauptman-Woodward Medical Research Institute High-Throughput Screening (HTS) laboratory. The HTS library tests 1536 different chemical conditions for crystallization *via* the microbatch-under-oil method. When very few hits resulted in crystals from the initial screen, crystallization trials of OleC were repeated in the presence of 2.2 m*M* 5'-AMP. The inclusion of 5'-AMP was based on the success of cocrystallization of other LuxE-superfamily proteins with an acyl-adenylate or acyl group and 5'-AMP substrate (121). OleC crystals grew under multiple conditions (>100) in a week when 5'-AMP was included in the OleC sample, suggesting conformational heterogeneity in the absence of this substrate. The crystals used for X-ray data collection were grown in HTS condition 1303 (10% PEG 8000 and 10% PEG 1000), which contains no buffer. This condition was transferred to hanging-drop vapor diffusion in VDX plates (Hampton Research) using equal amounts (1.0 µl) of protein solution and reservoir solution.

Optimization was achieved by a fine screen around the PEG 8000 and PEG 1000 concentrations.

7.2.4 X-ray diffraction data collection, processing and structure solution

The crystal was soaked in a cryoprotectant solution consisting of 50% Paratone-N and 50% paraffin oil (Hampton Research). When all of the aqueous solution had been washed away, the crystal was mounted in a nylon loop and flash-frozen in liquid nitrogen. Data were collected at a temperature of 100 K on beamline 4.2.2 at the Advanced Light Source, Berkeley, California. The resulting diffraction data were processed with *HKL*-2000 (85). Further diffraction data analysis was conducted with the *CCP*4 suite (28).

7.3 Results and discussion

7.3.1 Crystallization

The final optimized conditions for crystal growth were 10% PEG 8000 and 11% PEG 1000 at 293 K. Crystals suitable for X-ray analysis appeared after 1 d by hanging-drop vapor diffusion. Crystals of OleC only grew in the presence of 5'-AMP. The crystals were clear, with a bipyramid-like morphology (Fig. 7-2).

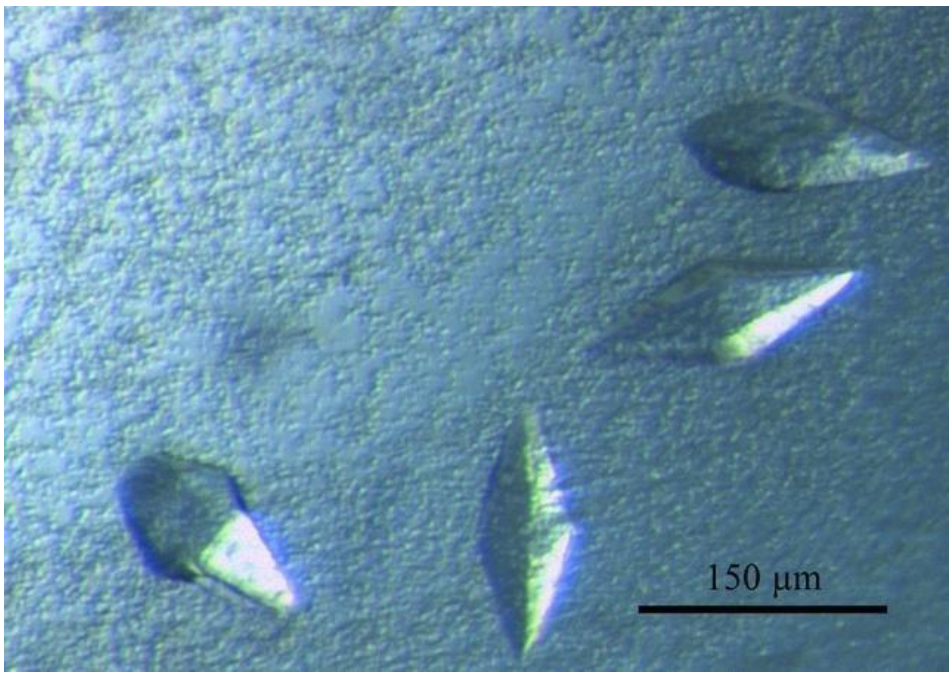


Figure 7-2. OleC crystals grown in the presence of 5'-AMP. Average crystal dimensions are 150 × 40 × 40 µm.

Table 1
Data-collection statistics.

Values in parentheses are for the highest resolution shell.

Space group	$P3_121$ or $P3_221$
Unit-cell parameters	
$a = b$ (Å)	98.8
c (Å)	141.0
Wavelength (Å)	0.9784
Resolution range	50–3.40 (3.46–3.40)
No. of unique reflections	115600 (11407)
R_{merge}† (%)	11.9 (40.9)
Completeness (%)	100 (100)
Redundancy	10.0 (9.8)

† $R_{merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the observed intensity and $\langle I(hkl)\rangle$ is the average intensity of multiple measurements.

Table 7-1. Data-collection statistics

Values in parentheses are for the highest resolution shell.

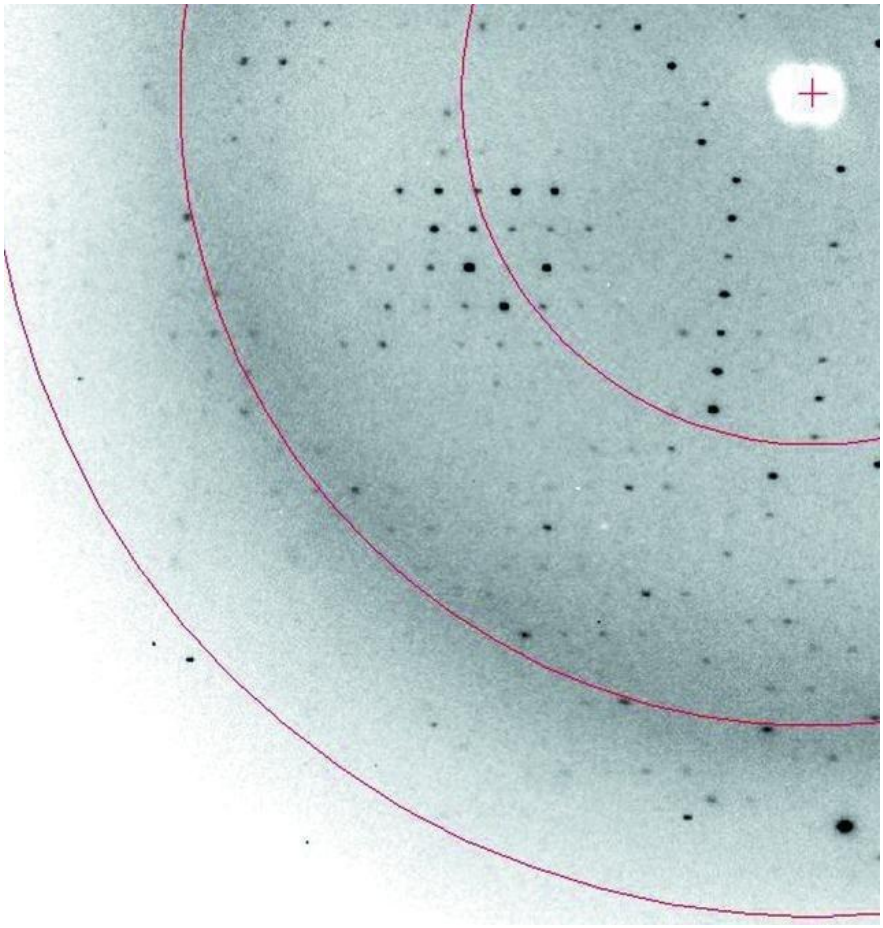


Figure 7-3. Section of an X-ray diffraction image from OleC crystals grown in the presence of 5'-AMP. The red-colored arcs indicate resolution (from left to right the arcs are at 3.5, 4.5 and 8.0 Å resolution).

7.3.2 Crystallization and X-ray diffraction data collection

The majority of OleC crystals grown in the presence of 5'-AMP did not diffract. The problem is likely to be one of crystal fragility towards handling, as the use of alternative cryoprotectants (glycerol, xylitol, 2-methyl-1,3-propanediol, PEG 400, PEG 600 and perfluoropolyether) had no impact in improving crystal diffraction. Overall, one out of every 30 crystals was suitable for data collection, with diffraction to 10 Å resolution in-house and diffraction to 3.4 Å resolution at the synchrotron. Data-collection statistics are provided in Table 7-1 and a diffraction image is shown in Fig. 7-3. Initial data processing demonstrated that the crystals belonged to the trigonal space group $P3_121$ or $P3_221$, with unit-cell parameters $a = b = 98.8$, $c = 141.0$ Å. Analysis of the Matthews coefficient (75) suggested the presence of one OleC monomer (60 kDa) per asymmetric unit, with a solvent content of 62.73%. The data are of moderate quality, with an overall R_{merge} of 11.9% and an estimated mosaicity of 0.93°. The average *B* factor derived from the Wilson plot is 70.1 $Å^2$ (120). Attempts to solve the structure by molecular replacement are currently under way.

7.4 Acknowledgements

This research was supported by the National Institutes of Health grant GM-90260 to CMW, Chemistry-Biology Interface Training Grant GM-008700 to JAF and a Graduate School doctoral dissertation fellowship to JAF. This material is partly based upon work supported by the Department of Energy ARPA-E under Award No. DE-

AR0000007 to LPW. X-ray data were collected at the Kahlert Structural Biology Laboratory (KSBL) at The University of Minnesota (supported by Minnesota Partnership for Biotechnology and Medical Genomics grant SPAP-05-0013-P-FY06) and beamline 4.2.2, Molecular Biology Consortium, Advanced Light Source (ALS), Berkeley, California, USA. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences of the US Department of Energy under Contract No. DE-AC02-05CH11231. Computer resources were provided by the Basic Sciences Computing Laboratory (BSCL) of the University of Minnesota Supercomputing Institute. We thank Ed Hoeffner for KSBL support, Jay Nix and the staff at beamline 4.2.2, ALS for their support and Can Ergenekan for BSCL support.

7.5 Publication of thesis work

This work has been previously published: Frias*, Janice A., Brandon R. Goblirsch*, Lawrence P. Wackett, and Carrie M. Wilmot. Cloning, purification, crystallization and preliminary X-ray diffraction of the OleC protein from *Stenotrophomonas maltophilia* involved in head-to-head hydrocarbon biosynthesis**.** Acta Crystallographica F. 66 (9) Acta Crystallographica F. 2010. *Equal contribution. Dr. Carrie Wilmot and graduate student Brandon Goblirsch brought the crystallography expertise to this work.

CHAPTER 8

Conclusions and Recommendations for Future Work

8.1 Thesis conclusions

Developing a new feasible biofuel is a great challenge. Understanding the potential tools that can be useful in biofuel efforts will be immensely helpful in overcoming this challenge. This thesis focused on understanding microbial biosynthesis of hydrocarbons that are known in the literature. With the decreased cost of genome sequencing and data mining ramping up in recent years, our progress on pathway elucidation of long chain hydrocarbons in microbes has taken a large step forward. Although the phenomenon of bacteria producing long chain alkenes had been reported more than 50 years ago (6, 109, 112), the genes involved in the pathway have just been identified in the last few years (12, 49).

The identification of the genes in hydrocarbon biosynthesis has allowed for the study of the enzymes in the pathway. It was shown in this work that OleA, OleC, and OleD added together *in vitro* with cofactors and acyl-CoAs will produce alkenes. The reaction product of OleA was determined to be an unstable β-keto acid condensation product that non-enzymatically decarboxylates to a long chain ketone. Substrate specificity of OleA and OleC show a strong preference for long chains, but each organism may likely have a more narrow preferred range. From the preliminary

studies on OleD, it was determined that the second enzymatic step is the reduction of the hydroxyl of the β-keto acid. The *in vitro* studies described in this thesis can be useful in determining how to tailor the *in vivo* pathway for the synthesis of potential biofuels, especially regarding the specificity of the enzyme.

8.2 Future work

8.2.1 OleA

Although we now have a much greater understanding of OleA, there is still a lot of investigation being done. Preliminary work has been completed in growing OleA crystals with and without the inhibitors, cerulenin and iodoacteamide, and Brandon Goblirsch of the Wilmot lab has solved the structures. The structures are currently being used to study mutants of residues potentially involved in the catalysis and to understand the reaction mechanism by which OleA condenses two long chain fatty acid residues. Two avenues the structure can also help us with are to tailor the enzymes binding sites to limit the length or the type of fatty acyl-CoA that are accepted by the enzyme for the most desirable production of hydrocarbons. Another avenue that the structure and mechanistic studies can provide insights into is the hydrolysis reaction of fatty acyl-CoAs, which leads to inefficiency in the production of long chain molecules. Perhaps the enhanced understanding can lead to a mutant that will only condense fatty acyl-CoA rather than hydrolyze without condensation.

Another important question that remains about OleA is the kinetics of the enzyme. Currently with the hydrolysis and the condensation reactions happening simultaneously, it has been difficult to gain an understanding of how fast the condensation reaction is happening and how the reaction compares with various substrates. Being able to perform kinetic studies will also aide in the evaluation of various mutants, and perhaps an enzyme with higher rates of condensation can be identified by random mutagenesis. A typical assay of the kinetics of thiolase enzymes found in the literature, involve using the subsequent enzyme in the pathway such as the reductase in the fatty acid synthesis pathway (32). Another assay involves measuring the thiolytic cleavage reaction of the condensed product, such as using the *Zoogloea* thiolase and acetoacetyl-CoA (33). Neither of these assays has been feasible to date in the Ole pathway. The reverse reaction has also not been feasible with OleA. The OleA product is different in nature in that an additional hydrolysis occurs in which the second CoA is cleaved off, which is likely an irreversible step. Using a coupled assay has also been unsuccessful, as the purification of OleD has not been high enough in yield to be able to add the enzyme in excess to obtain OleA rate information.

8.2.3 OleD

Obtaining higher yield of OleD soluble protein would be desirable for studies of both the OleD reaction as well as OleA kinetic studies. To date, I have studied 5 OleD enzymes. Two of these enzymes are active in alkene formation, *X. campestris*

and *S. maltophilia*, and worth pursuing further. Extensive studies on the purification of the *X. campestris* OleD have been performed, and proved to be unsuccessful in obtaining high yields of soluble OleD. The stability of the *X. campestris* OleD appeared to be more of an issue, as OleD appeared to be falling out of solution over time (Appendix VI). From Chapter 5, it was seen that there was relatively high soluble expression levels of OleD from *S. maltophilia*. Future work includes the purification and study of the *S. maltophilia* OleD *in vitro*.

8.2.4 OleC

The hypothetical AMP ligase, OleC, raises many questions still. The superfamily appears to be fairly consistent in the overall reaction of ligating AMP from ATP to the substrate and then transferring a CoA for the final product (102). OleC does not align well with known AMP ligases in the C terminal domain, but based on the much larger N terminal domain sequence it is identified into the LuxE/AMP Ligase family. From the sequence information and homologous enzymes, OleC was able to be studied preliminarily, but the mechanism is still not understood. Once the preceding enzyme in the pathway, OleD, is obtained in higher yields, this could potentially allow for more definitive studies of the OleC mechanism.

8.2.5 OleB

OleB has been scarcely mentioned in this thesis. Preliminary studies on OleB appeared in the patent literature (48, 49), but no mechanism or defined role had been identified. Due to soluble expression problems and the non-necessity of OleB in alkene formation *in vitro* and *in vivo*, OleB was not a focus of this thesis. The conservation of OleB in the genomes within this pathway, suggests that it does play a role in alkene synthesis. The variety of reactions included in this enzyme family does not lead to a straightforward hypothesis of the function of OleB. The lab is pursuing a potential role of OleB in assisting in the proper stereoselectivity of the OleA product for the OleD enzyme. No experimentation can be done regarding this hypothesis until the purification yield and OleB stability is improved.

REFERENCES

1. **Adams, R., W. E. Bachman, L. F. Fieser, J. R. Johnson, and H. R. Snyder (ed.).** 1942. Organic Reactions, 9th ed, vol. 1. John Wiley & Sons, Inc., New York.

2. **Albro, P. W., and J. C. Dittmer.** 1969. The biochemistry of long-chain, nonisoprenoid hydrocarbons. 3. Metabolic relationship of long-chain fatty acids and hydrocarbons and other aspects of hydrocarbon metabolism in *Sarcinea lutea*. Biochemistry **8:**1913-1918.

3. **Albro, P. W., and J. C. Dittmer.** 1969. The biochemistry of long-chain, nonisoprenoid hydrocarbons. 1. Characterization of hydrocarbons of *Sarcinea lutea* and isolation of possible intermediates of biosynthesis. Biochemistry **8:**394-405.

4. **Albro, P. W., and J. C. Dittmer.** 1969. The biochemistry of long-chain, nonisoprenoid hydrocarbons. 4. Characteristics of synthesis by a cell-free preparation of *Sarcinea lutea*. Biochemistry **8:**3317-3324.

5. **Albro, P. W., and J. C. Dittmer.** 1969. The biochemistry of long-chain, nonisoprenoid hydrocarbons. 2. The incorporation of acetate and the aliphatic chains of isoleucine and valine into fatty acids and hydrocarbon by *Sarcina lutea in vivo*. Biochemistry **8:**953-959.

6. **Albro, P. W., and C. K. Huston.** 1964. Lipids of *Sarcinea lutea* .2. Hydrocarbon content of lipid extracts. Journal of Bacteriology **88:**981-986.

7. **Albro, P. W., T. D. Meehan, and J. C. Dittmer.** 1970. Intermediate steps in the incorporation of fatty acids into long-chain, nonisoprenoid hydrocarbons by lysates of *Sarcina lutea.* Biochemistry **9:**1893-1898.

8. **Alexson, S. E. H., and J. Nedergaard.** 1988. A novel type of short-chain and medium-chain acyl-CoA hydrolases in brown adipose-tissue mitochondria. Journal of Biological Chemistry **263:** 13564-13571.

9. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. Journal of Molecular Biology **215:**403-410.

10. **Alvarez, H. M., and A. Steinbuchel.** 2002. Triacylglycerols in prokaryotic microorganisms. Applied Microbiology and Biotechnology **60:**367-376.

11. **Bassler, B. L., C. Yu, Y. C. Lee, and S. Roseman.** 1991. Chitin utilization by marine bacteria – degradation and catablolism of chitin oligosaccharides by *Vibrio furnisii*. Journal of Biological Chemistry **266:**24276-24286.

12. **Beller, H. R., E. B. Goh, and J. D. Keasling.** 2010. Genes Involved in Long-Chain Alkene Biosynthesis in *Micrococcus luteus*. Applied and Environmental Microbiology **76:**1212-1223.

13. **Bligh, E. G., and W. J. Dyer.** 1959. A rapid method of total lipid extraction and purification. Canadian Journal of Biochemistry and Physiology **37:**911-917.

14. **Bobik, T. A.** 2006. Polyhedral organelles compartmenting bacterial metabolic processes. Applied Microbiology and Biotechnology **70:**517-525.

15. **Bognar, A. L., G. Paliyath, L. Rogers, and P. E. Kolattukudy.** 1984. Biosynthesis of alkanes by particulate and solubilized enzyme preparations from pea leaves (*Pisum sativum*) Archives of Biochemistry and Biophysics **235:**8-17.

16. **Brenner, D. J., F. W. Hickmanbrenner, J. V. Lee, A. G. Steigerwalt, G. R. Fanning, D. G. Hollis, J. J. Farmer, R. E. Weaver, S. W. Joseph, and R. J. Seidler.** 1983. *Vibrio furnissii* (formerly aerogenic biogroup of *Vibrio fluvialis*), a new species isolated from human feces and the environment. Journal of Clinical Microbiology **18:**816-824.

17. **Briese, R. R., and S. M. McElvain.** 1933. The Acetoacetic Ester Condensation V. The Condensation of Higher Esters. J Am Chem Soc **55:**1697-1700.

18. **Bugg, T. D. H.** 2004. Diverse catalytic activities in the alpha beta-hydrolase family of enzymes: activation of H_2O, HCN, H_2O_2, and O_2. Bioorganic Chemistry **32:**367-375.

19. **Cardona, C. A., and O. J. Sanchez.** 2007. Fuel ethanol production: Process design trends and integration opportunities. Bioresource Technology **98:**2415-2457.

20. **Cassagne, C., and R. Lessire.** 1974. Studies on alkane biosynthesis in epidermis of *Allum porrum* L. leaves: Direct synthesis of tricosane from lignoceric acid. Archives of Biochemistry and Biophysics **165:**274-280.

21. **Caudales, R., C. Forni, and J. M. Wells.** 1998. Cellular fatty acid composition of rod and coccus forms of *Arthrobacter globiformis, A. crystallopoietes* and *A. nicotianae* isolated from the water fern Azolla. Journal of Applied Microbiology **84:**784-790.

22. **Chang, M. C. Y., and J. D. Keasling.** 2006. Production of isoprenoid pharmaceuticals by engineered microbes. Nature Chemical Biology **2:**674-681.

23. **Channon, H. J., and A. C. Chibnall.** 1929. The Ether-Soluble Substances of Cabbage Leaf Cytoplasm. Biochem J. **23:**168-175.

24. **Chen, C. Y., K. M. Wu, Y. C. Chang, C. H. Chang, H. C. Tsai, T. L. Liao, Y. M. Liu, H. J. Chen, A. B. T. Shen, J. C. Li, T. L. Su, C. P. Shao, C. T. Lee, L. I. Hor, and S. F. Tsai.** 2003. Comparative genome analysis of *Vibrio vulnificus*, a marine pathogen. Genome Research **13:**2577-2587.

25. **Chen, J. S., and S. F. Hiu.** 1986. Acetone, butanol, isopropanol production by *Clostridium beijerinckii* (Synonym, *Clostridium butylicum*). Biotechnology Letters **8:**371-376.

26. **Chibnall, A. C., and S. H. Piper.** 1934. The Metabolism of Plant and Insect Waxes. Biochem J **28:**2209-2219.

27. **Clinkenbeard, K. D., T. Sugiyama, W. D. Reed, and M. D. Lane.** 1975. Cytoplasmic 3-hydroxy-3-methylglutaryl Coenzyme A synthase from liver – purification, properties, and role in cholesterol synthesis. Journal of Biological Chemistry **250:**3124-3135.

28. **Collaborative Computation Project, N.** 1994. The CCP4 suite: programs for protein crystallography. Acta Cryst. **D50:**760-763.

29. **Colwell, R. R.** 2004. Infectious disease and environment: cholera as a paradigm for waterborne disease. International Microbiology **7:**285-289.

30. **Connor, M. R., and J. C. Liao.** 2009. Microbial production of advanced transportation fuels in non-natural hosts. Current Opinion in Biotechnology **20:**307-315.

31. **Coyle, W. T.** May 2010. Next-Generation Biofuels. *In* U. S. D. o. A. Economic Research Service (ed.), www.ers.usda.gov.

32. **Davies, C., R. J. Heath, S. W. White, and C. O. Rock.** 2000. The 1.8 angstrom crystal structure and active-site architecture of beta-ketoacyl-acyl carrier protein synthase III (FabH) from Escherichia coli. Structure **8:**185-195.

33. **Davis, J. T., R. N. Moore, B. Imperiali, A. J. Pratt, K. Kobayashi, S. Masamune, A. J. Sinskey, C. T. Walsh, T. Fukui, and K. Tomita.** 1987. Biosynthetic thiolase from Zoogloea ramigera. 1. Preliminary characterization and analysis of proton transfer reaction. Journal of Biological Chemistry **262:**82-89.

34. **Demirbas, A.** 2011. Competitive liquid biofuels from biomass. Applied Energy **88:**17-28.

35. **Dennis, M., and P. E. Kolattukudy.** 1992. A cobalt porphyrin enzyme converts a fatty aldehyde to a hydrocarbon and CO. Proceedings of the National Academy of Sciences of the United States of America **89:**5306-5310.

36. **Department of Energy, U. G.**, posting date. www.eia.doe.gov. [Online.]

37. **Detalle, J. F., A. Riahi, V. Steinmetz, F. Henin, and J. Muzart.** 2004. Mechanistic insights into the palladium-induced domino reaction leading to ketones from benzyl beta-ketoesters: First characterization of the enol as an intermediate. Journal of Organic Chemistry **69:**6528-6532.

38. **Dodge, A. G.** 2008. Insights into Microbial Adaptive Potential from Case Studies of Novel Functional Group Biocatalysis and Microbe-Metal Interactions. University of Minnesota - Twin Cities, St. Paul, MN.

39. **Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky.** 2000. Use of repetitive DNA sequences and the PCR to differentiate Escherichia coli isolates from human and animal sources. Applied and Environmental Microbiology **66:**2572-2577.

40. **Eddy, S. R.** 1998. Profile hidden Markov models. Bioinformatics **14:**755-763.

41. **Eglinton, G., and R. J. Hamilton.** 1967. Leaf Epicuticular Waxes. Science **156:**1322-1335.

42. **Ellman, G. L.** 1958. A colorimetric method for determining low concentrations of mercaptans. Arch Biochem Biophys **74:**443-450.

43. **Esther, C. R., H. M. Jasin, L. B. Collins, J. A. Swenberg, and G. Boysen.** 2008. A mass spectrometric method to simultaneously measure a biomarker and dilution marker in exhaled breath condensate. Rapid Communications in Mass Spectrometry **22:**701-705.

44. **Fakhr, M. K., L. K. Nolan, and C. M. Logue.** 2005. Multilocus sequence typing lacks the discriminatory ability of pulsed-field gel electrophoresis for typing *Salmonella enterica* serovar Typhimurium. Journal of Clinical Microbiology **43:**2215-2219.

45. **Filling, C., K. D. Berndt, J. Benach, S. Knapp, T. Prozorovski, E. Nordling, R. Ladenstein, H. Jornvall, and U. Oppermann.** 2002. Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. Journal of Biological Chemistry **277:**25677-25684.

46. **Frias, J. A., B. R. Goblirsch, L. P. Wackett, and C. M. Wilmot.** 2010. Cloning, purification, crystallization and preliminary X-ray diffraction of the OleC protein from *Stenotrophomonas maltophilia* involved in head-to-head hydrocarbon biosynthesis. Acta Crystallographica Section F-Structural Biology and Crystallization Communications **66:**1108-1110.

47. **Frias, J. A., J. E. Richman, and L. P. Wackett.** 2009. C-29 Olefinic Hydrocarbons Biosynthesized by Arthrobacter Species. Applied and Environmental Microbiology **75:**1774-1777.

48. **Friedman, L., and B. Da Costa (LS9, Inc.).** 2008. USA. WO 2008/147781 A2.

49. **Friedman, L., and M. Rude.** 2008. USA. WO 2008/113041.

50. **Gavalda, S., M. Leger, B. van der Rest, A. Stella, F. Bardou, H. Montrozier, C. Chalut, O. Burlet-Schiltz, H. Marrakchi, M. Daffe, and A. Quemard.** 2009. The Pks13/FadD32 Crosstalk for the Biosynthesis of Mycolic Acids in *Mycobacterium tuberculosis*. Journal of Biological Chemistry **284:**19255-19264.

51. **Haapalainen, A. M., G. Merilainen, and R. K. Wierenga.** 2006. The thiolase superfamily: condensing enzymes with diverse reaction specificities. Trends in Biochemical Sciences **31:**64-71.

52. **Hansen, A. C., Q. Zhang, and P. W. L. Lyne.** 2005. Ethanol-diesel fuel blends - a review. Bioresource Technology **96:**277-285.

53. **Heath, R. J., and C. O. Rock.** 2002. The Claisen condensation in biology. Natural Product Reports **19:**581-596.

54. **Heidelberg, J. F., J. A. Eisen, W. C. Nelson, R. A. Clayton, M. L. Gwinn, R. J. Dodson, D. H. Haft, E. K. Hickey, J. D. Peterson, L. Umayam, S. R. Gill, K. E. Nelson, T. D. Read, H. Tettelin, D. Richardson, M. D. Ermolaeva, J. Vamathevan, S. Bass, H. Y. Qin, I. Dragoi, P. Sellers, L. McDonald, T. Utterback, R. D. Fleishmann, W. C. Nierman, O. White, S. L. Salzberg, H. O. Smith, R. R. Colwell, J. J. Mekalanos, J. C. Venter, and C. M. Fraser.** 2000. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. Nature **406:**477-483.

55. **Holmquist, M.** 2000. Alpha/Beta-Hydrolase Fold Enzymes: Structures, Functions and Mechanisms. Current Protein & Peptide Science **1:**209-235.

56. **Huber, G. W., and A. Corma.** 2007. Synergies between bio- and oil refineries for the production of fuels from biomass. Angewandte Chemie-International Edition **46:**7184-7201.

57. **Ishii, S., T. Yan, D. A. Shively, M. N. Byappanahalli, R. L. Whitman, and M. J. Sadowsky.** 2006. Cladophora (Chlorophyta) spp. harbor human bacterial pathogens in nearshore water of Lake Michigan. Applied and Environmental Microbiology **72:**4545-4553.

58. **Itaya, K., and M. Ui.** 1966. A New Micromethod for the Colorimetric Determination of Inorganic Phosphate. Clinica Chimica Acta **14:**361-366.

59. **Jerke, K., C. H. Nakatsu, F. Beasley, and A. Konopka.** 2008. Comparative analysis of eight Arthrobacter plasmids. Plasmid **59:**73-85.

60. **Johnson, L. K., M. B. Brown, E. A. Carruthers, J. A. Ferguson, P. E. Dombek, and M. J. Sadowsky.** 2004. Sample size, library composition, and genotypic diversity among natural populations of *Escherichia coli* from different animals influence accuracy of determining sources of fecal pollution. Applied and Environmental Microbiology **70:**4478-4485.

61. **Jones, D. T., and D. R. Woods.** 1986. Acetone-butanol fermentation revisited. Microbiological Reviews **50:**484-524.

62. **Keddie, R. M., and D. Jones.** 2006. The genus *Arthrobacter*. *In* M. Dworkin (ed.), The prokaryotes, 3rd ed., vol. 3. Springer-Verlag, New York, NY.

63. **Kerfeld, C. A., M. R. Sawaya, S. Tanaka, C. V. Nguyen, M. Phillips, M. Beeby, and T. O. Yeates.** 2005. Protein structures forming the shell of primitive bacterial organelles. Science **309:**936-938.

64. **Kessler, D., I. Leibrecht, and J. Knappe.** 1991. Pyruvate formate lyase deactivase and acetyl-CoA reductase activities of *Escherichia coli* reside on a polymeric protein particle encoded by *adhe*. Febs Letters **281:**59-63.

65. **Keyhani, N. O., and S. Roseman.** 1996. The chitin catabolic cascade in the marine bacterium *Vibrio furnissii* - Molecular cloning, isolation, and characterization of a periplasmic chitodextrinase. Journal of Biological Chemistry **271:**33414-33424.

66. **Kloos, W. E., T. G. Tornabene, and K. H. Schleifer.** 1974. Isolation and characterization of micrococci from human skin, including two new species: *Micrococcus lylae* and *Micrococcus kristinae*. J. Syst. Bacteriol. **24:**79-101.

67. **Kolattukudy, P.** 1968. Biosynthesis of surface lipids. Biosynthesis of long-chain hydrocarbons and waxy esters is discussed. Science **159:**498-505.

68. **Kolattukudy, P. E.** 1966. Biosynthesis of Wax in *Bassica oleracea*. Relation of Fatty Acids to Wax.*. Biochemistry **5:**2265-2275.

69. **Kolattukudy, P. E., J. S. Buckner, and L. Brown.** 1972. Direct evidence for a decarboxylation mechanism in the biosynthesis of alkanes in *B. oleracea*. Biochem Biophys Res Commun **47:**1306-1313.

70. **Koskinen, M., P. Vodicka, and K. Hemminki.** 2000. Adenine N3 is a main alkylation site of styrene oxide in double-stranded DNA. Chemico-Biological Interactions **124:**13-27.

71. **Lee, S. K., H. Chou, T. S. Ham, T. S. Lee, and J. D. Keasling.** 2008. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. Current Opinion in Biotechnology **19:**556-563.

72. **Li, X. B., and S. Roseman.** 2004. The chitinolytic cascade in Vibrios is regulated by chitin oligosaccharides and a two-component chitin catabolic sensor/kinase. Proceedings of the National Academy of Sciences of the United States of America **101:**627-631.

73. **Luque, R., L. Herrero-Davila, J. M. Campelo, J. H. Clark, J. M. Hidalgo, D. Luna, J. M. Marinas, and A. A. Romero.** 2008. Biofuels: a technological perspective. Energy & Environmental Science **1:**542-564.

74. **Markey, S. P., and T. G. Tornabene.** 1971. Characterization of branched monounsaturated hydrocarbons of Sarcina lutea and Sarcina flava. Lipids **6:**190-195.

75. **Matthews, B. W.** 1968. Solvent content of protein crystals. J. Mol. Biol. **33:**491-497.

76. **Meyer, F., A. Goesmann, A. C. McHardy, D. Bartels, T. Bekel, J. Clausen, J. Kalinowski, B. Linke, O. Rupp, R. Giegerich, and A. Puhler.** 2003. GenDB - an open source genome annotation system for prokaryote genomes. Nucleic Acids Research **31:**2187-2195.

77. **Miyamoto, K.** 2001. Microorganisms and process for producing petroleum substitute oil by using these microorganisms. Japan.

78. **Mongodin, E. F., N. Shapir, S. C. Daugherty, R. T. Deboy, J. B. Emerson, A. Shvartzbeyn, D. Radune, J. Vamathevan, F. Riggs, V. Grinberg, H. Khouri, L. P. Wackett, K. E. Nelson, and M. J. Sadowsky.** 2006. Secrets of soil survival revealed by the genome sequence of Arthrobacter aurescens TC1. Plos Genetics **2:**2094-2106.

79. **Muntendam, R., E. Melillo, A. Ryden, and O. Kayser.** 2009. Perspectives and limits of engineering the isoprenoid metabolism in heterologous hosts. Applied Microbiology and Biotechnology **84:**1003-1019.

80. **Nordin, K.** 2004. Stockholm University, Stockholm, Sweden.

81. **Novagen.** 2003. pET System Manual, www.novagen.com, 10th ed.

82. **Okada, K., T. Iida, K. Kita-Tsukamoto, and T. Honda.** 2005. Vibrios commonly possess two chromosomes. Journal of Bacteriology **187:**752-757.

83. **Oppermann, U., C. Filling, M. Hult, N. Shafqat, X. Q. Wu, M. Lindh, J. Shafqat, E. Nordling, Y. Kallberg, B. Persson, and H. Jornvall.** 2003. Short-chain dehydrogenases/reductases (SDR): the 2002 update. Chemico-Biological Interactions **143:**247-253.

84. **Osman, K. T., L. Q. Du, Y. J. He, and Y. Luo.** 2009. Crystal Structure of Bacillus cereus D-Alanyl Carrier Protein Ligase (DltA) in Complex with ATP. Journal of Molecular Biology **388:**345-355.

85. **Otwinowski, Z., and W. Minor.** 1997. Processing of X-ray diffraction data collected in oscillation mode, p. 307-326, Macromolecular Crystallography, Pt A, vol. 276.

86. **Park, M. O.** 2005. New pathway for long-chain n-alkane synthesis via 1-alcohol in Vibrio furnissii M1. Journal of Bacteriology **187:**1426-1429.

87. **Park, M. O., K. Heguri, K. Hirata, and K. Miyamoto.** 2005. Production of alternatives to fuel oil from organic waste by the alkane-producing bacterium, Vibrio furnissii M1. Journal of Applied Microbiology **98:**324-331.

88. **Park, M. O., M. Tanabe, K. Hirata, and K. Miyamoto.** 2001. Isolation and characterization of a bacterium that produces hydrocarbons extracellularly which are equivalent to light oil. Applied Microbiology and Biotechnology **56:**448-452.

89. **Peralta-Yahya, P. P., and J. D. Keasling.** 2010. Advanced biofuel production in microbes. Biotechnology Journal **5:**147-162.

90. **Piper, S., A. Chibnall, S. Hopkins, A. Pollard, J. Smith, and E. Williams.** 1931. Synthesis and crystal spacings of certain long-chain paraffins, ketones, and secondary alcohols. Biochem J **25:**2095-2110.

91. **Rieley, G., M. A. Teece, T. M. Peakman, A. M. Raven, K. J. Greene, T. P. Clarke, M. Murray, J. W. Leftley, C. N. Campbell, R. P. Harris, R. J. Parkes, and J. R. Maxwell.** 1998. Long-chain alkenes of the haptophytes *Isochrysis galbana* and *Emiliania huxleyi*. Lipids **33:**617-625.

92. **Ruby, E. G., M. Urbanowski, J. Campbell, A. Dunn, M. Faini, R. Gunsalus, P. Lostroh, C. Lupp, J. McCann, D. Millikan, A. Schaefer, E. Stabb, A. Stevens, K. Visick, C. Whistler, and E. P. Greenberg.** 2005. Complete genome sequence of *Vibrio fischeri*: A symbiotic bacterium with pathogenic congeners. Proceedings of the National Academy of Sciences of the United States of America **102:**3004-3009.

93. **Rude, M. A., and A. Schirmer.** 2009. New microbial fuels: a biotech perspective. Current Opinion in Microbiology **12:**274-281.

94. **Sadowsky, M. J., R. E. Tully, P. B. Cregan, and H. H. Keyser.** 1987. Genectic diversity in *Bradyrhizobium japonicum* serogroup 123 and its relation to genotype specific nodulation of soybean. Applied and Environmental Microbiology **53:**2624-2630.

95. **Sambrook, J., and D. W. Russell.** 2001. Molecular cloning: a laboratory manual, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, HY.

96. **Schirmer, A., M. A. Rude, X. Z. Li, E. Popova, and S. B. del Cardayre.** 2010. Microbial Biosynthesis of Alkanes. Science **329:**559-562.

97. **Schneider-Belhaddad, F., and P. Kolattukudy.** 2000. Solubilization, partial purification, and characterization of a fatty aldehyde decarbonylase from a higher plant, *Pisum sativum*. Archives of Biochemistry and Biophysics **377:**341-349.

98. **Sivakumar, G., D. R. Vail, J. F. Xu, D. M. Burner, J. O. Lay, X. M. Ge, and P. J. Weathers.** 2010. Bioethanol and biodiesel: Alternative liquid fuels for future generations. Engineering in Life Sciences **10:**8-18.

99. **Skaff, D. A., and H. M. Miziorko.** 2010. A visible wavelength spectrophotometric assay suitable for high-throughput screening of 3-hydroxy-3-methylglutaryl-CoA synthase. Analytical Biochemistry **396:**96-102.

100. **Sleeman, M. C., and C. J. Schofield.** 2004. Carboxymethylproline synthase (CarB), an unusual carbon-carbon bond-forming enzyme of the crotonase superfamily involved in carbapenem biosynthesis. Journal of Biological Chemistry **279:**6730-6736.

101. **Sonnhammer, E. L. L., S. R. Eddy, and R. Durbin.** 1997. Pfam: A comprehensive database of protein domain families based on seed alignments. Proteins-Structure Function and Genetics **28:**405-420.

102. **Starai, V. J., and J. C. Escalante-Semerena.** 2004. Acetyl-coenzyme A synthetase (AMP forming). Cellular and Molecular Life Sciences **61:**2020-2030.

103. **Steen, E. J., Y. S. Kang, G. Bokinsky, Z. H. Hu, A. Schirmer, A. McClure, S. B. del Cardayre, and J. D. Keasling.** 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature **463:**559-U182.

104. **Steinbuchel, A., and S. Hein.** 2001. Biochemical and molecular basis of microbial synthesis of polyhydroxyalkanoates in microorganisms. Adv. Biochem. Eng. Biotechnol. **71:**81-123.

105. **Strong, L. C., C. Rosendahl, G. Johnson, M. J. Sadowsky, and L. P. Wackett.** 2002. Arthrobacter aurescens TC1 metabolizes diverse s-triazine ring compounds. Applied and Environmental Microbiology **68:**5973-5980.

106. **Suen, Y., G. U. Holzer, J. S. Hubbard, and T. G. Tornabene.** 1988. Biosynthesis of acyclic methyl branched poly unsaturated hydrocarbons in *Pseudomonas maltophilia*. Journal of Industrial Microbiology **2:**337-348.

107. **Suen, Y., G. U. Holzer, J. S. Hubbard, and T. G. Tornabene.** 1988. Biosynthesis of acyclic methyl branched polyunsaturated hydrocarbons in *Pseudomonas maltophilia*. Journal of Industrial Microbiology & Biotechnology **2:**337-348.

108. **Sukovich, D. J., J. L. Seffernick, J. E. Richman, J. A. Gralnick, and L. P. Wackett.** 2010. Widespread Head-to-Head Hydrocarbon Biosynthesis in Bacteria and Role of OleA. Applied and Environmental Microbiology **76:**3850-3862.

109. **Sukovich, D. J., J. L. Seffernick, J. E. Richman, K. A. Hunt, J. A. Gralnick, and L. P. Wackett.** 2010. Structure, Function, and Insights into the Biosynthesis of a Head-to-Head Hydrocarbon in *Shewanella oneidensis* Strain MR-1. Applied and Environmental Microbiology **76:**3842-3849.

110. **Tettelin, H., V. Masignani, M. J. Cieslewicz, C. Donati, D. Medini, N. L. Ward, S. V. Angiuoli, J. Crabtree, A. L. Jones, A. S. Durkin, R. T. DeBoy, T. M. Davidsen, M. Mora, M. Scarselli, I. M. Y. Ros, J. D. Peterson, C. R. Hauser, J. P. Sundaram, W. C. Nelson, R. Madupu, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, S. A. Sullivan, S. C. Daugherty, D. H. Haft, J. Selengut, M. L. Gwinn, L. W. Zhou, N. Zafar, H. Khouri, D. Radune, G. Dimitrov, K. Watkins, K. J. B. O'Connor, S. Smith, T. R. Utterback, O. White, C. E. Rubens, G. Grandi, L. C. Madoff, D. L. Kasper, J. L. Telford, M. R. Wessels, R. Rappuoli, and C. M. Fraser.** 2005. Genome analysis of multiple pathogenic isolates of Streptococcus

agalactiae: Implications for the microbial "pan-genome". Proceedings of the National Academy of Sciences of the United States of America **102:**13950-13955.

111. **Tornabene,T. G., E. O. Bennett, and J. Oro.** 1967. Fatty acid and aliphatic hydrocarbon composition of *Sarcina lutea* grown in 3 different media. Journal of Bacteriology **94:**344-348.

112. **Tornabene,T. G., E. Gelpi, and J. Oro.** 1967. Identification of fatty acids and aliphatic hydrocarbons in *Sarcina lutea* by gas chromatography and combined gas chromatography-mass spectrometry. Journal of Bacteriology **94:**333-343.

113. **Tornabene, T. G., and J. Oro.** 1967. 14-C incorporation into the fatty acids and aliphatic hydrocarbons of *Sarcina lutea*. J. Bacteriol. **94:**349-358.

114. **Tornabene, T. G., and S. L. Peterson.** 1978. *Pseudomonas maltophilia* – identification of hydrocarbons, glycerides, and glycolipoproteins of cellular lipids. Canadian Journal of Microbiology **24:**525-582.

115. **Trivedi, O. A., P. Arora, V. Sridharan, R. Tickoo, D. Mohanty, and R. S. Gokhale.** 2004. Enzymic activation and transfer of fatty acids as acyl-adenylates in mycobacteria. Nature **428:**441-445.

116. **Unell, M., N. Kabelitz, J. K. Jansson, and H. J. Heipieper.** 2007. Adaptation of the psychrotroph *Arthrobacter chlorophenolicus* A6 to growth temperature and the presence of phenols by changes in the anteiso/iso ratio of branched fatty acids. FEMS Microbiology Letters **266:**138-143.

117. **Vioque, J., and P. E. Kolattukudy.** 1997. Resolution and purification of an aldehyde-generating and an alcohol-generating fatty acyl-CoA reductase from pea leaves (*Pisum sativum* L). Archives of Biochemistry and Biophysics **340:**64-72.

118. **Wackett, L. P.** 2008. Microbial-based motor fuels: science and technology. Microb. Biotechnol. **1:**211-225.

119. **Wackett, L. P., J. A. Frias, J. L. Seffernick, D. J. Sukovich, and S. M. Cameron.** 2007. Genomic and biochemical studies demonstrating the absence of an alkane-producing phenotype in *Vibrio furnissii* M1. Applied and Environmental Microbiology **73:**7192-7198.

120. **Wilson, A. J. C.** 1949. The probability distribution of X-ray intensities. Acta Cryst. **2:**318-321.

121. **Wu, R., J. Cao, X. F. Lu, A. S. Reger, A. M. Gulick, and D. Dunaway-Mariano.** 2008. Mechanism of 4-chlorobenzoate: Coenzyme a ligase catalysis. Biochemistry **47:**8026-8039.

122. **Yashiro, K., Y. Kameyama, M. Mizunokamiya, S. O. Shin, and A. Fujita.** 1995. Substrate specificity of microsomal 1-acyl-sn-glycero-3-phosphoinositol acyltransferase in rat submandibular gland for polyunsaturated long chain acyl-CoAs. Biochimica Et Biophysica Acta-Lipids and Lipid Metabolism **1258:**288-296.

123. **Yeung, P. S., and K. J. Boor.** 2004. Epidemiology, pathogenesis, and prevention of foodborne *Vibrio parahaemolyticus* infections. Foodborne Pathog. Dis. **1:**74-88.

Appendix I: Thiolase (OleA) sequence from Arthrobacter, AAur_1998

DNA:

```
   1 ttggcaggga atgcgacctt ccggcacagc aacaccgcgc tgctctcggt gagcagcgtc
  61 gaggctccga ggatagtgag ctccacggat ttcgaccgga gattggcttc gaccctgcag
 121 cggctgaagt ttccaccgcg attgcttgaa cgtgttgcag ggatcacaca ccggcgctgg
 181 tgggccgccg ggacatcgtt cgatgacgct gccgtggaag cgggcgcaaa agccttggct
 241 gaagccggag tcgaagcctc ggaagtcggc ttgctgatca acacctcggt taccaggcgc
 301 aacctggaac cgtccgtagc ggtcaagata caccacgaac ttggactgcc gtcctcggcc
 361 atgaactttg acctggccaa cgcctgtctt ggatttgtga acggcttgat cctcgcagcc
 421 aacatgatcg attcgggcca gatcaggtat gcggtcatcg ttaacggaga ggacgcgcaa
 481 ggcactcagg aagccacgct ggcgcgactc caaaggcctg agaccacgcg tgaagacttt
 541 aatcgggaat ttgccaccct gacgctcgga tccggagctg ccgcggccgt tctgggcccc
 601 cgtgatgagt accccggggc gcaccggctt gtgggtgggg tcatgcgtgc cggcaccgag
 661 catcacgaac tgtgtgttgg tggaatcgac ggcatgtcca ccgacacgaa aggcttgctg
 721 gacggcggcc tgcagcttgt ggtggacgcc tggcatgaag cccagcctga gtgggactgg
 781 gccagcatgg accgctacgt cacacatcaa gtcagtaatg cctataccca ggcgatcatc
 841 aatgccatcg acctggaccc ggataaggtt cccatcactt tcccgcactg gggcaatgtg
 901 ggtcctgcat cgcttcccat gacgctcgca gcccaggcgc agtcattgac ctcgggagac
 961 cgggtcctgt gcatgggcgt gggctcaggc ctgaacgctg gaatggtgga aatcgtttgg
1021 tag
```

Protein:

```
   1 magnatfrhs ntallsvssv eaprivsstd fdrrlastlq rlkfpprlle rvagithrrw
  61 waagtsfdda aveagakala eagveasevg llintsvtrr nlepsvavki hhelglpssa
 121 mnfdlanacl gfvnglilaa nmidsgqiry avivngedaq gtqeatlarl qrpettredf
 181 nrefatltlg sgaaaavlgp rdeypgahrl vggvmragte hhelcvggid gmstdtkgll
 241 dgglqlvvda wheaqpewdw asmdryvthq vsnaytqaii naidldpdkv pitfphwgnv
 301 gpaslpmtla aqaqsltsgd rvlcmgvgsg lnagmveivw
```

Appendix II: Stenotrophomonas gene sequences (as ordered through GenScript)

OlnA (OleA) rev: 1029 bp, 1016 bp cut

CATATGTTATTCAAAAATGTATCTATCGCTGGTCTGGCTCATGTGGATGCTCCGCATACC
CTGACCACCAAGGAGATCAACGAACGCCTGCAACCGACGCTGGATCGCCTGGGTATCCGT
ACCGACGTACTGGGCGACATCGCTGGTATCCACGCACGTCGTCTGTGGGACAATGGCGTA
CTGGCTTCTGATGCTGCAACTATGGCAGGCCGTAAAGCACTGGAGGACGCTGGTATTAAC
GCGACTCAGGTAGGCCTGCTGGTAAACACCTCCGTAAGCCGCGATTACCTGGAGCCTTCC
ACCGCCTCCATTGTGTCTGGCAACCTGGGCGTGTCTGATGAATGCATGACCTTCGATGTT
GCTAACGCTTGCCTGGCATTCATTAACGGTATGGATATCGCAGCACGCATGATTGAACGC
GGTGACATCGATTACGCCCTGGTAGTAGACGGTGAAACCGCAAATCTGGTGTACGAAAAG
ACCCTGGAACGTATGACTGCACCGGATGTGACTGCAGATGACTTCCGTAACGAGCTGGCT
GCTCTGACCACCGGTTCTGGCGCGGCAGCGATGGTGATGGCTCGCTCCGAGCTGGTCCCT
GATGCGCCGCGTTACAAGGGTGGTGTGACCCGTAGCGCCACCGAATGGAACCAACTGTGT
CTGGGTAACCTGGACCGTATGGTTACCGACACTCGTCTGCTGCTGATCGAAGGTATCAAG
CTGGCTCAGAAAACCTTTACCGCCGCTAAAATTGCTCTGGGTTGGGCAGTGGAAGAACTG
GACCAGTTCGTGATTCACCAGGTCAGCCAACCGCACACTGCGGCGTTCATCAAGAACTTC
GGTATCGACCCTAAAAAGGTGATGACCATTTTCGGTGAGCACGGCAACATCGGCCCAGCT
TCTGTACCGATCGTTCTGTCTAAACTGAAACAGCTGGGCAAACTGAAAAAGGGCGACCGC
ATCGCGCTGCTGGGCATCGGTAGCGGCCTGAACTGTAGCATGGCTGAGGTTGTGTGGAAG
CTTCTCGAG

Protein in pET30b

ATGTTATTCAAAAATGTATCTATCGCTGGTCTGGCTCATGTGGATGCTCCGCATACC
CTGACCACCAAGGAGATCAACGAACGCCTGCAACCGACGCTGGATCGCCTGGGTATCCGT
ACCGACGTACTGGGCGACATCGCTGGTATCCACGCACGTCGTCTGTGGGACAATGGCGTA
CTGGCTTCTGATGCTGCAACTATGGCAGGCCGTAAAGCACTGGAGGACGCTGGTATTAAC
GCGACTCAGGTAGGCCTGCTGGTAAACACCTCCGTAAGCCGCGATTACCTGGAGCCTTCC
ACCGCCTCCATTGTGTCTGGCAACCTGGGCGTGTCTGATGAATGCATGACCTTCGATGTT
GCTAACGCTTGCCTGGCATTCATTAACGGTATGGATATCGCAGCACGCATGATTGAACGC
GGTGACATCGATTACGCCCTGGTAGTAGACGGTGAAACCGCAAATCTGGTGTACGAAAAG
ACCCTGGAACGTATGACTGCACCGGATGTGACTGCAGATGACTTCCGTAACGAGCTGGCT
GCTCTGACCACCGGTTCTGGCGCGGCAGCGATGGTGATGGCTCGCTCCGAGCTGGTCCCT
GATGCGCCGCGTTACAAGGGTGGTGTGACCCGTAGCGCCACCGAATGGAACCAACTGTGT
CTGGGTAACCTGGACCGTATGGTTACCGACACTCGTCTGCTGCTGATCGAAGGTATCAAG
CTGGCTCAGAAAACCTTTACCGCCGCTAAAATTGCTCTGGGTTGGGCAGTGGAAGAACTG
GACCAGTTCGTGATTCACCAGGTCAGCCAACCGCACACTGCGGCGTTCATCAAGAACTTC
GGTATCGACCCTAAAAAGGTGATGACCATTTTCGGTGAGCACGGCAACATCGGCCCAGCT
TCTGTACCGATCGTTCTGTCTAAACTGAAACAGCTGGGCAAACTGAAAAAGGGCGACCGC
ATCGCGCTGCTGGGCATCGGTAGCGGCCTGAACTGTAGCATGGCTGAGGTTGTGTGG
AAGCTTGCGGCCGCACTCGAGCACCACCACCACCACCACTGA

```
MLFKNVSIAGLAHVDAPHTLTTKEINERLQPTLDRLGIRTDVLGDIAGIHARRLWDNGVL  60
ASDAATMAGRKALEDAGINATQVGLLVNTSVSRDYLEPSTASIVSGNLGVSDECMTFDVA  120
NACLAFINGMDIAARMIERGDIDYALVVDGETANLVYEKTLERMTAPDVTADDFRNELAA  180
LTTGSGAAAMVMARSELVPDAPRYKGGVTRSATEWNQLCLGNLDRMVTDTRLLLIEGIKL  240
AQKTFTAAKIALGWAVEELDQFVIHQVSQPHTAAFIKNFGIDPKKVMTIFGEHGNIGPAS  300
VPIVLSKLKQLGKLKKGDRIALLGIGSGLNCSMAEVVWKLAAALEHHHHHH  351
Molecular Weight:  37.77 kDa
```

OlnB (OleB) rev: 903 bp, 890 bp cut

CATATGTCCCAGCTTCCCGGTTACCCCGCCCACCCGCAGCGCTTCGAGGTACGCCCCGGC
CTGTCGATGAACTATCTCGACGAAGGCCCGCGCGATGGCGAAGTGGTGGTGATGGTGCAC
GGCAACCCGTCGTGGAGCTATTACTGGCGCACGCTGGTCGCCGGCCTGTCGGACACGTAC
CGCTGCATCGTGCCGGACCACATCGGCATGGGCCTGTCGGACAAGCCCGATGACAGCCGC
TACGAGTACACGCTGCAGTCGCGCGTTGACGACCTCGATGCGCTGCTCAAGCATCTGGGC
ATAACCGGCCCGGTGACCCTGGCGGTGCACGACTGGGGCGGCATGATCGGTTTCGGCTGG
GCGCTGTCGCACCACGACCAGGTCAAGCGCCTGGTGGTGCTCAATACCGCTGCATTCCCG
ATGCCGGCGGCGAAGAAGATGCCGTGGCAGATCGCGCTGGGCCGCCACTGGAAGATCGGC
GAGTGGATCATCCGCACCTTCAACGCGTTCTCCTCCGGTGCCTCGTGGCTGGGCGTGGAG
CGGAAGATGCCGGCCGACGTGCGCCGCGCCTACGTGTCGCCGTACAACAGCTGGGCCAAC
CGCATCAGCACCATCCGCTTCATGCAGGACATCCCGCTGTCGCCGGCCGACAAGGCGTGG
TCGCTGCTGGAGCATGCCGGCAAGGCGCTGCCGTCGTTCGCCGACCGGCCGGCCTTCCTC
GGCTGGGGCCTGCGCGACTTCGTGTTCGACCACCACTTCCTGAAGGGCTTCCAGGCCGCG
CTGCCGCAGGCCCAGGTACATGCGTTCGAGGACGCCGGCCACTACGTGCTGGAAGACAAG
CACGAAGTGCTGGTGCCGGAAATCCGCGCGTTCCTGGACAAGAACCCGATCAAGCTTCTC
GAG

Protein

ATGTCCCAGCTTCCCGGTTACCCCGCCCACCCGCAGCGCTTCGAGGTACGCCCCGGC
CTGTCGATGAACTATCTCGACGAAGGCCCGCGCGATGGCGAAGTGGTGGTGATGGTGCAC
GGCAACCCGTCGTGGAGCTATTACTGGCGCACGCTGGTCGCCGGCCTGTCGGACACGTAC
CGCTGCATCGTGCCGGACCACATCGGCATGGGCCTGTCGGACAAGCCCGATGACAGCCGC
TACGAGTACACGCTGCAGTCGCGCGTTGACGACCTCGATGCGCTGCTCAAGCATCTGGGC
ATAACCGGCCCGGTGACCCTGGCGGTGCACGACTGGGGCGGCATGATCGGTTTCGGCTGG
GCGCTGTCGCACCACGACCAGGTCAAGCGCCTGGTGGTGCTCAATACCGCTGCATTCCCG
ATGCCGGCGGCGAAGAAGATGCCGTGGCAGATCGCGCTGGGCCGCCACTGGAAGATCGGC
GAGTGGATCATCCGCACCTTCAACGCGTTCTCCTCCGGTGCCTCGTGGCTGGGCGTGGAG
CGGAAGATGCCGGCCGACGTGCGCCGCGCCTACGTGTCGCCGTACAACAGCTGGGCCAAC
CGCATCAGCACCATCCGCTTCATGCAGGACATCCCGCTGTCGCCGGCCGACAAGGCGTGG
TCGCTGCTGGAGCATGCCGGCAAGGCGCTGCCGTCGTTCGCCGACCGGCCGGCCTTCCTC
GGCTGGGGCCTGCGCGACTTCGTGTTCGACCACCACTTCCTGAAGGGCTTCCAGGCCGCG
CTGCCGCAGGCCCAGGTACATGCGTTCGAGGACGCCGGCCACTACGTGCTGGAAGACAAG
CACGAAGTGCTGGTGCCGGAAATCCGCGCGTTCCTGGACAAGAACCCGATCAAGCTTGCG
GCCGCACTCGAGCACCACCACCACCACCACTGA

MSQLPGYPAHPQRFEVRPGLSMNYLDEGPRDGEVVVMVHGNPSWSYYWRTLVAGLSDTYR 60
CIVPDHIGMGLSDKPDDSRYEYTLQSRVDDLDALLKHLGITGPVTLAVHDWGGMIGFGWA 120
LSHHDQVKRLVVLNTAAFPMPAAKKMPWQIALGRHWKIGEWIIRTFNAFSSGASWLGVER 180
KMPADVRRAYVSPYNSWANRISTIRFMQDIPLSPADKAWSLLEHAGKALPSFADRPAFLG 240
WGLRDFVFDHHFLKGFQAALPQAQVHAFEDAGHYVLEDKHEVLVPEIRAFLDKNPIKLAA 300
ALEHHHHHH 309
Molecular Weight = 34.99 kDa

OlnC (OleC) rev: 1671 bp, 1658 bp cut

CATATGAACCGACCCTGCAATATTGCCGCCCGCTTGCCTGAACTGGCGCGTGAACGGCCC
GACCAGATCGCCATCCGCTGCCCGGGCCGCCGTGGCGCCGGCAACGGCATGGCCGCCTAC
GACGTGACCCTGGATTACCGCCAGCTGGACGCCCGCAGCGACGCCATGGCGGCTGGCCTG
GCAGGCTACGGGATCGGTCGCGGGGTGCGCACGGTGGTGATGGTGCGGCCGTCGCCGGAG
TTCTTCCTGTTGATGTTTGCCCTGTTCAAACTGGGTGCGGTGCCGGTGCTGGTCGACCCG
GGCATCGACAAACGCGCGCTGAAGCAGTGCCTGGATGAGGCGCAGCCGGAGGCTTTCATT
GGCATTCCGCTGGCACACGTGGCGCGGTTGGCGCTGCGCTGGGCGCGCTCGGCGACCCGC
CTGGTCACCGTCGGCCGCCGCCTCGGCTGGGGCGGGACCACGCTGGCCGCGCTGGAGCGC
GCCGGTGCCAACGGGGGGGCGATGCTGGCCGCCACCGATGGCGAGGACATGGCCGCGATC
CTGTTCACCAGCGGCTCCACCGGCGTGCCCAAGGGCGTGGTCTACCGCCACCGTCATTTC
GTCGGCCAGATCCAGCTGCTGGGCAGTGCCTTTGGCATGGAAGCGGGCGGGGTGGACCTG
CCGACCTTCCCGCCGTTTGCCTTGTTCGACCCGGCGCTGGGCCTGACCTCGGTGATCCCG
GACATGGACCCGACCCGGCCGGCACAGGCCGACCCGGCACGCCTGCATGACGCCATCCAG
CGCTTCGGCGTGACCCAGCTGTTTGGCTCGCCTGCGCTGATGCGGGTGCTGGCCCGTCAC
GGCCGGCCGTTGCCGACGGTGACCCGGGTGACCTCGGCCGGTGCGCCGGTGCCGCCGGAT
GTGGTCGCCACCATCCGCAGCCTGTTGCCGGCCGATGCGCAGTTCTGGACCCCGTACGGC
GCCACCGAGTGCCTGCCGGTGGCGGTGGTGGAAGGCCGCGAGCTGGAGCGCACCCGTGCC
GCCACCGAGGCCGGTGCCGGTACCTGCGTGGGCAGCGTGGTTGAACCGAATGAGGTGCGC
ATCATCGGCATCGATGACGGACCGTTGGCCGACTGGTCGCAGGCGCGCGTGCTGGCCACC
GGCGAGGTGGGCGAGATCACCGTGGCCGGACCGACCGCTACCGACAGTTACTTCAACCGC
CCGCAGGCCACCGCGGCGGCGAAGATCCGCGAGACCCTGGCCGATGGCAGCACGCGCGTG
GTGCACCGCATGGGTGACGTGGGCTACTTCGATGCGCAGGGCCGCCTGTGGTTCTGCGGC
CGCAAGACCCATCGCGTGGAAACCGCACGGGGGCCGCTGTACACCGAACAGGTGGAGCCG
GTGTTCAATACCGTGCCCGGGGTGGCGCGCACGGCGCTGGTCGGTGTTGGTCCGGCGGGA
GCACAGGTGCCGGTGCTGTGCGTGGAGCTGCAGCGTGGCCAGTCAGACAGCCCGGCGCTG
CAGGAGGCGCTGCGTGCGCACGCCGCAGCCCGCGCACCGGAAGCCGGCCTGCAGCACTTC
CTGGTGCATCCGGCGTTCCCCGTCGATATCCGTCACAACGCCAAGATCGGCCGCGAGAAG
CTCGCCGTCTGGGCCAGCGCCGAACTGGAGAAGCGCGCAAAGCTTCTCGAG

Sequence encoding for protein (once cloned into pET30b):

ATGAACCGACCCTGCAATATTGCCGCCCGCTTGCCTGAACTGGCGCGTGAACGGCCC
GACCAGATCGCCATCCGCTGCCCGGGCCGCCGTGGCGCCGGCAACGGCATGGCCGCCTAC
GACGTGACCCTGGATTACCGCCAGCTGGACGCCCGCAGCGACGCCATGGCGGCTGGCCTG
GCAGGCTACGGGATCGGTCGCGGGGTGCGCACGGTGGTGATGGTGCGGCCGTCGCCGGAG
TTCTTCCTGTTGATGTTTGCCCTGTTCAAACTGGGTGCGGTGCCGGTGCTGGTCGACCCG
GGCATCGACAAACGCGCGCTGAAGCAGTGCCTGGATGAGGCGCAGCCGGAGGCTTTCATT
GGCATTCCGCTGGCACACGTGGCGCGGTTGGCGCTGCGCTGGGCGCGCTCGGCGACCCGC
CTGGTCACCGTCGGCCGCCGCCTCGGCTGGGGCGGGACCACGCTGGCCGCGCTGGAGCGC
GCCGGTGCCAACGGGGGGGCGATGCTGGCCGCCACCGATGGCGAGGACATGGCCGCGATC
CTGTTCACCAGCGGCTCCACCGGCGTGCCCAAGGGCGTGGTCTACCGCCACCGTCATTTC
GTCGGCCAGATCCAGCTGCTGGGCAGTGCCTTTGGCATGGAAGCGGGCGGGGTGGACCTG
CCGACCTTCCCGCCGTTTGCCTTGTTCGACCCGGCGCTGGGCCTGACCTCGGTGATCCCG
GACATGGACCCGACCCGGCCGGCACAGGCCGACCCGGCACGCCTGCATGACGCCATCCAG
CGCTTCGGCGTGACCCAGCTGTTTGGCTCGCCTGCGCTGATGCGGGTGCTGGCCCGTCAC
GGCCGGCCGTTGCCGACGGTGACCCGGGTGACCTCGGCCGGTGCGCCGGTGCCGCCGGAT
GTGGTCGCCACCATCCGCAGCCTGTTGCCGGCCGATGCGCAGTTCTGGACCCCGTACGGC
GCCACCGAGTGCCTGCCGGTGGCGGTGGTGGAAGGCCGCGAGCTGGAGCGCACCCGTGCC

GCCACCGAGGCCGGTGCCGGTACCTGCGTGGGCAGCGTGGTTGAACCGAATGAGGTGCGC
ATCATCGGCATCGATGACGGACCGTTGGCCGACTGGTCGCAGGCGCGCGTGCTGGCCACC
GGCGAGGTGGGCGAGATCACCGTGGCCGGACCGACCGCTACCGACAGTTACTTCAACCGC
CCGCAGGCCACCGCGGCGGCGAAGATCCGCGAGACCCTGGCCGATGGCAGCACGCGCGTG
GTGCACCGCATGGGTGACGTGGGCTACTTCGATGCGCAGGGCCGCCTGTGGTTCTGCGGC
CGCAAGACCCATCGCGTGGAAACCGCACGGGGGCCGCTGTACACCGAACAGGTGGAGCCG
GTGTTCAATACCGTGCCCGGGGTGGCGCGCACGGCGCTGGTCGGTGTTGGTCCGGCGGGA
GCACAGGTGCCGGTGCTGTGCGTGGAGCTGCAGCGTGGCCAGTCAGACAGCCCGGCGCTG
CAGGAGGCGCTGCGTGCGCACGCCGCAGCCCGCGCACCGGAAGCCGGCCTGCAGCACTTC
CTGGTGCATCCGGCGTTCCCCGTCGATATCCGTCACAACGCCAAGATCGGCCGCGAGAAG
CTCGCCGTCTGGGCCAGCGCCGAACTGGAGAAGCGCGCAAAGCTTGCGGCCGCACTCGAG
CACCACCACCACCACCACTGA

MNRPCNIAARLPELARERPDQIAIRCPGRRGAGNGMAAYDVTLDYRQLDARSDAMAAGLA
GYGIGRGVRTVVMVRPSPEFFLLMFALFKLGAVPVLVDPGIDKRALKQCLDEAQPEAFIG
IPLAHVARLALRWARSATRLVTVGRRLGWGGTTLAALERAGANGGAMLAATDGEDMAAIL
FTSGSTGVPKGVVYRHRHFVGQIQLLGSAFGMEAGGVDLPTFPPFALFDPALGLTSVIPD
MDPTRPAQADPARLHDAIQRFGVTQLFGSPALMRVLARHGRPLPTVTRVTSAGAPVPPDV
VATIRSLLPADAQFWTPYGATECLPVAVVEGRELERTRAATEAGAGTCVGSVVEPNEVRI
IGIDDGPLADWSQARVLATGEVGEITVAGPTATDSYFNRPQATAAAKIRETLADGSTRVV
HRMGDVGYFDAQGRLWFCGRKTHRVETARGPLYTEQVEPVFNTVPGVARTALVGVGPAGA
QVPVLCVELQRGQSDSPALQEALRAHAAARAPEAGLQHFLVHPAFPVDIRHNAKIGREKL
AVWASAELEKRAKLAAALEHHHHHH
Molecular Weight: 60.18 kDa

OlnD (OleD) rev: 1005 bp, 992 bp cut

CATATGAAGATCCTGGTCACCGGTGGTGGTGGTTTCCTTGGCCAGGCGCTGTGCCGTGGG
CTGGTCGAACGTGGCCACCAGGTGCTGGCGTTCAACCGCAGCCACTACCCGGAACTGCAG
GCCATGGGCGTGGGCCAGATCCGCGGCGATCTGGCCGATGCGCAGGCGGTGCTGCATGCA
GTGGCCGGCGTCGATGCGGTATTTCACAACGGCGCCAAGGCCGGCGCCTGGGGCAGCTAT
GACAGCTACCACCAGGCCAACGTGGTCGGCACCGACAACGTGATCGCCGCCTGCCGCGCG
CACGGCATCAGCCGGCTGGTCTACACCTCCACGCCCAGCGTGACCCACCGCGCGACCCAC
CCGGTGGAAGGCCTGGGGGCTGATGAGGTGCCGTACGGCGAGGACTTCCAGGCGCCGTAT
GCGGCGACCAAGGCGATTGCCGAACAGCGCGTGCTGGCCGCCAACGATGCCACGCTGGCG
ACGGTGGCGCTGCGCCCGCGCCTGATCTGGGGCCCGGGCGATCAGCAACTGGTGCCACGC
CTGGCAGAACGTGCGCGGCAGGGCCGCCTGCGTCTGGTGGGCGATGGCAACAACAAGGTG
GATACCACTTACATCGACAACGCCGCGCTCGCGCATTTCCTCGCCTTCGAAGCGTTGGCA
CCGGGTGCCGCGTGTGCGGGCAAGGCCTACTTCATTTCCAACGGCGAACCGCTGCCCATG
CGCGAGCTGGTCAACAAGCTGCTGGCCGCGGTTGGCGCGCCGACGGTGGACAAGGCGATC
AGTTTCAAGACCGCCTATCGCATCGGTGCGGTCTGCGAGCGGCTGTGGCCGCTGCTGCGC
CTGCGTGGCGAGCCGCCGCTGACCCGCTTCCTGGCCGAGCAGCTGTGCACGCCGCACTGG
TACAGCATGGAGCCGGCGCGCCGTGACTTCGGCTACGTGCCGCAGGTCAGCATTGAAGAA
GGGCTGCGCAGGCTGAAGGCTTCATCTGCCGCAAAGCTTCTCGAG

Sequence encoding for protein (once cloned into pET30b):

ATGAAGATCCTGGTCACCGGTGGTGGTGGTTTCCTTGGCCAGGCGCTGTGCCGTGGG
CTGGTCGAACGTGGCCACCAGGTGCTGGCGTTCAACCGCAGCCACTACCCGGAACTGCAG
GCCATGGGCGTGGGCCAGATCCGCGGCGATCTGGCCGATGCGCAGGCGGTGCTGCATGCA
GTGGCCGGCGTCGATGCGGTATTTCACAACGGCGCCAAGGCCGGCGCCTGGGGCAGCTAT
GACAGCTACCACCAGGCCAACGTGGTCGGCACCGACAACGTGATCGCCGCCTGCCGCGCG
CACGGCATCAGCCGGCTGGTCTACACCTCCACGCCCAGCGTGACCCACCGCGCGACCCAC
CCGGTGGAAGGCCTGGGGGCTGATGAGGTGCCGTACGGCGAGGACTTCCAGGCGCCGTAT
GCGGCGACCAAGGCGATTGCCGAACAGCGCGTGCTGGCCGCCAACGATGCCACGCTGGCG
ACGGTGGCGCTGCGCCCGCGCCTGATCTGGGGCCCGGGCGATCAGCAACTGGTGCCACGC
CTGGCAGAACGTGCGCGGCAGGGCCGCCTGCGTCTGGTGGGCGATGGCAACAACAAGGTG
GATACCACTTACATCGACAACGCCGCGCTCGCGCATTTCCTCGCCTTCGAAGCGTTGGCA
CCGGGTGCCGCGTGTGCGGGCAAGGCCTACTTCATTTCCAACGGCGAACCGCTGCCCATG
CGCGAGCTGGTCAACAAGCTGCTGGCCGCGGTTGGCGCGCCGACGGTGGACAAGGCGATC
AGTTTCAAGACCGCCTATCGCATCGGTGCGGTCTGCGAGCGGCTGTGGCCGCTGCTGCGC
CTGCGTGGCGAGCCGCCGCTGACCCGCTTCCTGGCCGAGCAGCTGTGCACGCCGCACTGG
TACAGCATGGAGCCGGCGCGCCGTGACTTCGGCTACGTGCCGCAGGTCAGCATTGAAGAA
GGGCTGCGCAGGCTGAAGGCTTCATCTGCCGCAAAGCTTGCGGCCGCACTCGAGCACCAC
CACCACCACCACTGA

MKILVTGGGGFLGQALCRGLVERGHQVLAFNRSHYPELQAMGVGQIRGDLADAQAVLHAV
AGVDAVFHNGAKAGAWGSYDSYHQANVVGTDNVIAACRAHGISRLVYTSTPSVTHRATHP
VEGLGADEVPYGEDFQAPYAATKAIAEQRVLAANDATLATVALRPRLIWGPGDQQLVPRL
AERARQGRLRLVGDGNNKVDTTYIDNAALAHFLAFEALAPGAACAGKAYFISNGEPLPMR
ELVNKLLAAVGAPTVDKAISFKTAYRIGAVCERLWPLLRLRGEPPLTRFLAEQLCTPHWY
SMEPARRDFGYVPQVSIEEGLRRLKASSAAKLAAALEHHHHHH
Molecular Weight: 36.94 kDa

Appendix III: Detailed methods and scheme of chemical synthesis and analysis (provided by Jack E. Richman)

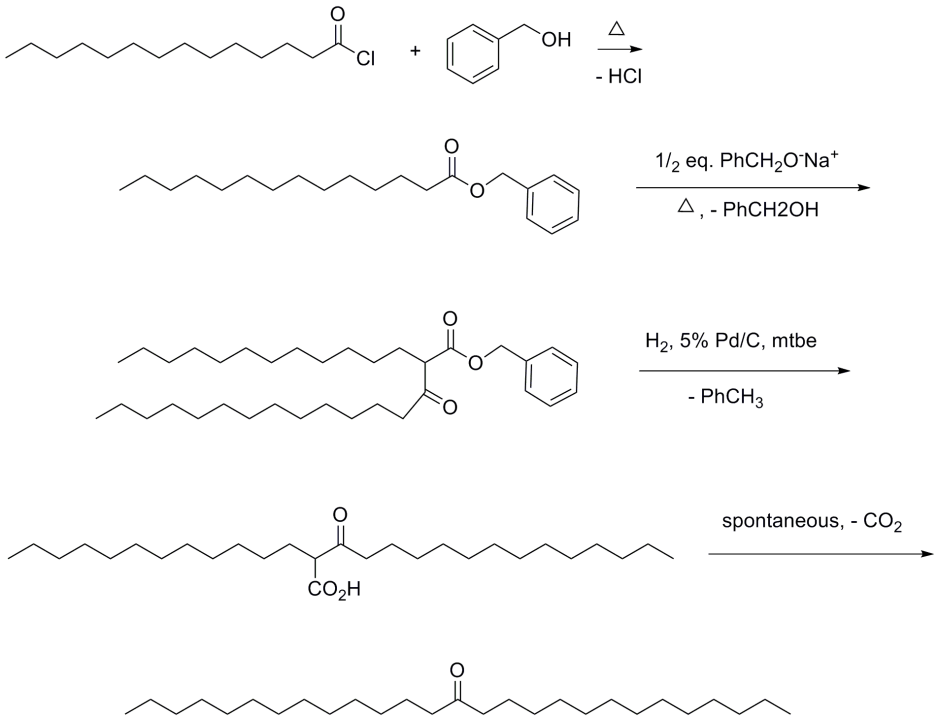


Figure III-1. Scheme showing chemical synthesis of 2-myristoyl myristic acid (2MMA).

General. ^{1}H and ^{13}C NMR were obtained at 400 and 100 MHz, respectively.

Benzyl myristate. Benzyl myristate was prepared essentially by the method of Shonle and Row[1] by the neat (solventless) reaction of myristoyl chloride and slight excess benzyl alcohol. After evacuation to remove volatiles, NMR analysis indicated that no myristoyl

chloride and only benzyl myristate, the excess benzyl alcohol and a trace of benzyl chloride remained. This product was used without further purification.

Benzyl 2-Myristoylmyristate (B2MM). A distilling flask was charged with 6 mL of benzyl alcohol and 1 mL of this was distilled (97°C/20 mmHg) to dry the residue and remove volatiles. Sodium hydride powder (0.12 g, 5 µmol, pentane washed oil dispersion, filtered and N_2 dried) added to the cooled dry benzyl alcohol under N_2 evolved hydrogen but fully dissolved only after warming and stirring. The sodium benzyl alkoxide/benzyl alcohol solution was added under N_2 to a reaction flask charged with 3.2 g (10 µmol) of benzyl myristate, a stir bar and steam-heated (100°C) reflux condenser with vacuum takeoff to a cold trap. The reaction mixture was evacuated (0.1 mmHg) and slowly heated to remove excess benzyl alcohol. Finally, the reaction was heated over two hours from 120 to 132°C/0.1 mmHg leaving a semi-solid residue. Neutralization (HOAc) and NMR analysis indicated that the reaction had proceeded to about 70% conversion. Bulb-to-bulb distillation of 90% of the crude product mixture to 150°C/0.05 mmHg left 1.35 g (*ca*. 70% yield) of reasonably pure B2MM, which slowly crystallized to a waxy solid (mp 34.5-35.8°C). ^{1}H NMR integrations of the aryl and high-field signals are in 10% excess of theory, suggesting that this product is approximately 90% pure B2MM. There is no evidence for tautomeric (enolic) content, which may account in part for the integral disparities.

^{1}H NMR ($CDCl_3$): 7.3-7.4 (m, 5H, aryl), 5.17 and 5.14 (prochiral benzylic AB, J=12.3 Hz, 2H), 3.46 (t, 1H, CH, J=7.4), 2.46 and 2.40 (each dt, 2H, prochiral CH_2, J=17.2 and 7.2 Hz), 1.84 (m, 2H, α-CH_2), 1.52 (tt, 2H, β-CH_2), 1.34-1.16 (m, 10H, CH_2) and 0.88 ppm (t, 6H, CH_3).

^{13}C NMR ($CDCl_3$, assigned as 77.0 ppm): 205.1, 169.7, 135.4, 128.5, 128.3, 128.2, 66.8, 59.1, 41.8, 31.9, 29.63, 29.60, 29.55, 29.45, 29.4, 29.30, 29.28, 29.26, 28.9, 28.2, 27.4, 23.4, 22.6 and 14.0 ppm. (Twenty four of 33 possible carbon singlets are resolved.)

Hydrogenolysis of Benzyl 2-Myristoylmyristate: 2-Dodecyl-3-ketohexadecanoic Acid (2-Myristoylmyristic Acid, 2MMA) and Methyl 2-Dodecyl-3-ketohexadecanoate (Methyl 2-Myristoylmyristate, M2MM). A shielded apparatus with gas inlet (top) and outlet that could be positioned to near the bottom (for complete gas purging) was charged with 17 mg B2MM, 3.4 mg 5% Pd/C catalyst suspended in 3 mL methyl *t*-butyl ether (mtbe). The apparatus was thoroughly purged with N_2, then with H_2 (industrial grade) and maintained under H_2 at atmospheric pressure.[2] After 2.5 h, stirring was interrupted, and the settled mixture was sampled. The sample was immediately treated (shield) with slight excess of ethereal diazomethane (yellow persisted) and concentrated. Analyses by NMR and GC showed that the hydrogenolysis was 97% completed. After stirring an additional hour, the reaction mixture (N_2 purged) was filtered and the filtrate and washes were immediately cooled (dry ice), and stored in -80°C freezer. A sample of this product, warmed to -18°C, treated with diazomethane, and analyzed by GC, showed a 5:2 mixture of M2MM and the product of decarboxylation of the β-keto acid, 14-heptacosanone, plus only minor impurities. After two weeks storage at -80°C, gc analyses showed that the supernate had enriched to 8:1 β-keto acid (→ M2MM) and ketone. This indicates that partial crystallization of the ketone had occurred.

NMR of M2MM ($CDCl_3$): essentially identical patterns (slightly different shifts) to B2MM absent benzyl signals plus methyl ester singlet at 3.72 ppm.

^{1}H NMR of 14-heptacosanone ($CDCl_3$): 2.38 (t, 4H, α-CH_2, J= 7.4), 1.56 (m, 4H, β-CH_2), 1.32-1.22 (m, 40H), 0.88 (t, 6H, CH_3).

References and notes

[1] Shonle, H. A., and Row, P. Q. (1921) *J. Am. Chem. Soc.* ***43***, 361-5.

[2] Caution: all oxygen must be removed before hydrogen is introduced: H_2/O_2 or air within the explosive limits plus catalyst will detonate.

Appendix IV: OleA genes (as ordered from DNA 2.0)

***Congregibacter litoralis* KT71**

GGTACCGGGCCCCATATGTCTGGTAACGCTAAATTCACTTTAAACGATACTGCTATCGTTTCT
GTTACTGCTCACCACGCTCCAGAAGTTGTTACTTCTGCTTCTTTAGATGATCGTATCATGCAC
ACTTACGAACGTTTAGGTACTCAACCAGGTTTATTAGAATCTTTAGCTGGTATCTCTGAACGT
CGTTGGTGGCCAGAAGGTCACACTTTCACTGAAGCTGCTGCTGAAGCTGGTCGTAAAGCTATG
GCTGCTGCTAACATCAAACCAGAACAAGTTGGTTTATTAATCGATACTTCTGTTTCTCGTGAT
CGTTTAGAACCATCTTCTGCTGTTACTGTTCACCACTTATTAGATTTACCATCTTCTTGTTTA
AACTTCGATATGGCTAACGCTTGTTTAGGTTTCATGAACGCTATGCAAGTTGCTGGTATGATG
TTAGATTCTCGTCAAATCGATTTCGCTTTAATCGTTGATGGTGAAGGTTCTCGTCAACCACAA
GAAAAAACTTTAGAACGTTTAGCTTCTGATGAAGCTACTGTTGCTGATTTATTCGCTGATTTC
GCTACTTTAACTTTAGGTTCTGGTGCTGCTGGTATGGTTTTAGGTCGTCACTCTGAAAACGCT
GGTTCTCACAAAATCATCGGTGGTATCAACCGTGCTAACACTTCTCACCACAAATTATGTGTT
GGTACTTTAGATCAAATGCGTACTGATACTGCTGCTTTATTAGAAGCTGGTTTAGATGTTTCT
GAACGTGCTTGGGCTAACGCTGAAGAATACGGTTGGTTAGATATGGATCGTTACGTTATCCAC
CAAATCTCTTCTGTTCACACTTCTATGTTATGTGAACGTTTAGGTATCGATGTTGATAAAGTT
CCATTAACTTACCCAAAATTAGGTAACACTGGTCCAGCTGCTGTTCCATTAACTTTAGCTCAA
GAATCTGAATCTTTAAACCCAGGTGATCGTGTTTTATGTTTAGGTATGGGTTCTGGTATCAAC
GCTATGGCTTTAGAAATCGCTTGGTAGGGATCCACTAGTCCCGGG

MSGNAKFTLNDTAIVSVTAHHAPEVVTSASLDDRIMHTYERLGTQPGLLESLAGI
SERRWWPEGHTFTEAAAEAGRKAMAAANIKPEQVGLLIDTSVSRDRLEPSSAVTVHHLLD
LPSSCLNFDMANACLGFMNAMQVAGMMLDSRQIDFALIVDGEGSRQPQEKTLERLASDEA
TVADLFADFATLTLGSGAAGMVLGRHSENAGSHKIIGGINRANTSHHKLCVGTLDQMRTD
TAALLEAGLDVSERAWANAEEYGWLDMDRYVIHQISSVHTSMLCERLGIDVDKVPLTYPK
LGNTGPAAVPLTLAQESESLNPGDRVLCLGMGSGINAMALEIAW

Blast = ZP_01103251.1

***Xanthomonas campestris* pv. campestris str. ATCC 33913**

GGTACCGGGCCCCATATGTTATTCCAAAACGTTTCTATCGCTGGTTTAGCTCACATCGATGCT
CCACACACTTTAACTTCTAAAGAAATCAACGAACGTTTACAACCAACTTACGATCGTTTAGGT
ATCAAAACTGATGTTTTAGGTGATGTTGCTGGTATCCACGCTCGTCGTTTATGGGATCAAGAT
GTTCAAGCTTCTGATGCTGCTACTCAAGCTGCTCGTAAAGCTTTAATCGATGCTAACATCGGT
ATCGAAAAAATCGGTTTATTAATCAACACTTCTGTTTCTCGTGATTACTTAGAACCATCTACT
GCTTCTATCGTTTCTGGTAACTTAGGTGTTTCTGATCACTGTATGACTTTCGATGTTGCTAAC
GCTTGTTTAGCTTTCATCAACGGTATGGATATCGCTGCTCGTATGTTAGAACGTGGTGAAATC
GATTACGCTTTAGTTGTTGATGGTGAAACTGCTAACTTAGTTTACGAAAAAACTTTAGAACGT
ATGACTTCTCCAGATGTTACTGAAGAAGAATTCCGTAACGAATTAGCTGCTTTAACTTTAGGT
TGTGGTGCTGCTGCTATGGTTATGGCTCGTTCTGAATTAGTTCCAGATGCTCCACGTTACAAA
GGTGGTGTTACTCGTTCTGCTACTGAATGGAACAAATTATGTCGTGGTAACTTAGATCGTATG

GTTACTGATACTCGTTTATTATTAATCGAAGGTATCAAATTAGCTCAAAAAACTTTCGTTGCT
GCTAAACAAGTTTTAGGTTGGGCTGTTGAAGAATTAGATCAATTCGTTATCCACCAAGTTTCT
CGTCCACACACTGCTGCTTTCGTTAAATCTTTCGGTATCGATCCAGCTAAAGTTATGACTATC
TTCGGTGAACACGGTAACATCGGTCCAGCTTCTGTTCCAATCGTTTTATCTAAATTAAAAGAA
TTAGGTCGTTTAAAAAAAGGTGATCGTATCGCTTTATTAGGTATCGGTTCTGGTTTAAACTGT
TCTATGGCTGAAGTTGTTTGGTAGGGATCCACTAGTCCCGGG

MLFQNVSIAGLAHIDAPHTLTSKEINERLQPTYDRLGIKTDVLGDVAGIHARRLWDQDVQ
ASDAATQAARKALIDANIGIEKIGLLINTSVSRDYLEPSTASIVSGNLGVSDHCMTFDVA
NACLAFINGMDIAARMLERGEIDYALVVDGETANLVYEKTLERMTSPDVTEEEFRNELAA
LTLGCGAAAMVMARSELVPDAPRYKGGVTRSATEWNKLCRGNLDRMVTDTRLLLIEGIKL
AQKTFVAAKQVLGWAVEELDQFVIHQVSRPHTAAFVKSFGIDPAKVMTIFGEHGNIGPAS
VPIVLSKLKELGRLKKGDRIALLGIGSGLNCSMAEVVW

Blast = NP_635607.1

***Xylella fastidiosa* 9a5c**

GGTACCGGGCCCCATATGTTATTCAACAACGTTTCTATCGCTGGTTTAGCTCACATCGATGCT
CCATGTACTTTAACTTCTCAAGAAATCAACGCTCGTTTACAACCAATGTTAGAACGTATCGGT
ATCAAATCTGATGTTTTCGCTGATATCGTTGGTATCAACGCTCGTCGTTTATGGAACACTAAC
GTTCAAACTTCTGATGTTGCTACTATGGCTGCTCGTAAAGCTTTACAAGATGCTGGTGTTGCT
GTTGATCGTATCGGTTTAGTTGTTAACACTTCTGTTTCTCGTGATTACTTAGAACCATCTACT
GCTTCTATCGTTTCTGGTAACTTAGGTGTTGGTGAACAATGTATCGCTTTCGATGTTGCTAAC
GCTTGTTTAGCTTTCTTAAACGGTATGGATATCGCTGGTCAAATGTTAGAACGTGGTGATATC
GATTACGCTTTAGTTGTTAACGCTGAAACTGCTAACCGTGTTTACGAAAAAACTTTAGAACGT
ATGTCTGCTCCAGGTGTTACTGAACAAGAATTCCGTGAAGAAATGGCTGCTTTAACTTTAGGT
TGTGGTGCTGTTGCTATGGTTTTAGCTCGTACTGCTTTAGTTCCAGATGCTCCACAATACAAA
GGTGGTGTTACTCGTTCTGCTACTGAATGGAACAAATTATGTTGTGGTAACTTAGATCGTATG
GTTACTGATACTCGTTTAATGTTAATCGAAGGTATCAAATTAGCTAAAAAAACTTTCGTTGTT
GCTAAACAAGTTTTAGGTTGGGCTGTTGAAGAATTAGATCAATTCGTTATCCACCAAGTTTCT
CGTCCACACACTGAAGCTTTCATCAAATCTTTCGGTATCGATCCAGCTAAAGTTATGACTATC
TTCCGTGAATACGGTAACATCGGTCCAGCTTCTGTTCCAATCGTTTTATCTAAATTAAAAGAA
TTAGGTCGTTTAAAAAAAGGTGATCGTATCGCTTTATTAGGTATCGGTTCTGGTTTAAACTGT
TCTATGGCTGAAGTTGTTTGGTAGGGATCCACTAGTCCCGGG

MLFNNVSIAGLAHIDAPCTLTSQEINARLQPMLERIGIKSDVFADIVGINARRLWNTNVQ
TSDVATMAARKALQDAGVAVDRIGLVVNTSVSRDYLEPSTASIVSGNLGVGEQCIAFDVA
NACLAFLNGMDIAGQMLERGDIDYALVVNAETANRVYEKTLERMSAPGVTEQEFREEMAA
LTLGCGAVAMVLARTALVPDAPQYKGGVTRSATEWNKLCCGNLDRMVTDTRLMLIEGIKL
AKKTFVVAKQVLGWAVEELDQFVIHQVSRPHTEAFIKSFGIDPAKVMTIFREYGNIGPAS
VPIVLSKLKELGRLKKGDRIALLGIGSGLNCSMAEVVW

Blast = NP_299252.1

***Plesiocystis pacifica* SIR-1**

GGTACCGGGCCCCATATGCGTTTCGCTAACGTTTCTATCTGTTCTGTTGCTCACGTTGATGCT
CCATACCGTGTTTCTTCTACTGATTTAGAAAACCGTTTAGCTGCTCCAATGCAACGTTTAGGT
TTACCACCAGGTATCTTAGAAACTTTAACTGGTATCAAAGCTCGTCGTATGTGGCCAGCTTCT
GTTTCTCCATCTGATGCTGCTACTTTAGCTGCTCGTCGTGCTATCGCTGAATCTGGTGTTGAT
CCAGAACGTATCGGTGTTTTAATCTCTACTTCTGTTTGTCGTGATTTCGTTGAACCATCTACT
GCTTGTTTAGTTCACGGTAAATTAGGTTTACCACCAACTTGTTTAAACTTCGATGTTGGTAAC
GCTTGTTTAGGTTTCATCAACGGTATGGATATCATCGGTAACATGATCGAACGTGGTCAATTA
GATTACGGTATCGTTGTTGATGGTGAAGATTCTCGTTACGTTATCGATAAAACTATCGAACGT
TTATCTGCTCCAGATTCTACTCGTGAAGATTTCTGGTCTAACTTCGCTACTTTAACTTTAGGT
GGTACTGCTGCTGCTATGGTTTTAGCTCGTACTGATTTAGCTCAAGCTTTAGCTGAAAAACGT
GCTGAAGGTGGTTACTCTCACCAATTCTTAGGTTCTGTTATCGTTGCTGCTACTCAACACTCT
GGTTTATGTCGTGGTCAAGTTGATCGTATGGAAACTGATTCTGCTGAATTATTAACTGCTGGT
TTACGTGTTGCTAAAGAAGCTTGGCGTGCTGCTCAACGTGAATTCGGTTGGACTCCAGGTGCT
TTAGATGAATGTGTTATCCACCAAGTTTCTCGTACTCACACTGATAAATTCTGTGAAACTTTC
GAATTAGATCCAGCTAAATTATTAGCTACTTACCCAGAATTCGGTAACGTTGGTCCAGCTGGT
GTTCCAATGGTTTTATCTAAAGCTGCTTCTTCTGGTCGTTTAGGTCGTGGTGATCGTGTTGGT
TTAATGGGTATCGGTTCTGGTTTAAACTGTGCTATGGCTGAAGTTGTTTGGTAGGGATCCACT
AGTCCCGGG

MRFANVSICSVAHVDAPYRVSSTDLENRLAAPMQRLGLPPGILETLTGIKARRMWPASVS
PSDAATLAARRAIAESGVDPERIGVLISTSVCRDFVEPSTACLVHGKLGLPPTCLNFDVG
NACLGFINGMDIIGNMIERGQLDYGIVVDGEDSRYVIDKTIERLSAPDSTREDFWSNFAT
LTLGGTAAAMVLARTDLAQALAEKRAEGGYSHQFLGSVIVAATQHSGLCRGQVDRMETDS
AELLTAGLRVAKEAWRAAQREFGWTPGALDECVIHQVSRTHTDKFCETFELDPAKLLATY
PEFGNVGPAGVPMVLSKAASSGRLGRGDRVGLMGIGSGLNCAMAEVVW

Blast= ZP_01906524.1

***gamma proteobacterium* NOR5-3**

GGTACCGGGCCCCATATGCACTTCGAATCTGTTGTTATCTTATCTTTAGCTGCTGCTGATGCT
CCAATCTCTTTAACTTCTAAAGAAATCTCTCAACGTTTAAAACCAACTATGGATCGTTTAGGT
GTTCGTGAAAACTTATTAGAAGAAATCTCTGGTATCGCTTCTCGTCGTATCTGGAACCCAGAA
ACTTCTCCATCTGATGCTGCTACTTTAGCTGCTGAAAAAGCTATCCAAGATTCTGGTATCGAT
CGTTCTCGTATCGGTGTTATCATCTCTACTTCTGTTTCTCGTGATTTCTTAGAACCATCTGCT
GCTTGTATGGTTCACGGTAACTTAGGTTTAGCTTCTGATTGTTTAAACTTCGATGTTGCTAAC
GCTTGTTTAGGTTTCTTAAACGGTATGGATATCGCTGCTCGTATGATCGAACGTGAAGAATTA
GATTACGCTTTAGTTGTTGCTGGTGAATCTTCTCGTCCATTAATCGAAGCTACTACTGAACGT
TTATTAGATCAAGATGTTGGTGCTGCTCAATTCCGTGAAGAATTCGCTTCTTTAACTTTAGGT
TCTGGTGCTGCTGCTATGATCATGACTCGTCGTGAATTAGCTCCAGGTGGTCACACTTACCGT
GGTTCTGTTACTCGTTCTGCTACTCAATTCAACCGTTTATGTCAAGGTAACATGGATCGTATG
CGTACTGATACTGGTATGTTATTATCTGCTGGTTTAGAATTAGCTGCTCAAACTTTCGAAGCT
TCTTGTTCTACTTTAGATTGGTCTGTTGATGAAATGGATCAATTCATCATCCACCAAGTTTCT

AAAGTTCACACTGAATCTTTAGTTAAAACTTTAGGTTTAAACCCAGATAAAGTTCACGCTATC
TACCCACACATGGGTAACATCGGTCCAGCTTCTGTTCCAATCGTTTTAGCTAAAGTTGAAGAA
GCTGGTAAATTAAAAAAAGGTGATCGTATCGCTTTATTAGGTATCGGTTCTGGTTTAAACTGT
GCTATGGCTGAAGTTGTTTGGTAGGGATCCACTAGTCCCGGG

MHFESVVILSLAAADAPISLTSKEISQRLKPTMDRLGVRENLLEEISGIASRRIWNPETS
PSDAATLAAEKAIQDSGIDRSRIGVIISTSVSRDFLEPSAACMVHGNLGLASDCLNFDVA
NACLGFLNGMDIAARMIEREELDYALVVAGESSRPLIEATTERLLDQDVGAAQFREEFAS
LTLGSGAAAMIMTRRELAPGGHTYRGSVTRSATQFNRLCQGNMDRMRTDTGMLLSAGLEL
AAQTFEASCSTLDWSVDEMDQFIIHQVSKVHTESLVKTLGLNPDKVHAIYPHMGNIGPAS
VPIVLAKVEEAGKLKKGDRIALLGIGSGLNCAMAEVVW

Blast=ZP_05127044.1

Appendix V: OleD genes ordered through DNA 2.0

>D-Xcamp_optEc1
ACCATGGGCAGCTCGCATCATCATCATCATCACAGCAGCGGTCTGGTGCCGCGTGGTAGCCATATG
AAAATCCTGGTTACCGGTGGTGGTGGTTTTCTGGGCCAAGCCCTGTGTCGTGGTTTGGTCGCACGT
GGTCACGAGGTTGTCAGCTTTCAGCGCGGTGACTACCCGGTCCTGCACACGTTGGGCGTGGGCCAA
ATCCGTGGTGACCTGGCAGACCCTCAGGCGGTCCGTCACGCTTTGGCAGGTATTGATGCCGTTTTT
CACAATGCCGCCAAAGCGGGTGCATGGGGCAGCTATGATTCTTATCATCAAGCGAATGTCGTTGGT
ACTCAAAATGTCCTGGATGCGTGTCGCGCGAACGGCGTCCCGCGTTTGATCTACACCTCCACCCCG
TCGGTGACGCATCGTGCGACGAATCCGGTTGAGGGTTTGGGTGCGGATGAAGTTCCGTACGGTGAG
GACTTGCGTGCGCCGTACGCTGCGACCAAGGCTATCGCGGAGCGTGCGGTCCTGGCAGCCAACGAC
GCGCAATTGGCAACCGTTGCGCTGCGCCCACGCCTGATTTGGGGTCCGGGTGACAATCACCTGCTG
CCGCGTCTGGCAGCGCGTGCCCGTGCCGGTCGCCTGCGTATGGTCGGTGATGGCAGCAACCTGGTG
GACTCTACCTATATCGATAATGCAGCCCAGGCCCACTTCGATGCGTTTGCGCACCTGGCGCCTGGT
GCAGCTTGCGCGGGTAAGGCATACTTCATTAGCAACGGCGAACCGCTGCCGATGCGTGAGCTGCTG
AACCGTCTGCTGGCAGCGGTGGATGCCCCAGCGGTGACCCGTAGCCTGAGCTTCAAAACCGCGTAC
CGCATCGGCGCTGTGTGCGAAACCCTGTGGCCGCTGCTGCGCCTGCCGGGTGAGGTTCCGCTGACG
CGTTTCTTGGTTGAACAGCTGTGCACTCCGCACTGGTACAGCATGGAACCAGCACGTCGCGACTTC
GGCTATGTTCCGCAGATTTCTATCGAGGAAGGCCTGCAGCGTTTGCGTTCCAGCAGCAGCCGCGAC
ATTAGCATTACGCGCTGACTCGAGGATCC

MGSSHHHHHHSSGLVPRGSHMKILVTGGGGFLGQALCRGLVARGHEVVSFQRGDYPVLH
TLGVGQIRGDLADPQAVRHALAGIDAVFHNAAKAGAWGSYDSYHQANVVGTQNVLDACRA
NGVPRLIYTSTPSVTHRATNPVEGLGADEVPYGEDLRAPYAATKAIAERAVLAANDAQLA
TVALRPRLIWGPGDNHLLPRLAARARAGRLRMVGDGSNLVDSTYIDNAAQAHFDAFAHLA
PGAACAGKAYFISNGEPLPMRELLNRLLAAVDAPAVTRSLSFKTAYRIGAVCETLWPLLR
LPGEVPLTRFLVEQLCTPHWYSMEPARRDFGYVPQISIEEGLQRLRSSSSRDISITR-LED
MW = 38.3 kDa
Blast = NP_635614.1

>D-Xfast_optEc1
ACCATGGGTAGCTCGCATCATCATCATCATCACAGCAGCGGCCTGGTGCCTCGCGGCAGCCATATG
CGTATTCTGGTCACGGGCGGTAGCGGTTTCCTGGGTGAGGCCCTGTGTCGTGGCCTGTTGAAACGC
GGTTACCAGGTGGTGAGCTTTCAGCGTAGCCACTATCAAGCATTGCAGGCCCTGGGTGTTGTCCAA
ATTTGTGGCGATTTGTCCGATTTCCACGCCGTCCGTCACGCGGTTCGTGGTGTCGACGCGGTTTTT
CACAACGCGGCAAAAGTTGGTGCGTGGGGCTCTTATACGTCCTATCACCAGATTAACGTTATCGGC
ACGCAACATGTCCTGGACGCGTGCCGTGCAGAAAACATCAATAAACTGGTGTATACCAGCACCCCT
AGCGTTATTCATCGCAGCAATTACCCGGTCGAAGGTCTGGACGCGGACCAGGTTCCGTACAGCAAC
GCTGTGAAAGTGCCGTACGCAGCCACTAAGGCTATGGCGGAACAAGCAGTTCTGGCTGCAAATAGC
GTAGATCTGACCACCGTCGCACTGCGTCCGCGTATGATCTGGGGTCCGGGCGATCCGCATCTGATG
CCGCGCTTGGTCGCGCGTGCGCGTGCCGGTCGTCTGCGTCTGATTGGTGACGGTCGTAACCTGGTC
GACAGCACCTACATCGATAATGCAGCGCAGGCCCACTTCGACGCTTTTGAGCACCTGATGCCGGGT
GCGGCATGTGCCGGCAAGGCTTATTTCATTTCCAATGGTGAGCCTCTGCAGATGCGCGAGCTGATC
AACAAGTTGCTGGCAACGACGAATGCACCGCCGGTGACCCAAAGCCTGAGCTTCAAGACCGGCTAC
TGCATTGGTGCCTTCTGCGAGATGCTGTGGAGCCTGCTGCCGCTGCCGGGTGAACCGCTGTTGACC
CGCTTTCTGGTCGAGCAAATGAGCACCCCGCACTGGTATTCCATCGAACCGGCTAAACGTGATTTT
GGCTACGTTCCGCGTGTCAGCATTGAAGAGGGTCTGGTGCGTCTGCTGAGCGAGACTCGCGTGACC
TGCTGACTCGAGGATCC

Protein

MRILVTGGSGFLGEALCRGLLKRGYQVVSFQRSHYQALQ
ALGVVQICGDLSDFHAVRHAVRGVDAVFHNAAKVGAWGSYTSYHQINVIGTQHVLDACRA
ENINKLVYTSTPSVIHRSNYPVEGLDADQVPYSNAVKVPYAATKAMAEQAVLAANSVDLT
TVALRPRMIWGPGDPHLMPRLVARARAGRLRLIGDGRNLVDSTYIDNAAQAHFDAFEHLM
PGAACAGKAYFISNGEPLQMRELINKLLATTNAPPVTQSLSFKTGYCIGAFCEMLWSLLP
LPGEPLLTRFLVEQMSTPHWYSIEPAKRDFGYVPRVSIEEGLVRLLSETRVTC
Blast = NP_779252.1

>D-Gprot_optEc1
ACCATGGGTAGCTCGCATCATCATCATCATCACAGCTCGGGTCTGGTTCCGCGTGGTAGCCATATG
AAGATCTTGGTCACCGGTGGCGGTGGTTTCCTGGGTCAAGAAATCTGCCACATGCTGCTGGCGCAA
GGCGACGAGCCGGTCGCCTTCCAACGCGGCGAGGCACGCGCGCTGGCCCAAGCGGGTATCGAAGTC
CGTCGTGGTGATATTGGTCGTCTGCAGGACGTGCTGGCAGCGGCGGAGGGTTGTGAGGCGGTGATC
CACACGGCGGGTAAGGCCGGTGCATGGGGTGACGCGCAACTGTACCGTGCGGTCAACGTAAGCGGT
ACCCAAAACGTTCTGCAAGCGTGCGAAGCGCTGGGTATCCAACGCCTGGTTTTTACCAGCTCCCCG
AGCGTGGCGCATTGCGGTGGCGACATCGCCGGTGGCGATGAAAGCCTGCCGTACCCGCGTCATTAT
GCGGCACCGTATCCGCAGACTAAAGCTGCGGCAGAACAGCTGGTTATGGCGGCATCCGGCAGCGGT
CTGAATACCGTGTCCCTGCGTCCGCACCTGGTGTGGGGTCCTGGTGACAATCAGCTGTTGCCGCGT
TTGGTCGAGCGTGCGCGTCGTGGTACCCTGCGCCTGCCGGGTGCGGATAAGTTGATCGATGCAACT
TACATTTACAATGCCGCTCGCGCACACCTGCTGGCATTGGCAGCACTGGACAATAACGAAGCGTGT
CACGGTAAAACCTATTTCATCAGCAACGGTGAACCGTGGCCGCAAGCCAAGATTATTGCCGCACTG
CTGAACGCAGTGGGCGTGAACGCTGATATCAAGCCGATTGCGGCAGGTGCTGCAAAACTGGCGGGT
ATTTTGGCGGAGTCTTGGTGGCGCTTGAGCCAGCGCGACGATGAGCCTCCGGTGACCCGCTGGAGC
GCGGAACAACTGGCGACGGCGCACTGGTACGACATTAGCGCGGCACGTAAGGATTTGGGTTACGAA
CCTGTTATCAGCATGGCAGAGGGTCTGAAACGTCTGGCCCAGAGCGCTGAGAATGCCCGTCTGGCT
GACGATATCCAGACGAAATGACTCGAGGATCC

Protein
MKILVTGGGGFLGQEICHMLLAQGDEPVAFQRGEARALA
QAGIEVRRGDIGRLQDVLAAAEGCEAVIHTAGKAGAWGDAQLYRAVNVSGTQNVLQACEA
LGIQRLVFTSSPSVAHCGGDIAGGDESLPYPRHYAAPYPQTKAAAEQLVMAASGSGLNTV
SLRPHLVWGPGDNQLLPRLVERARRGTLRLPGADKLIDATYIYNAARAHLLALAALDNNE
ACHGKTYFISNGEPWPQAKIIAALLNAVGVNADIKPIAAGAAKLAGILAESWWRLSQRDD
EPPVTRWSAEQLATAHWYDISAARKDLGYEPVISMAEGLKRLAQSAENARLADDIQTK
Blast= ZP_05127041.1

>D-Chloro_optEc1
ACCATGGGTAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTCCCGCGTGGCAGCCATATG
ATCGCGCTGGTCACCGGCGGTAACGGTTTCGTTGGTCGTTACATTGTTGAGCAACTGCTGGCCCGT
GGTGATCACGTTCGTGTGATTGGTCGTGGTGCGTATCCAGAGCTGCAGTCCCTGGGCGCAGAAACC
TACCAGGCAGATCTGACGTTGCCTGAATCTGCGCCGGTGCTGGCACGTGCAATGCGTGGTGTCACG
ACCGTGTTTCACGTCGCGGCAAAAGCAGGTCTGTGGGGTTCGTACGATGACTTTTACCGTGCGAAC
GTATCCGCGACCCAGCGTGTTGTTAAAGCGGCCATCCGTGCCGGCGTTCCGAAACTGGTGTATACC
AGCACCCCGAGCGTTGTTATTGGTCATGAAGATATCCATGGTGGCGATGAACACCTGCCGTATCCT
CGCCGTTACCTGGCCCCGTACCCACACACGAAGGCGATCGCAGAGCGCTATGTGCTGGCGCAAACG
GACATCGCGACGGTCAGCCTGCGTCCGCACCTGATTTGGGGTCCGCGCGACCCGCACATTCTGCCG
CGCTTGCTGCGTCGTGCGCGTCGCCGTATGTTGTTCCAAATCGGTGACGGCACGAACCTGGTCGAC
GTCTGCTATGTGGAAAATGCGGCAACCGCGCATATCCAGGCGGCCAGCGCCCTGAATGAACGTAGC
CCGCTGCGTGGCCGTGCGTACTTCATTGGTCAGGAGCGTCCGGTGAATTTGTGGCAATTCATTGGT
GAGATCCTGAAGGCTGCGAATTGTCCGCCTGTTCGTGGCCGCATTTCCGCGAGCGCTGCCACCATT
CTGGCTACCGGCCTGGAACTGCTGTATACTATCTTGCGTCTGCCGGGTGAGCCGCCACTGACTCGC

CTGATGGTCCATGAGCTGTCTCACTCCCACTGGTTCAGCCACGCTGCGGCCGAGCGTGACTTTGGC
TACACGCCTCGTATTAGCATCGAGGAAGGTCTGGAACGTACCTTTGCACTGACCGGCTCCCAACCG
TGACTCGAGGATCC

Protein
MIALVTGGNGFVGRYIVEQLLARGDHVRVIGRGAYPELQ
SLGAETYQADLTLPESAPVLARAMRGVTTVFHVAAKAGLWGSYDDFYRANVSATQRVVKA
AIRAGVPKLVYTSTPSVVIGHEDIHGGDEHLPYPRRYLAPYPHTKAIAERYVLAQTDIAT
VSLRPHLIWGPRDPHILPRLLRRARRRMLFQIGDGTNLVDVCYVENAATAHIQAASALNE
RSPLRGRAYFIGQERPVNLWQFIGEILKAANCPPVRGRISASAATILATGLELLYTILRL
PGEPPLTRLMVHELSHSHWFSHAAAERDFGYTPRISIEEGLERTFALTGSQP
Blast = Caur_3530

Appendix VI: OleD Solubility Test

OleD appeared to be falling out of solution over time after the soluble fraction was obtained from crude cell lysate. A solubility test was performed using the OptiSol™ Protein Solubility Screening Kit provided by Dilyx Biotechnologies, LLC (Seattle, Washington). The kit contained included a 96 well plate with a variety of solutions to understand pH, salt, and additive effects on the solubility of the protein of interest. After the solubility was assessed following the test kit manual and Biorad protein concentration assay, the solutions with the highest solubility were tested for alkene production using myristoyl-CoA as described in section 6.2.7. The buffer system that resulted in the highest solubility of OleD is ranked 1 in Sol. Rank in Table VI-1, and the alkene and ketone detected, in relative amounts, are listed for each of the buffer systems with the highest solubility of OleD.

	Buffer			**Additive**			**Products**	
Sol. Rank	Name	Conc. (mM)	pH	NAME	Conc	unit	Alkene	Ketone
1	CHES	50	9.5	Tween 20	1	% (w/v)	0	tiny
2	Imidazole	50	8	Tween 20	1	% (w/v)	0	0
3	MOPS	50	7	Tween 20	1	% (w/v)	XXX	X
4	CHES	50	9.5	NaCl	150	mM	0	X
5	HEPES	50	7.5	NaCl	500	mM	0	tiny
6	CAPS	50	10	TMAO	500	mM	XXX	XXX
7	Tris	50	8.5	TMAO	500	mM	0	tiny
8	CAPS	50	10	NaCl	500	mM	0	tiny
9	CAPS	100	10				0	tiny
10	Tris	50	8.5	NaCl	500	mM	XXX	X
11	CHES	100	9.5				XX	XXX
12	CHES	100	9				0	tiny
13	Imidazole	50	8	NaCl	150	mM	0	XX
14	ADA	50	6.5	TMAO	500	mM	0	tiny
15	Tris	100	8.5				X	X
16	HEPES	50	7.5	TMAO	500	mM	XX	XX
17	ADA	50	6.5	NaCl	500	mM	tiny	tiny

Table VI-1. OptiSol Solubility Screen top solubility results (Data provided by J. S. Erickson)

Appendix VII: OleC Preliminary Enzyme Assay Data

The role of OleC is hypothesized to be different than most members of the superfamily. From sequence alignments, the smaller C terminal domain does not align to other members of the superfamily. The product of the OleC reaction likely does not have a CoA moiety ligated to the carboxyl, as the CoAs have been hydrolyzed by the first enzyme in the pathway, OleA. A half reaction of an AMP ligation to the substrate is proposed in Figure VII-1, and was preliminarily investigated in this chapter.

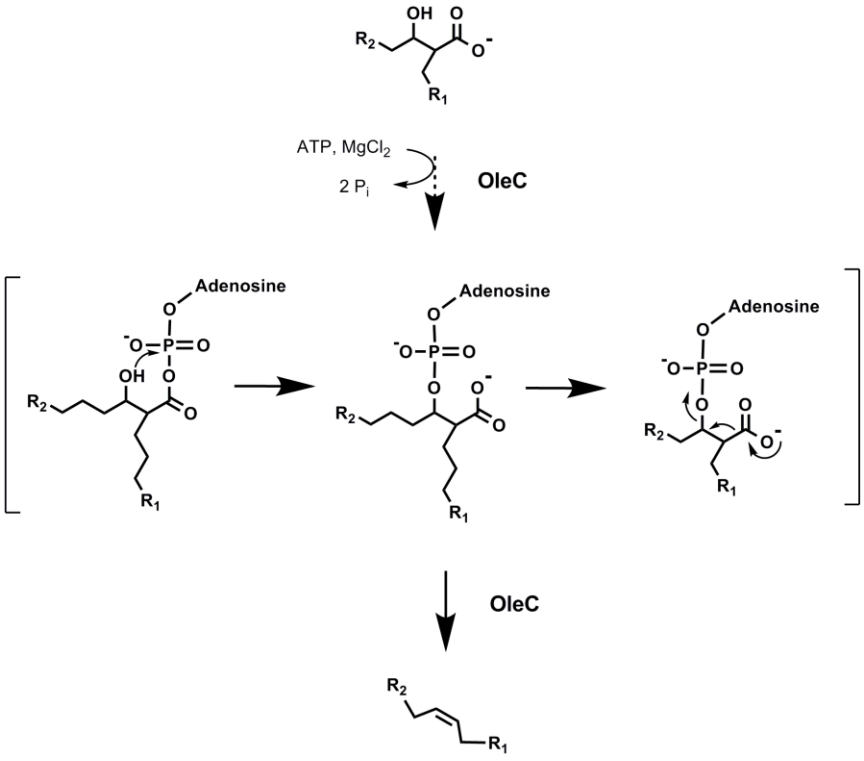


Figure VII-1. Proposed reaction mechanism of OleC in alkene biosynthesis

Experimental

Activity Assays

Phosphate (PO_4) was determined using a modified malachite green assay (58, 84) with reagents made up as described below.The Malachite green reagent contained: 36.8 ml H_2O, 6.5 ml concentrated HCl, 16.5 mg (malachite green solution is clear and yellow). In making a homogenous ammonium molybdate solution, I added 0.65 mg ammonium molybdate $(NH_4)Mo_7O_{24}\bullet 4H_2O$ (Fluka, ≥99% purity), let it sit overnight to dissolve ammonium molybdate, and filtered the solution through Whatman #2. The solution was not stable on introduction of contaminants; care must be taken to avoid contamination with phosphate by rinsing glassware with deionized water and using clean pipette tips. Protein assays were set up in 96-well microtiter plates with gentle shaking at 30°C, with wells sampled at a variety of time points over several days. Standard curves were run with each new assay to be sure of the detectable range. The assay was diluted approximately 1:20 with water into a total of 100 µl in cuvettes and 400 µl of dye reagent was added. Cuvettes were quickly mixed and the absorbance determined at 620 nm. Using standard curve and dilutions, concentration of PO_4 released in the protein assay could be calculated. Protein assays were set up in 100 or 500 µl volumes with 10 mM ATP (dissolved in 200 mMTris pH 7.4) and 20 mM $MgCl_2$ (dissolved in 200 mMTris pH 7.4). When second substrate was used, levels ranged from 1 to 10 mM depending on solubility.

High pressure liquid chromatography (HPLC) (Hewlett Packard Series 1100) was used with the following method to separate ATP, ADP, AMP and the predicted product from OleC reaction with β-hydroxy octanoic acid (43, 70): 0-10 min 95% 50 mM ammonium acetate pH 5.0 (A): 5% methanol (B), 10-45 min ramp to 80% A: 20% B, 45-65 ramp to 100% B, 65-70 hold at 100% B, 70-80 min ramp to 95% A: 5% B and equilibrate using Adsorbosphere (Alltech) C18 reverse phase column (250 by 4.6 mm) with 5µ particle size. When looking exclusively for ATP, ADP, and AMP, an abbreviated program was used: 0-5 min 95% A: 5% B, 5-40 min ramp to 80% A: 20% B, 40-55 ramp back to 95% A: 5% B. Assays were performed in 500 µl total with 200 mM Tris, pH 7.4, 0.2 mg OleC from -80°C, 10 mM ATP, ADP, or AMP, and 20 mM $MgCl_2$, and incubated at 30°C gently shaking for 2 days. The additional substrate, β-hydroxyoctanoic acid, at 1 or 10 mM was added when analyzing for the putative OleC product derived from the hydroxy acid.

Results and Discussion

Activity Assays

AMP ligase/synthase proteins can be assayed by the detection of pyrophosphate (84). It was quickly determined that OleC was actively releasing phosphate. Initially experiments were conducted with pyrophosphatase as has been reported in the literature for similar enzymes (84), but the phosphate released was shown to be the same with or without pyrophosphatase. This led to additional experiments using

malachite green to detect phosphate and HPLC to detect organic products from OleC-catalyzed reactions. By HPLC, ATP, ADP, and AMP all appeared to be present in the enzyme assay samples (Fig.VII-4). Using malachite green and ADP as cofactor rather than ATP, phosphate release was still detected. OleC appears to be able to hydrolyze ATP and ADP to produce inorganic phosphate rather than pyrophosphate, which is different than well-studied ATP ligase/synthases (102). Perhaps OleC is able to utilize either cofactor to generate the AMP for ligation, which will be tested using ADP in place of ATP for alkene formation with OleA and OleD.

OleC is hypothesized to catalyze a reaction with an intermediate in the alkene biosynthesis pathway. Without having all enzymes expressed and active in a soluble form at the time of these studies, only commercially available substrates potentially similar in nature to the physiological substrate were tested. The tested substrates included short and long chain β-hydroxy acids, such as palmitic, octanoic, and butanoic acids, as well as an α-hydroxy acid, lactic acid, to begin to explore the active site geometry. The α-hydroxy acid, lactic acid, was found to inhibit the release of phosphate by malachite green detection, Fig VII-2. It was also found that the longer chain substrates increased the phosphate release. This led to the investigation into whether the hydroxyl group increased activity. The longer chain substrate was then investigated in a comparison to palmitic acid and palmitoyl-CoA. Palmitic acid and palmitoyl-CoA had no effect on activity over just OleC with ATP alone, but the addition of β-hydroxypalmitic acid resulted in an increased rate and total phosphate release in comparison, Fig VII-3. This supports the hypothesis that a hydroxyl group is

potentially involved in the chemistry of the AMP ligation. It was not determined whether the increase was due to an increase in binding affinity or whether the AMP is transferred to the hydroxyl at the β carbon or to the carboxyl group. Once intermediates can be generated enzymatically, this will be investigated further.

OleC product formation was also investigated by high pressure liquid chromatography to identify enzyme products. At the time of these studies the hypothesis of the role of OleC, without having OleD purified and active, was that OleC was activating the OleA product and could potentially be identified as a long chain molecule with AMP ligated to a hydroxyl group. From the assay with enzyme and ATP, it can be seen that both ADP and AMP were being detected (Fig VII-4). Assays with 10 mM β-hydroxy octanoic acid showed a new peak eluting at 55 min, 60% methanol (Fig VII-5). This compound with increased hydrophobicity was detected by UV absorbance at 259 nm. LC-MS investigation of this peak in the Center for Mass Spectrometry and Proteomics, using the same HPLC method and column, resulted in no identifiable compound that may be a direct result of the ligation of AMP and β-hydroxy octanoic acid or β-hydroxy palmitic acid. Background was quite high in the spectra and with the acids not having any absorbance on their own it was difficult to follow the compounds. Investigation into enzyme assays with the radiolabelled substrate, [1-^{14}C] myristic acid, followed this study. The radiolabelled HPLC experiments did not result in OleC intermediate elucidation, but did lead to the method development in Chapter 6 of the OleA intermediate identification. The lack of identification of a product in which AMP ligated to the fatty acid could be due to the

fact that the AMP ligates to the β hydroxyl group, which this compound does not have. Alternatively, the hydroxyl group is used for orientation in the binding site so the fatty acid was unable to orient properly in order to ligate the AMP to the carboxyl. With the recent identification and synthesis of the OleA intermediate, further investigations can be done by mass spectrometry of a potential OleC intermediate.

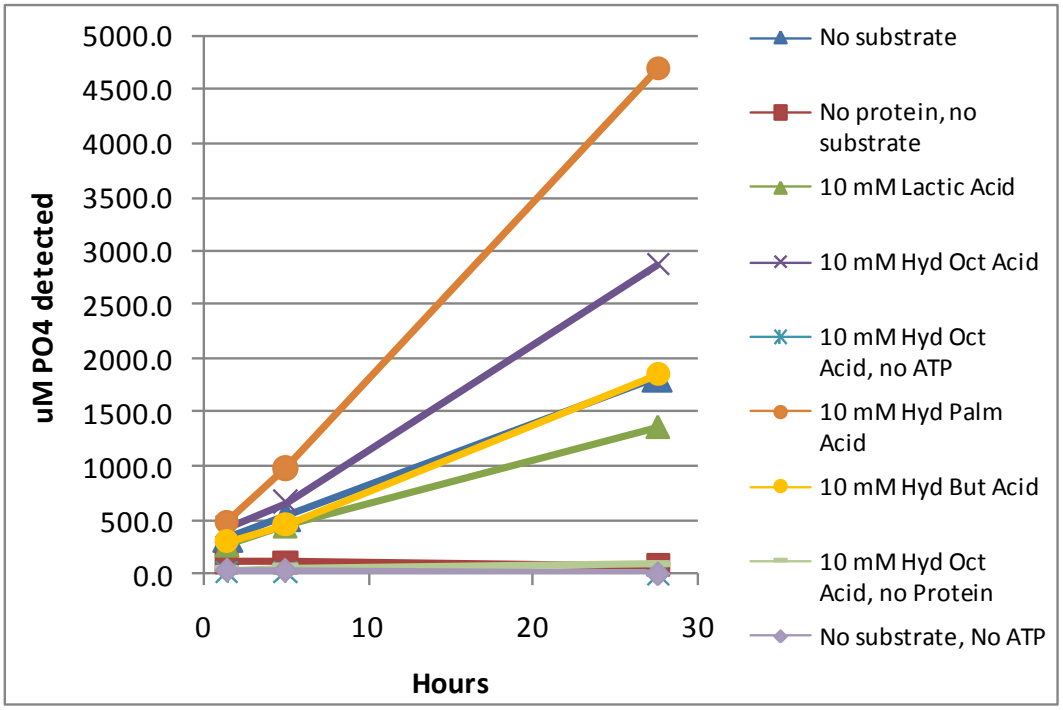


Figure VII-2. Substrate Effect on OleC Activity

β-hydoxy palmitic acid (Hyd Palm Acid), β-hydroxy octanoic acid (Hyd Oct Acid), β-hydroxy butyric acid (Hyd But Acid), and Lactic Acid were used to determine the effect, if any, of chain length on the release of phosphate from ATP by OleC. No phosphate release was detected in the following controls: "no OleC and no substrate", "hyd oct acid without OleC", "no substrate, no ATP", "Hyd Oct Acid without ATP". Lactic acid inhibited the activity of the OleC and ATP without additional substrate. Hydroxy butanoic acid had no effect above OleC and ATP. Increasing the chain length to C8 increased phosphate release, and increasing chain length to C16 resulted in the highest activity of phosphate released from ATP by OleC in this study.

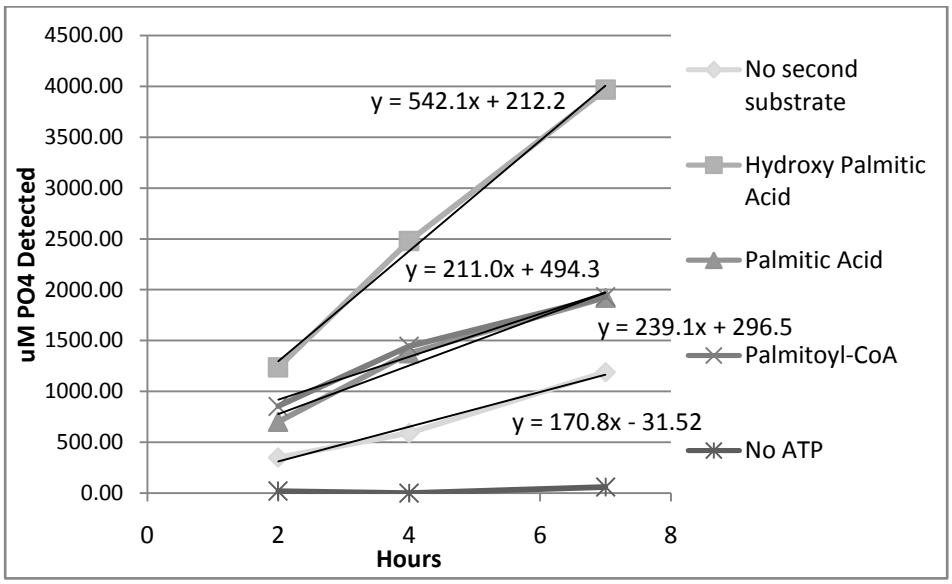


Figure VII-3. Investigating the Nature of the Substrate.

β-hydroxy palmitic acid addition resulted in an increased rate of phosphate release from ATP with OleC in comparison to substrates, palmitic acid and palmitoyl-CoA.

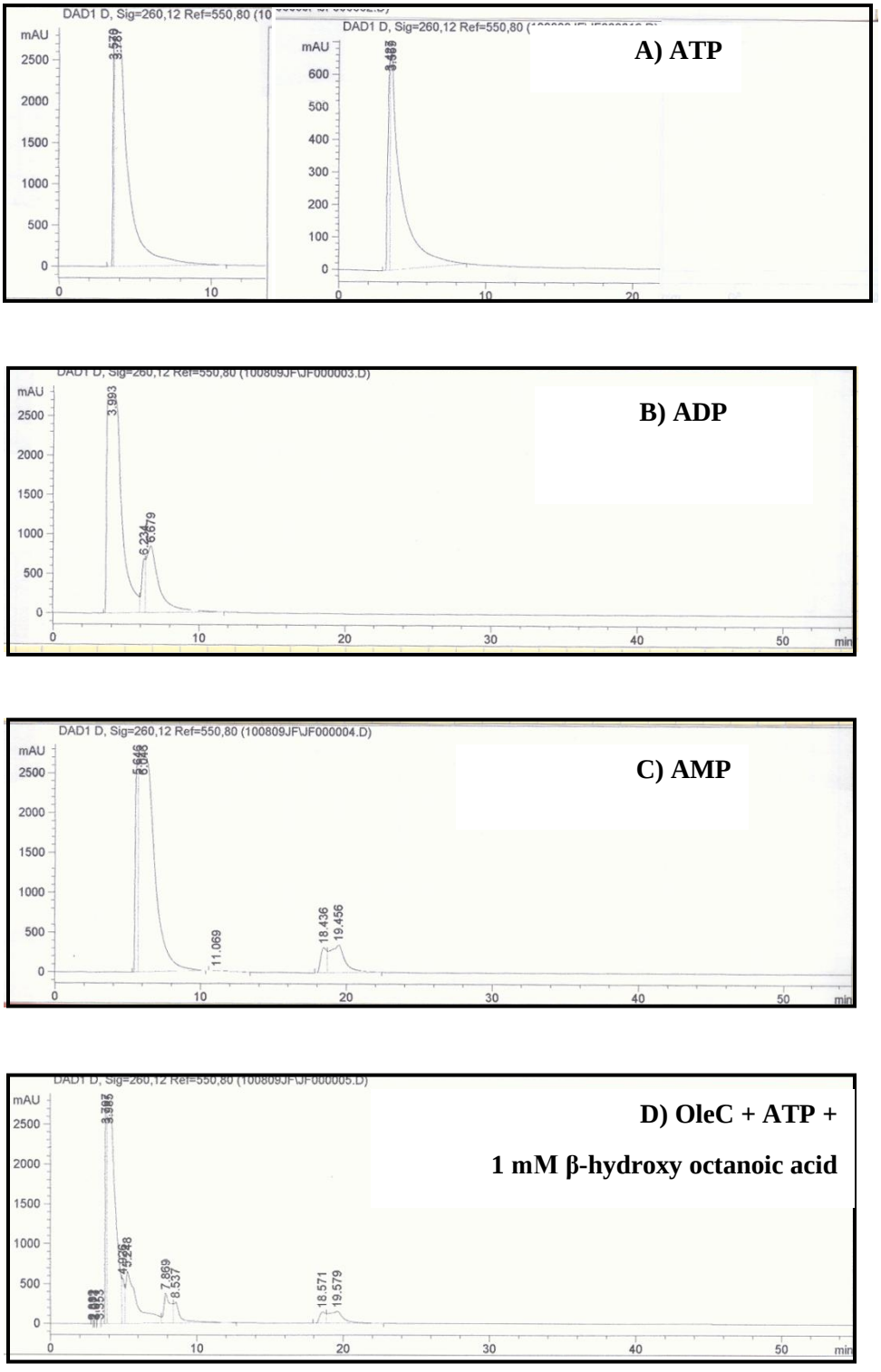


Figure VII-4. HPLC results of standards (A-C) and OleC reaction containing 10 mM ATP and 1 mM β-hydroxy octanoic acid, incubated for 2 days (D)

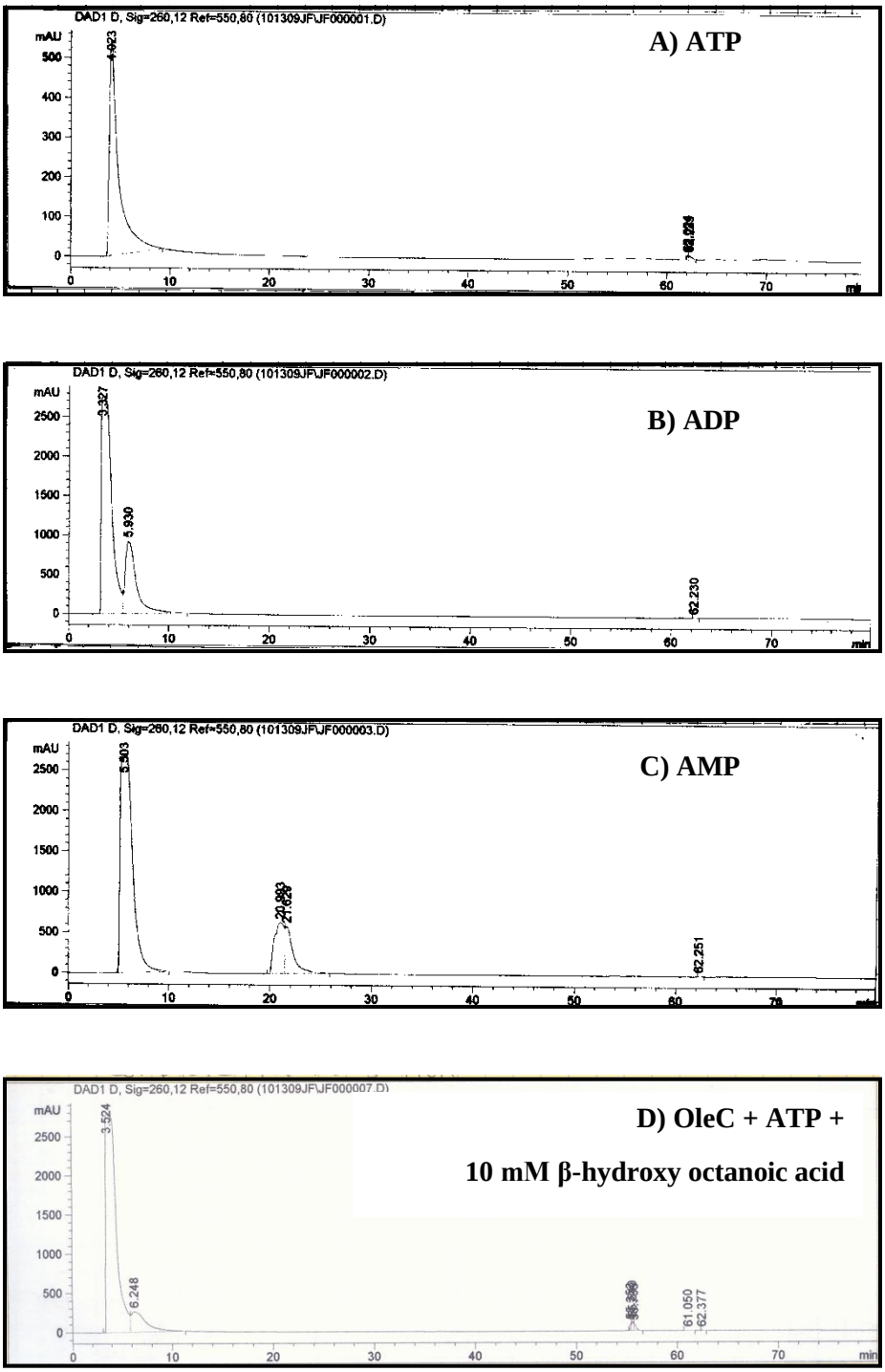


Figure VII-5. HPLC results, revised program, standards (A-C) and OleC reaction with 10 mM ATP and 10 mM β-hydroxy octanoic acid, incubated 2 days (D).