Hydrocarbon Biosynthesis by Bacteria: Genes and Hydrocarbon Products

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

David John Sukovich

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Lawrence P. Wackett

December 2010

Acknowledgements

I want to thank my advisor, Dr. Larry Wackett, for the opportunity to work in his laboratory, his support and patience, and for allowing me to head in research directions that did not always look promising. I would like to acknowledge my committee members for their guidance. Dr. Mike Sadowsky was enthusiastic and always available to answer questions and give support whenever I needed it. Dr. Jeffrey Gralnick provided valuable direction when I worked with *Shewanella*. Dr. Gary Dunny and Dr. Romas Kazlauskas were always willing to help me brainstorm whenever I asked. I would like to thank other investigators for providing important strains and genetic material: Dr. F. Srienc for providing *R. eutropha* and the pBBR vector, Dr. T.D. Schmidt for providing the TAV2 strain, and T.A. Burnes for providing the *X. campestris* strain used in this study. I would finally like to thank F. Schendel for allowing me to use the BioTechnology Institute's facilities to grow many of the organisms that needed special environments. My work was funded by the following sources: by grant LG-B13 from the Initiative for Renewable Energy and the Environment (IREE); grant DE-AR0000007 from the Advanced Research Project Agency – Energy (ARPA-E); a Dennis P. Watson Fellowship; and a Travel Award from the MICaB Program.

Thanks goes to Janice Frias for her development and optimization of the Bligh and Dyer Technique we used in almost every study as well as her support and advise; Jennifer Seffernick for her bioinformatics work and sound advise; Jack Richman for his synthesis of chemical standards and chemistry knowledge; Brandon Wood for his help with screening various chemical additives for potential use in ketone selection; Stephan

Cameron for her analysis of the *Vibrio* carboxysome; and other current and past Wackett Lab members Erik Reynolds, Naomi Kreamer, Alex Galaitsis, Neissa Pinzon, Kelly Aukema, Tony Dodge, Adam Erickson, Jasmine Erickson, Deep Golwalla, Amit Aggarwal, and Michael Turnbull. From the Gralnick Laboratory I have to thank Lee Robinson, Kristopher Hunt, and Christopher Flynn who were a great help with the mutational work. From the Sadowsky Laboratory I want to thank Dr. Matt Hamilton for his help with southern hybridization, entertainment, and sarcastic humor. From the Bond Laboratory I want to thank Misha Mehta for her help with anaerobic bacterial growth and her friendship. From the Srienc Laboratory I need to thank Dan Rouse for his help with constructing the *R. eutropha* strains. I would also like to thank Kerry Sokol, Janet Rollefson, Clint Remarcik, Jessica Eichmiller, Andrew Vale, Elaine Black, Dan Norat, Charlie Sawdey, Kris Lofgren, and last but not even close to being least Katie Ballering, for their advise and/or entertainment over the last five years.

Finally, I would like to thank Dr. David N. Cornfield, who was a mentor, role model, and friend prior to my joining the MICaB program.

Dedication

I would like to dedicate this dissertation to my family: Randy, Kay, and Ellen for their unwavering support throughout this process.

Abstract

Various publications have reported that microorganisms have the ability to produce hydrocarbons. One of these organisms, *Vibrio furnissii* M1, was reported to produce n-alkanes. Genomic analysis and biochemical studies revealed that the findings reported by Park *et al.* were not reproducible in our laboratory. Other heterotrophic bacteria were shown to produce hydrocarbons though. One of these organisms, *Shewanella oneidensis* MR-1, was found to produce 3,6,9,12,15,19,22,25,28-hentriacontanonaene. Hydrocarbon production in *S. oneidensis* was dependent upon the polyunsaturated fatty acid synthesis pathway and a relationship between temperature and hydrocarbon production was identified. Genomic analysis and mutation studies found that hydrocarbon production was dependent upon a gene cluster, designated *oleA*, *oleB*, *oleC*, and *oleD*. The OleA protein condenses two fatty acyl CoA chains in a head-to-head manner to produce a compound that, if the OleBCD proteins are not present, is spontaneously decarboxylated to a ketone. Homologs to the *oleABCD* genes were found in all heterotrophic bacteria reported to produce hydrocarbons. Searches of genomic databases found that 1.9% of all sequenced genomes have *oleABCD* gene homologs. These bacteria include members from the γ- and δ-*Proteobacteria*, *Actinobacteria*, *Verrucomicrobia*, *Planctomycetes*, and *Chloroflexi* Phyla. Bacteria containing the *oleABCD* homologs not previously characterized for hydrocarbons were obtained and tested for polyolefin production. It was found that if the genes were present, bacteria produced alkenes. A correlation between OleA amino acid sequence and product formation was also discovered. When different OleA proteins were expressed heterologously in non-native bacterial

backgrounds, the bacteria hosts were able to produce ketones. Ketone production could be increased using alternative plasmid promoters and regulation sequences. Preliminary experiments investigated strategies for cloning and expressing an *oleA* gene in cyanobacteria heterologously. Also, exploratory experiments were conducted to determine if ketone production by *Shewanella* might enhance cell growth when antibiotics, detergents, or other potentially inhibitory chemicals were added to the growth media.

Table of Contents

Chapter 3: Structure, function, and insights into the biosynthesis of a head-to-head hydrocarbon in *Shewanella oneidensis* strain MR-1

Chapter 4: Widespread Head-to-Head Hydrocarbon Biosynthesis in Bacteria and Role of OleA

List of Tables

List of Figures

Chapter 1: Introduction

1.1 The issues with traditional hydrocarbons and oil

Hydrocarbons are compounds that contain only carbon and hydrogen molecules. These compounds include alkanes (Figure 1.1A), alkenes (Figure 1.1B), and alkynes (Figure 1.1C) that can be linear, branched chained, or contain cyclic structures. Most hydrocarbons are aliphatic compounds, but there are important classes of aromatic and mixed aliphatic/aromatic hydrocarbons (86-88).

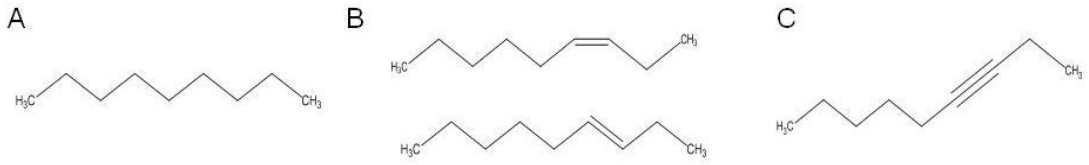


Figure 1.1 Structures of representative hydrocarbons. (A) *n*-nonane, a representative alkane, (B) *cis*- and *trans*-nona-3-ene, representative alkanes, (C) nona -3-yne, a representative alkyne .

Human society uses hydrocarbons derived from petroleum as a source for most of our fuels and chemicals (28, 119). Fuel hydrocarbons are used for heating, electricity generation, and as motor fuels. Since the fuels used are distillate fractions from petroleum, they are complex mixtures. The mixtures include some "ideal" fuels. For example, hexadecane is a desired diesel fuel, and isooctane is an excellent fuel for spark ignition engines (Figure 1.2). Not all alkanes are excellent fuels. For example, n-heptanes causes engine knocking (Figure 1.2) (119). Hydrocarbons are not only a

major source of energy, but are also refined to make plastics, tar, solvents, and other indispensible products (28).

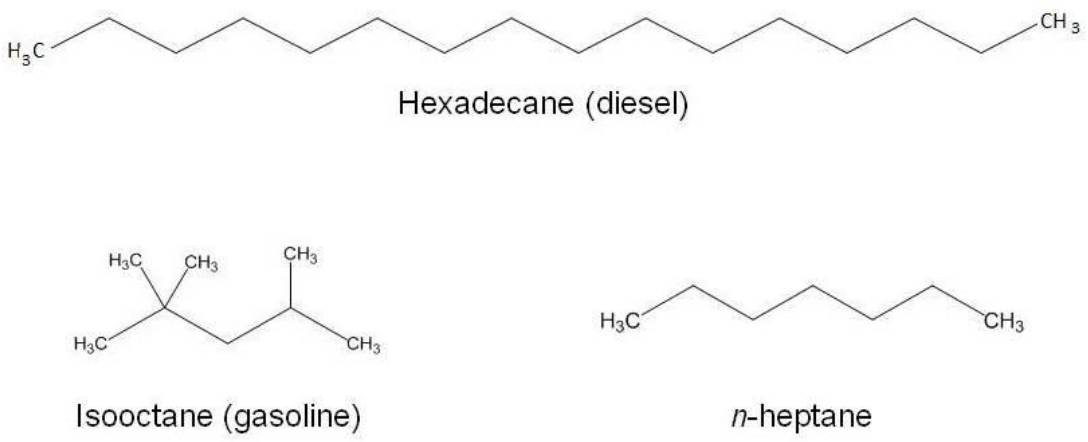


Figure 1.2 Hydrocarbons associated with standard fuels. Hexadecane is the "gold standard" for diesel, while isooctane is the "gold standard" for gasoline. Other hydrocarbons, such as n-heptane, cause the typical "knocking" of engines.

Traditionally hydrocarbons are obtained from oil and petroleum (156). These resources are originally derived from ancient biomass. Over eons, organic matter is subjected to pressure and is heated in an anoxic environment. This results in the diagenesis of the carbon-rich materials to more stable hydrocarbon molecules. Another process for petroleum formation has been proposed involving temperatures exceeding 300^{o}C. This second process is relatively fast in geological time, is often associated with geothermal vents, and results in a higher fraction of polynuclear aromatic hydrocarbons (140). Oil and petroleum are mined from various places around the world at a rate that exceeds the earth's ability to produce more. Over millions of years, the oil and

petroleum fields would be restocked (156), but for a growing world society dependent upon a steady flow of cheap energy and raw materials for stability, this long recovery period is not ideal.

While the abiotic processes forming petroleum have been our traditional source of hydrocarbons, it is known that insects (108, 155), plants (99, 118, 138), cyanobacteria (135), and various heterotrophic bacteria (2, 51, 147, 149, 159) biosynthesize hydrocarbons. This is currently being investigated for the potential to generate hydrocarbons from biomass or carbon dioxide.

1.2 Hydrocarbon biosynthesis in biological systems not investigated in this thesis: insect-, plant-, and cyanobacterial-derived hydrocarbons

It is well documented that hydrocarbons can be produced by living organisms. Various insects produce branched hydrocarbons (Figure 1.3A) as pheromones and for communication (58), while plants produce hydrocarbons (Figure 1.3B) in the form of isoprenoids. These compounds act as protectants against herbivores, as antimicrobial agents (68), induce thermotolerance (118), and to protect against reactive oxygen species (99). Cyanobacteria also produce long-chain hydrocarbons (Figure 1.3C) (45, 135), though their purpose is not well understood.

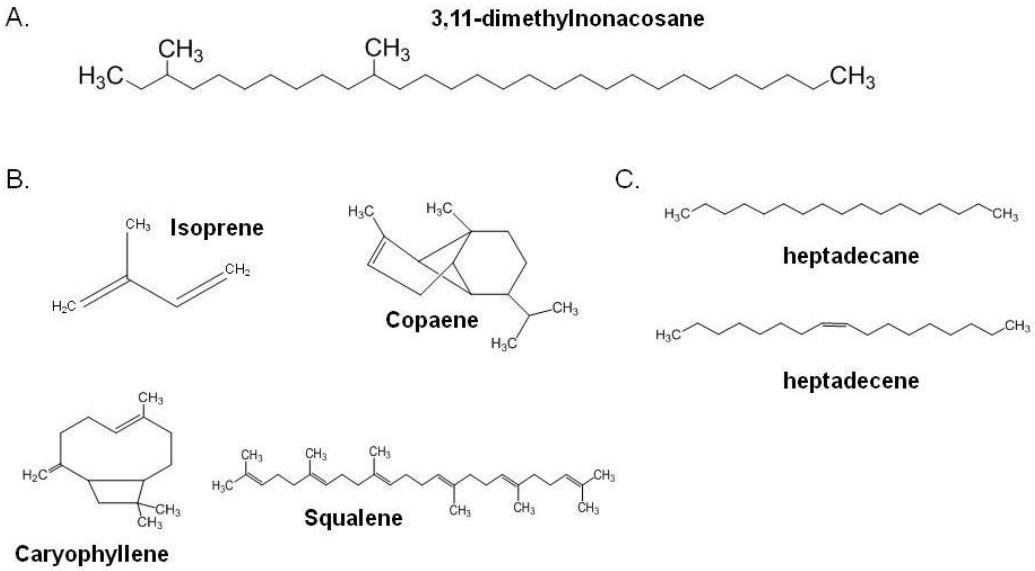


Figure 1.3 Hydrocarbons produced in nature. (A) A representative hydrocarbon produced by insects. 3,11-dimethylnonacosane is produced by *B. germanica* (108). (B) Representative hydrocarbons produced by various plant species. Isoprene is produced by numerous angiosperms. It's precursor is the building block to copaene and caryophyllene, produced by *Copaifera langsdorfii* (46); and squalene, produced by olive trees and tobacco (39, 93). (C) Representative hydrocarbons produced by cyanobacteria (125, 135). These compounds include long-chain alkanes and alkenes.

Insects have an abundance of methyl-branched hydrocarbons within the body and on their cuticular surfaces. These include mono-, di-, tri-, and tetramethyl-branched hydrocarbons ranging from 15 to 55 carbons in length (108). These compounds are proposed to arise from the elongation of appropriate fatty acyl CoA chains (80).

Branching arises from the substitution of methylmalonyl CoA in place of malonyl CoA at specific points during fatty acyl chain elongation (27). After obtaining a specific carbon length, the fatty acyl CoA is reduced to an aldehyde and decarboxylated to a n-1 hydrocarbon (Figure 1.4) (108). Cytochrome P450s are proposed to be involved in decarboxylation since the reaction not only requires NADPH and O_2, but also is inhibited by antibodies against cytochrome P450s (108, 123, 124).

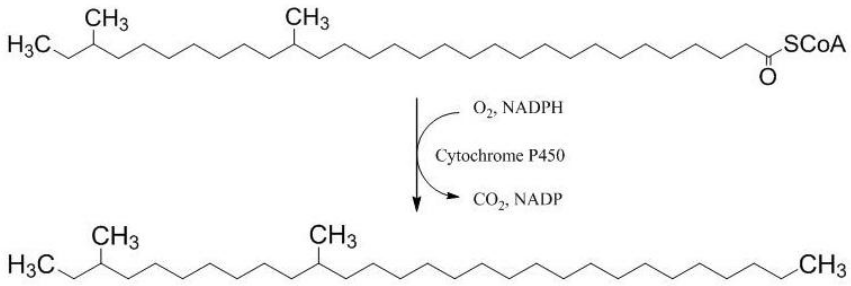


Figure 1.4 Synthesis of insect branched hydrocarbons. A branched fatty acyl CoA is synthesized, reduced to an aldehyde, and decarboxylated to the n-1 hydrocarbon (108, 123, 124).

Like insect hydrocarbon production, isoprenoid biosynthesis is still being deduced. Isoprenoids are produced by approximately 30% of all angiosperms as well as many bacteria and animals (92, 138). Isoprenoid synthesis utilizes the 2-C methyl-D-erythriotol 4-phosphate (MEP) (128, 173) or the mevalonic acid (MVA) (151, 173) pathways for their production (Figure 1.5). Isoprene synthase, the final protein in the isoprene-synthesis pathway, utilizes the final products (isopentenyl diphosphate and dimethylallyl diphosphate) in these pathways to produce isoprene. Many companies and laboratory groups are taking advantage of isoprenoid synthesis to produce

hydrocarbons for industrial purposes. For example, Genencor is currently optimizing a process to make isoprene (59). Since a bacterial isoprene synthase has not been purified and characterized (59, 172), the synthetic pathway is composed of a plant isoprene synthase and utilizes a bacterial-yeast MVA pathway engineered into *E. coli* to deliver isoprene at yields exceeding 60g/L (59). The resulting product can be reformed using polymerization reactions to create useful products. Isopentenyl diphosphate and dimethylallyl diphosphate are also the building blocks to a number of isoprenoid compounds (naturally and artificially produced) including squalene, caryophyllene, and copaene (32, 39, 46, 93) (Figure 1.3B). Other isoprenoid compounds can be utilized as antimicrobial agents (example artemisinin, an antimalarial drug (25, 173)).

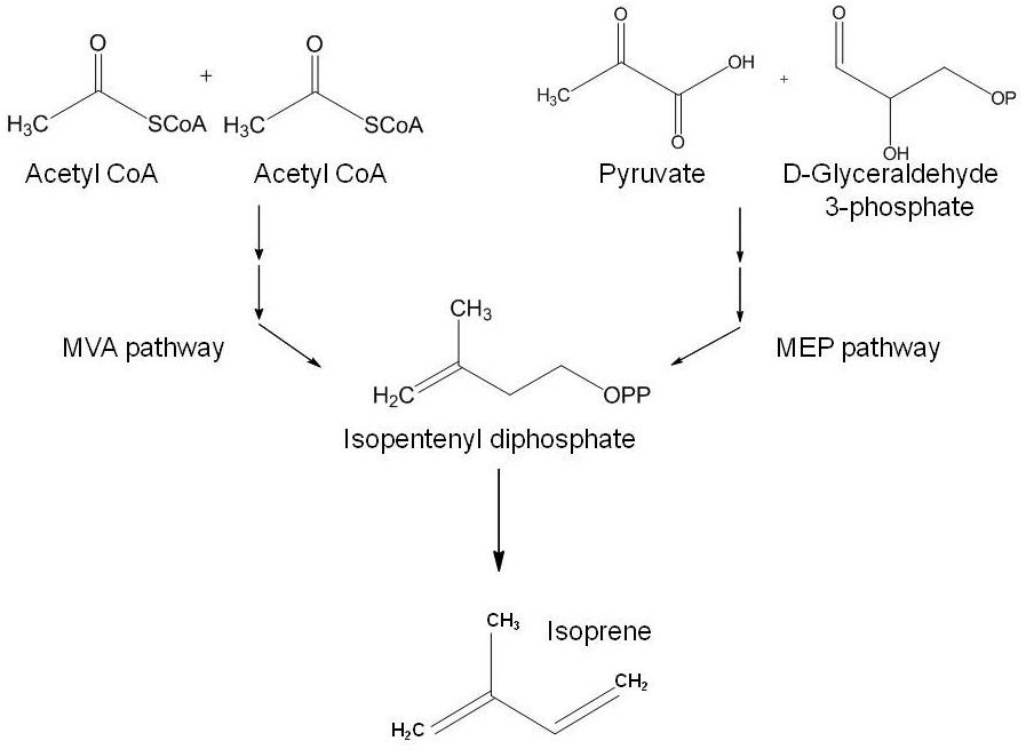


Figure 1.5 Isoprene biosynthesis. Two pathways lead to isoprene in bacteria. The best documented utilizes the MVA pathway while the second is the MEP

pathway. Both lead to the formation of isopentyl diphosphate and dimethylallyl disphosphate. Isoprene (and other terpenoid) synthase utilizes these substances to make isoprenoid monomers (59, 128, 151, 173).

Alkanes produced by cyanobacteria are also being studied. Various groups have shown that cyanobacteria can produce long chain hydrocarbons (45, 125, 135) from reduction of a fatty acyl-acyl carrier protein (ACP) followed by a proposed decarbonylation of the resulting aldehyde (Figure 1.6) (135). LS9, a startup company from California, is in the process of creating an *Escherichia coli*-based hydrocarbon-producing system using the alkane biosynthetic pathway from cyanobacteria. With the incorporation of the acyl-ACP reductase and aldehyde decarbonylase from various cyanobacteria strains into *E. coli*, the company was able to produce up to 80 mg/L of alkanes in preliminary studies (135). This provides a foundation for the understanding and development of the microbial alkane-biosynthetic pathway for fuel use.

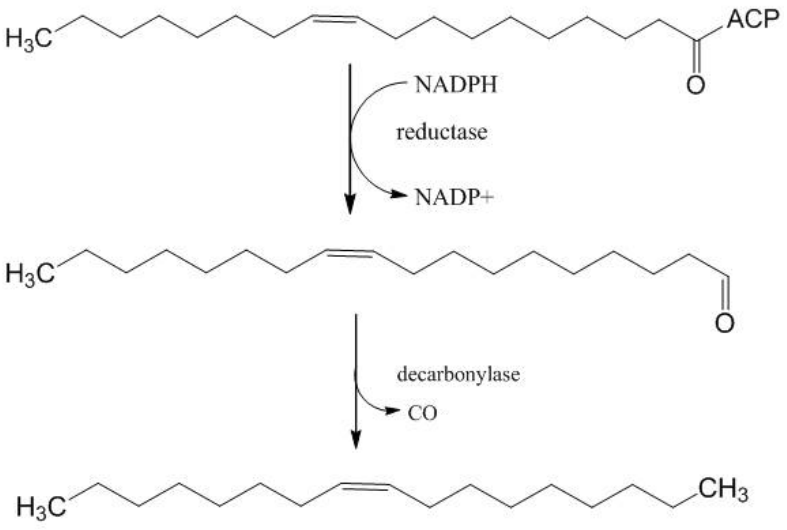


Figure 1.6 Alkane and alkene biosynthesis in cyanobacteria. A fatty acyl-ACP is reduced to an aldehyde prior to a proposed decarbonylation reaction, resulting in hydrocarbon formation (135).

1.3 Hydrocarbon biosynthesis by heterotrophic bacteria investigated in this thesis

The first heterotrophic bacteria reported to produce hydrocarbons were species found in the Micrococcus family (2, 157, 158). Albro *et al.* (1969) and Tornabene *et al.* (1967) reported that *Micrococcus luteus* produced long-chain branched mono-alkenes through the head-to-head condensation of two fatty acid-like compounds (3, 5, 159). The groups found that these products appeared to be constitutively made and were dependent upon fatty acyl availability for product specificity: *M. luteus* naturally condenses two branched fatty acyl CoAs, though when cell-free extracts were incubated with linear fatty acids, incorporation of these exogenous fatty acids into the hydrocarbons occurred along with condensation of the endogenous branched fatty acyl compounds (4). However, Albro *et al.* and Tornabene *et al.* did not identify the genes

and enzymes involved in the hydrocarbon biosynthetic pathway. More recently, Frias *et al.* (2009) expanded the bacterial hydrocarbon knowledge-base by showing that other Micrococcaea, in particular various *Arthrobacter* species, also produce long-chain branched mono-alkenes similar to the products made by *M. luteus* (51).

In 1971, Tornabene *et al.* reported that *Stenotrophomonas maltophilia* produced long chain alkenes, though unlike the *Micrococcus* species, the products found in the γ-*Proteobacteria* were not branched and contained multiple double bonds (161). These authors proposed that the bacteria were condensing two fatty acid-like compounds in a head-to-head manner to produce the alkenes (147).

More recently, Park *et al.* (2001) reported that *Vibrio furnissii* M1, and several other *Vibrio* species, produced diesel-length alkanes and alkenes (116). They reported that the bacteria utilized a wide range of carbon sources, including chitin and xylose, to produce alkenes and alkanes ranging from 14-27 carbons in length (115). Unlike the previously characterized organisms though, Park *et al.* (2005) reported that the *Vibrio* were producing the n-alkanes from the reduction of a fatty alcohol to an alkane (114). Like Albro *et al.* and Tornabene *et al.*, Park *et al.* did not isolate the genes involved in hydrocarbon production.

As of August 2005, no group had identified the genes necessary for the production of hydrocarbons in heterotrophic bacterial models. Also, only a few bacteria had been reported to have the ability to produce hydrocarbons.

1.4 Thesis rationale and goals

The work presented in this thesis represents a portion of the research being conducted in the Wackett Laboratory that focuses on renewable energy through the use of bacterial hosts. Two main hypotheses are presented here: (1) that many heterotrophic bacteria have the ability to produce hydrocarbons, and (2) that bacteria can be manipulated to produce long-chain carbon molecules. The broad goals of this thesis were to discover organisms with the potential to produce hydrocarbons, to identify the genes necessary for hydrocarbon production in heterotrophic bacteria, and to attempt to engineer various bacterial hosts to produce hydrocarbons. These studies are important for understanding intrinsic biological processes and for their potential to contribute to developing next generation biofuels.

1.5 Summary of thesis

The focus of this thesis is on the molecular basis of heterotrophic bacterial hydrocarbon biosynthesis. Chapter 2 describes a study of *Vibrio furnissii* M1, an organism described by others as producing copious amounts of hydrocarbons. In our studies, the results obtained by Park *et al.* were not reproducible and we presented genomic and biochemical evidence suggesting that *V. furnissii* did not produce the hydrocarbons reported. Chapter 3 describes hydrocarbon production by *Shewanella oneidensis* MR-1 and shows that the genes *oleABCD* encode enzymes involved in the head-to-head hydrocarbon biosynthetic pathway. The *oleA* genes from different organisms were heterologously expressed in *Shewanella* and shown to produce different hydrocarbons. Chapter 4 uses genomic analysis to identify *ole* gene sequences in

seventy other organisms that likely biosynthesize hydrocarbons via a head-to-head biosynthetic mechanism. Of those strains identified, selected organisms from different Phyla were grown, extracted and shown to produce long-chain alkenes. Chapter 5 details heterologous expression of the *Chloroflexus aurantiacus oleA* gene in *Shewanella oneidensis* and *Ralstonia eutropha*. This OleA protein in *C. aurantiacus* condenses two C16:1Δ9 fatty acyl CoA chains to make 9,15,22-hentriacontatriene but in alternative hosts condenses other fatty acyl chains to make a broad spectrum of ketone products. Chapter 6 describes the use of different promoters, vectors, and the *Stenotrophomonas maltophilia oleA* gene to increase hydrocarbon production in *S. oneidensis* MR-1. Appendix A describes efforts to create a cyanobacterium expressing an *oleA* gene heterologously. Appendix B describes the use of chemical agents in an effort to select for *Shewanella* strains that produce higher levels of hydrocarbons.

1.6 Note about thesis wording

Chapters 2, 3, and 4 were previously published. At the time of publication it was believed that fatty acids may be the precursors to hydrocarbon production in heterotrophic bacteria, as noted in the reports. Since publication, the actual precursor to hydrocarbon production were identified as fatty acyl CoA chains (J. Frias, University of Minnesota, personal communication).

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

Wackett, LP, J.A. Frias, J.L. Seffernick, D.J. Sukovich, and S.E Cameron. Applied and Environmental Microbiology, November 2007, volume 73, p. 7192-7198.

0099-2240/07/$08.00+0 doi: 10.1128/AEM.01785-07

Text and figures in Chapter 2 were previously published (except for Figure 2.6) and are reproduced here with permission of the authors and the publisher.

Contributions to the writing of this chapter were made by each author. Janice Frias performed the majority of the culture hydrocarbon extractions as well as all of the ratioactive cell free enzyme assays. Jennifer Seffernick did the majority of the genomic annotation and bioinformatics study of the *Vibrio furnissi* M1 genome. David Sukovich performed all the Pulse Field Gel experiments, the Southern Blot analyses, and the Rep-PCR experiments. Stephan Cameron studied the carboxysome region of the bacteria.

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

2.1 Overview

Vibrio furnissii M1 was recently reported to biosynthesize *n*-alkanes when grown on biopolymers, sugars, or organic acids (M. O. Park, J. Bacteriol. **187:**1426-1429, 2005). In the present study, *V. furnissii* M1 was subjected to genomic analysis and studied biochemically. The sequence of the 16S rRNA gene and repetitive PCR showed that *V. furnissii* M1 was not identical to other *V. furnissii* strains tested, but the level of relatedness was consistent with its assignment as a *V. furnissii* strain. Pulsed-field gel electrophoresis showed chromosomal bands at approximately 3.2 and 1.8 Mb, similar to other *Vibrio* strains. Complete genomic DNA from *V. furnissii* M1 was sequenced with 21-fold coverage. Alkane biosynthetic and degradation genes could not be identified. Moreover, *V. furnissii* M1 did not produce demonstrable levels of *n*-alkanes in vivo or in vitro. In vivo experiments were conducted by growing *V. furnissii* M1 under different conditions, extracting with solvent, and analyzing extracts by gas chromatography-mass spectrometry. A highly sensitive assay was used for in vitro experiments with cell extracts and [^{14}C]hexadecanol. The data are consistent with the present strain being a *V. furnissii* with properties similar to those previously described but lacking the alkane-producing phenotype. *V. furnissii* ATCC 35016, also reported to biosynthesize alkanes, was found in the present study not to produce alkanes.

2.2 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 (114-116).

V. furnissii strains were recognized as a distinct species in 1983 (18). Other *Vibrio* species, such as *V. cholerae* (33) and *V. parahaemolyticus* (175), have been more extensively studied because of their significant pathogenicity in humans. Both of the latter species (66), along with *V. vulnificus* (29) and *V. fischeri* (130), 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, 84). Marine *Vibrios* are prominent chitinolytic organisms (98).

Thus, the recent report of a *V. furnissii* strain biosynthesizing appreciable quantities of *n*-alkanes was unusual and interesting (114-116). The organism was isolated from activated sludge of a sewage disposal plant located in the Osaka prefecture of Japan (116). 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 (116), 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 (115). 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 (114). 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.

2.3 Materials and Methods

2.3.1 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.

2.3.2 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%.

2.3.3 Analytical methods

Analytical methods largely followed those described by Park *et al*. (115, 116). Cultures were extracted following the Bligh and Dyer protocol (15). 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.

2.3.4 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.

2.3.5 In vitro experiments

Cell-free enzyme fractions were obtained and assays were conducted as described previously by Park (114). 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.

2.3.6 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'-AAGTAAGTGACTGGGGTCAGCG-3') (41). 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 (41, 70, 76).

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).

2.3.7 Genome sequencing and annotation

DNA for sequencing was collected as previously described (131). *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'-AGAGTTTGATCMTGGCTCAG-3') and 785r (5'-GGACTACCIGGGTATCTAATCC-3'), 530f (5'-GTGCCAGCMGCCGCGG-3') and 1100r (5'-GGGTTGCGCTCGTTG-3'), and 926f (5'-AAACTYAAAKGAATTGACGG-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*. (153).

The resulting pseudochromosome was subjected to automated annotation via GenDB (104). Pfam (142) and hidden Markov models (HMM) (42) 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 (6).

2.3.8 Functional analysis of ORF 275

V. furnissii M1 genomic DNA and the primers GGATTATGGCATATGATGTTAGAT 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 $NADH_2$.

2.3.9 Nucleotide sequence accession number

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

2.4 Results

2.4.1 *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*. (116). 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 Figure 2.1. A dendogram 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 organisms of the same species, confirming that *V. furnissii* M1 is indeed a *V. furnissii* strain.

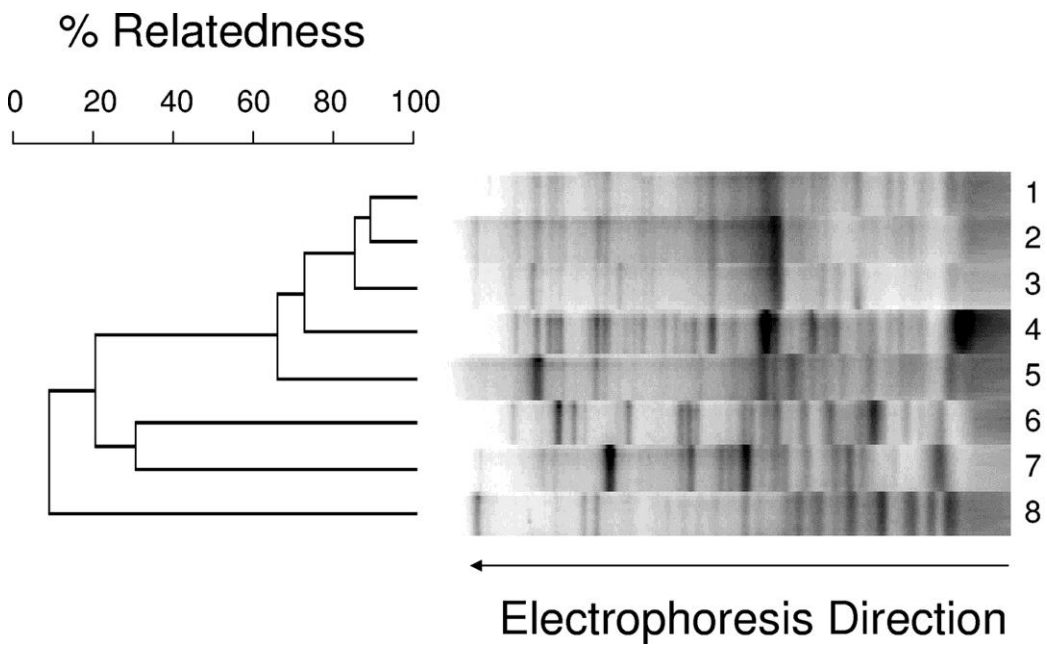


Figure 2.1 Rep-PCR of genomic DNA from *V. furnissii* M1 and control strains. Percent relatedness was determined as described in Materials and Methods. The lanes contained DNA from the following strains: 1, *V. furnissii* ATCC33841; 2, *V. furnissii* ATCC35627; 3, *V. furnissii* ATCC35016; 4. *V. furnissii* M1; 5, *V. furnissii* ATCC35628; 6, *V. parahaemolyticus* LM5312; 7, *V. harveyi* B392; 8, *E. coli* K-12.

2.4.2 PFGE

Genomic DNA of *V. furnissii* M1, other *V. furnissii* strains, and standard strains were analyzed by PFGE (Figure 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 Figure 2.2 (29, 113).

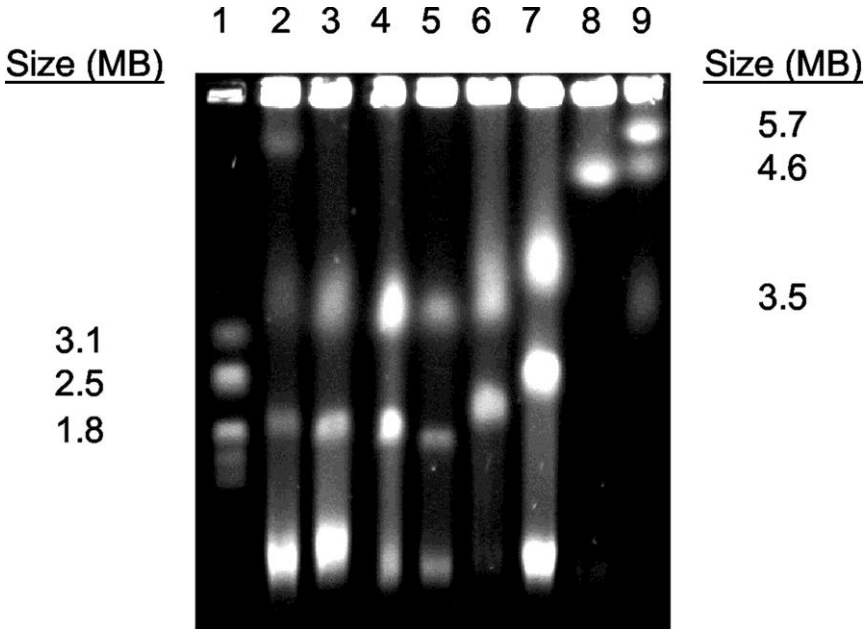


Figure 2.2 Comparative PFGE of genomic DNA from *V. furnissii* M1 (lane 2), *V. furnissii* 35016 (lane 3), *V. furnissii* 35627 (lane 4), *V. furnissii* 35628 (lane 5), *V. parahaemolyticus* LM5312 (lane 6), *V. harveyi* B392 (lane 7), and *E. coli* DH5α (lane 8). *Hansenula wingei* (lane 1) and *Schizosaccharomyces pombe* (lane 9) chromosomes were used as size markers.

2.4.3 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 (114).

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) (Figure 2.3). These functional genes were flanked by genes that were annotated as metabolosome (carboxysome) shell proteins (Figure 2.3). The flanking gene structure is consistent with the function of *pduJ* and *ccmO* genes in generating a carboxysome or metabolosome structure (16, 82), which is a bacterial intracellular protein shell involved in compartmentalizing metabolites or a set of reactions.

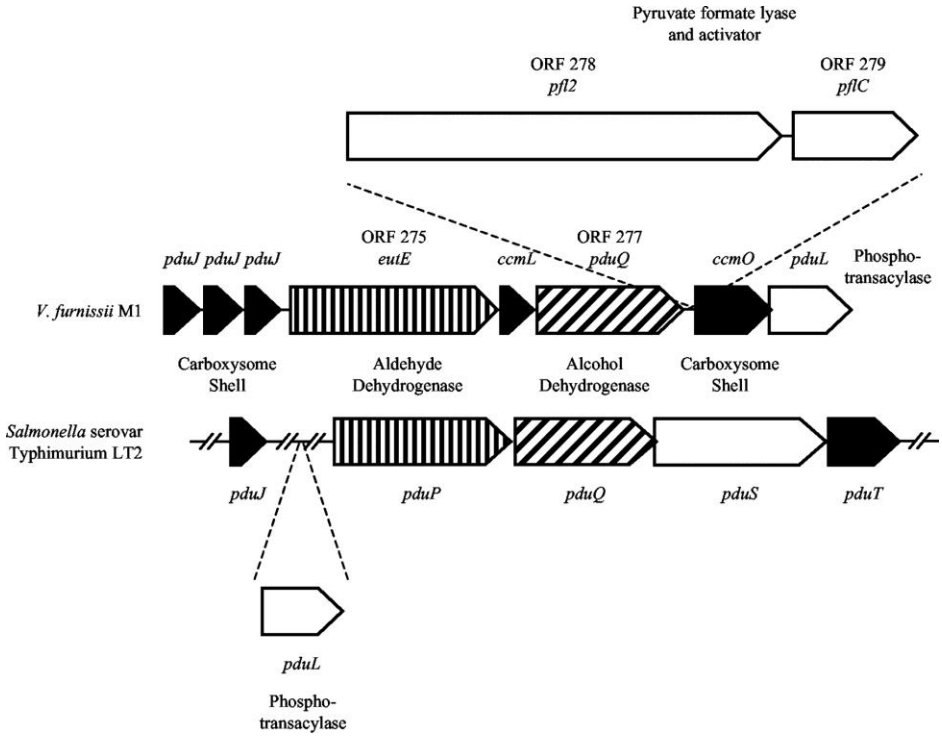


Figure 2.3 Genome region from *V. furnissii* M1 showing similarities to a gene region in *Salmonella*. Identified protein types are highlighted: carboxysome shell proteins (black), aldehyde dehydrogenases (vertical lines), and alcohol dehydrogenases (diagonal lines). The ORF numbers (275, 277, 278, and 279) are given for proteins discussed in the text.

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 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 (83). 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.

2.4.4 Whole-cell studies attempting to detect alkanes

V. furnissii M1 cells and media were extracted by methods described by Park *et al*. (115, 116). In preliminary experiments, alkanes were detected by GC and confirmed by MS (Figure 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 (Figure 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.

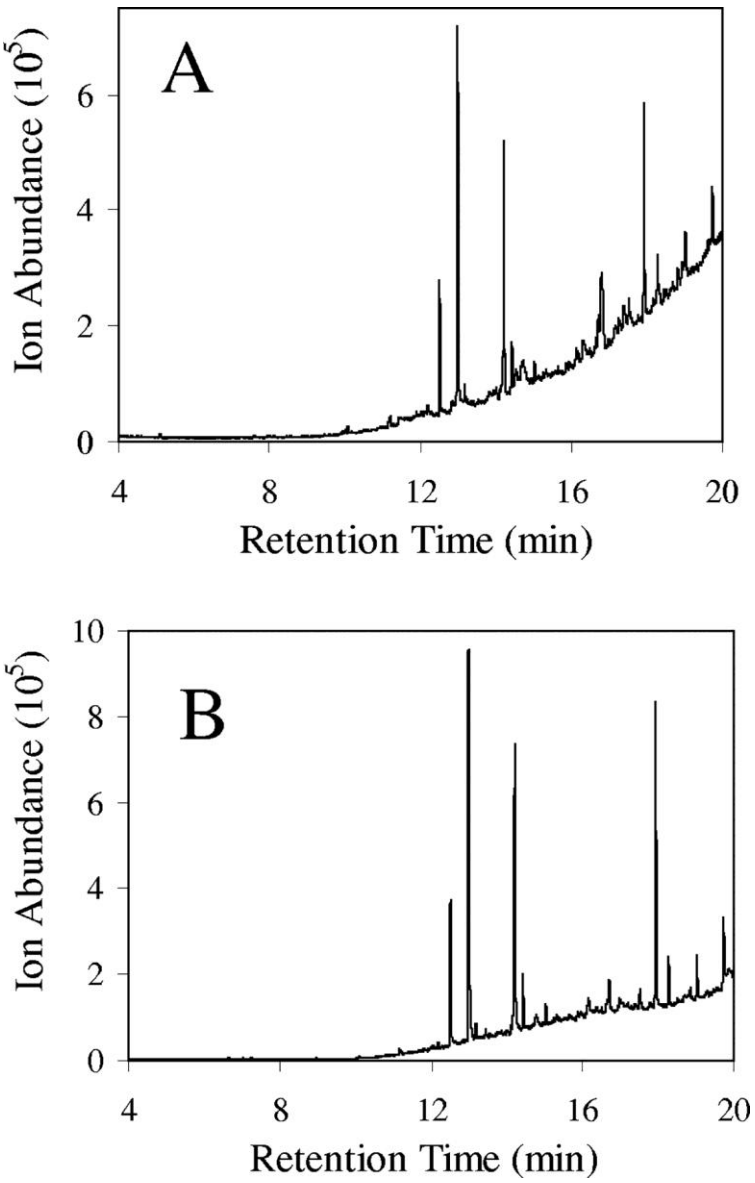


Figure 2.4 Initial gas chromatograms of *V. furnissii* M1 (A) and a chloroform blank showing the extent of contamination (B). The prominent peaks in both were identified by MS as methyl palmitate (12.5 min), dibutyl phthalate (13.0), octadecenoic acid methyl ester (14.2 min), diethylhexyl phthalate (17.9 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 (115, 116). 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*. (116). 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 (Figure 2.5A) and *E. coli* (Figure 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.

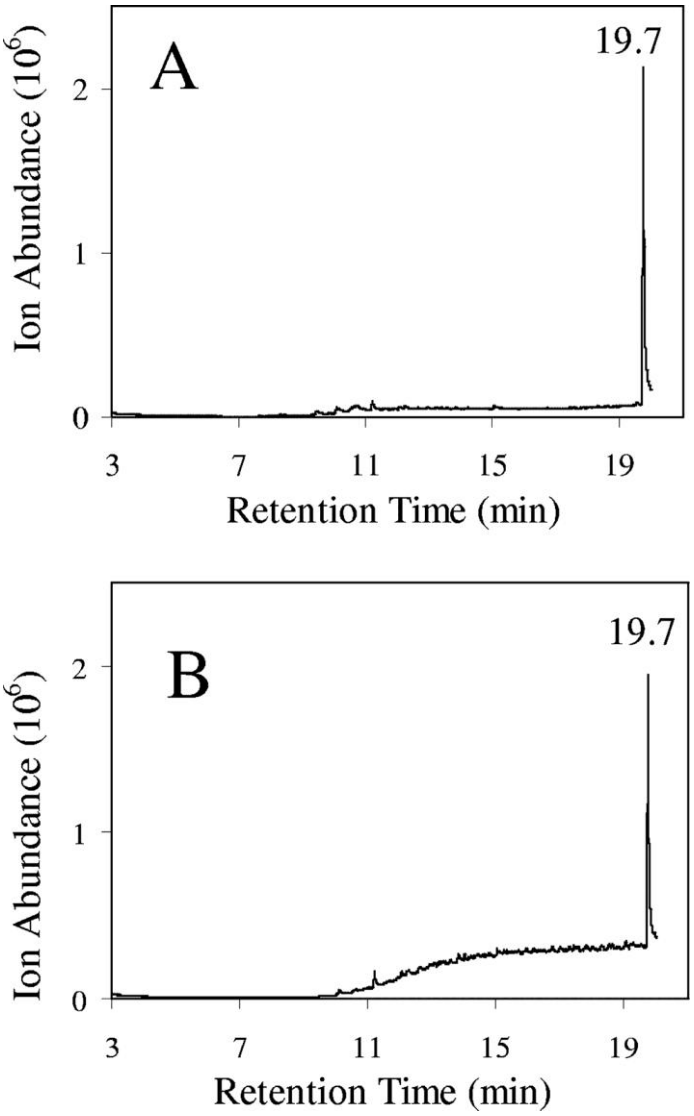


Figure 2.5 Gas chromatograms of extracts of *V. furnissii* M1 (A) and *E. coli* K-12 (B) using cleaner solvents and methods. Cultures were spiked with octacosane prior to extraction and workup. The 19.7-min peak was confirmed by MS to be octacosane.

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

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.

2.4.5 Cell-free enzyme assays for alkane formation

Cell-free enzyme preparations from *V. furnissii* M1, prepared as described by Park (114), 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 (114). A level of activity several percent of that reported by Park (114) would have been detected in this experiment. In separate experiments, *V. furnissii* ATCC 35028 was tested for reductase activity with hexadecanol, but no activity was detected.

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, membrane	799,030 ± 143,600	812 ± 162
E. coli, soluble	797,954 ± 246,564	642 ± 185
E. coli, membrane	795,810 ± 127,275	474 ± 152

2.4.6 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 (Figure 2.6). 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 (42) 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*.

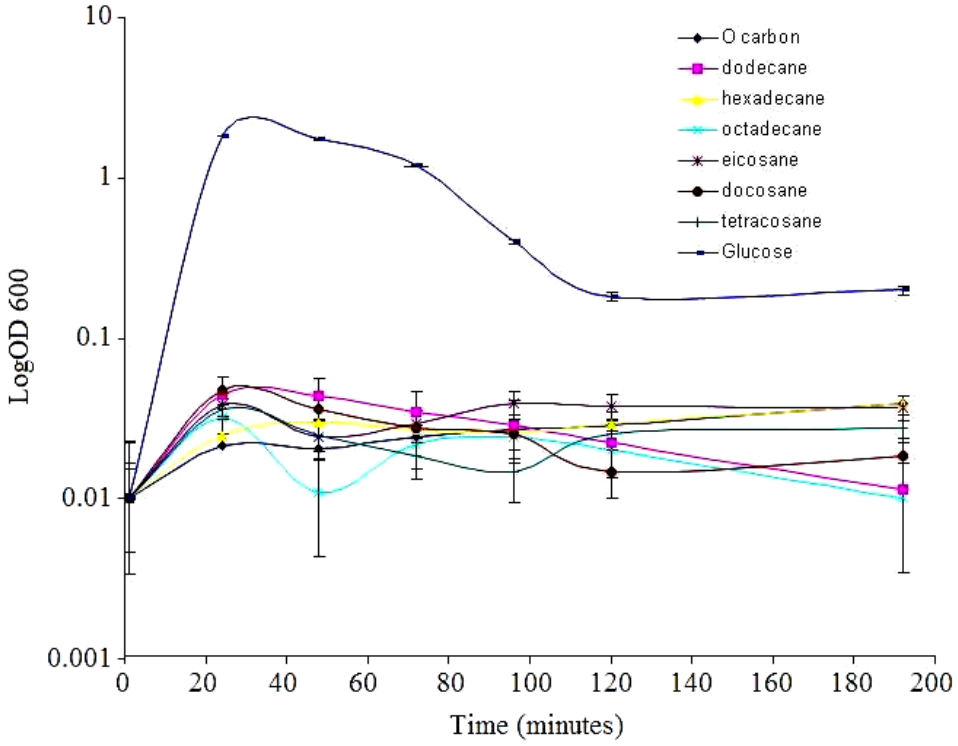


Figure 2.6 Growth of *Vibrio furnissii* M1 on glucose and select hydrocarbons. Bacteria were grown on 250mM carbon equivalents of the various carbon compounds in M3 media (n = 6). Error bars represent standard deviations.

2.4.7 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 (105). 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.

2.5 Discussion

The *V. furnissii* M1 strain used in this study strongly resembles the strain described previously (114-116), 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*. (114-116) 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 (114) 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 (105) 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 (7, 144). 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 (82). 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 (83). The channeling multidomain protein and a metabolosome may represent different biological mechanisms for channeling metabolic flux through a potentially toxic aldehyde intermediate.

2.6 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.

Structure function and insights into the biosynthesis of a head-to-head hydrocarbon in *Shewanella oneidensis* strain MR-1.

Sukovich, D.J., J.L. Seffernick, J.E. Richman, K.A. Hunt, J.A. Gralnick, and L.P. Wackett. June 2010, volume 76, p. 3842-3849.

0099-2240/10/$12.00 doi: 10.1128/AEM.00433-10

Text and figures in Chapter 3 were previously published (except for Chapter 3.6) and are reproduced here with permission of the authors and the publisher.

Contributions to the writing of this chapter were made by each author. David Sukovich constructed a number of the bacterial mutants as well as their complements; extracted, purified, and analyzed all the bacterial cultures; performed all the growth studies; and constructed the strains containing the heterologously expressed OleA proteins. With the help of Jack Richman, David Sukovich also performed the chemistry experiments. Jennifer Seffernick performed the bulk of the bioinformatics studies, identifying the genes necessary for hydrocarbon production. Jack Richman's expertise was used to identify the hydrocarbon product produced by *Shewanella oneidensis* through chemical manipulation of the purified substance and proton NMR. Kristopher Hunt constructed a number of the bacterial mutants.

Chapter 3: Structure function and insights into the biosynthesis of a head-to-head hydrocarbon in *Shewanella oneidensis* strain MR-1

3.1 Overview

A polyolefinic hydrocarbon was found in nonpolar extracts of *Shewanella oneidensis* MR-1 and identified as 3,6,9,12,15,19,22,25,28-hentriacontanonaene (compound I) by mass spectrometry, chemical modification, and nuclear magnetic resonance spectroscopy. Compound I was shown to be the product of a head-to-head fatty acid condensation biosynthetic pathway dependent on genes denoted as *ole* (for *ole*fin biosynthesis). Four *ole* genes were present in *S. oneidensis* MR-1. Deletion of the entire *oleABCD* gene cluster led to the complete absence of nonpolar extractable products. Deletion of the *oleC* gene alone generated a strain that lacked compound I but produced a structurally analogous ketone. Complementation of the *oleC* gene eliminated formation of the ketone and restored the biosynthesis of compound I. A recombinant *S. oneidensis* strain containing *oleA* from *Stenotrophomonas maltophilia* strain R551-3 produced at least 17 related long-chain compounds in addition to compound I, 13 of which were identified as ketones. A potential role for OleA in head-to-head condensation was proposed. It was further proposed that long-chain polyunsaturated compounds aid in adapting cells to a rapid drop in temperature, based on three observations. In *S. oneidensis* wild-type cells, the cellular concentration of polyunsaturated compounds increased significantly with decreasing growth temperature. Second, the *oleABCD* deletion strain showed a significantly longer lag phase than the

wild-type strain when shifted to a lower temperature. Lastly, compound I has been identified in a significant number of bacteria isolated from cold environments.

3.2 Introduction

Currently, there is industrial interest in nongaseous microbial hydrocarbons for specialty chemical applications and, more recently, as high-energy biofuels (95, 143, 169). Microbes produce hydrocarbons of different types, for example, aliphatic isoprenoid compounds (95) and alkanes from fatty aldehyde decarbonylation (38). Fatty aldehyde decarbonylation is not well understood but offers a clean route to diesel fuels from fatty acids.

Certain microbes also make a distinctly different class of long-chain hydrocarbons, generally C_{25} to C_{33} in chain length, that contain a double bond near the middle of the chain (2, 3, 14, 64, 158, 159, 169, 170). These long-chain olefinic hydrocarbons are thought to derive from processes different than isoprene condensation and decarbonylation mechanisms. This class of hydrocarbons has been shown by carbon-14-labeling studies (5) to derive from fatty acids. The process, described in 1929 by Channon and Chibnall (26), has become known as head-to-head hydrocarbon biosynthesis. Albro and Ditmar (3) defined the head-to-head condensation as coupling of the head (C_1) and the α-carbon (C_2) of two fatty acids with decarboxylation, a reaction that should not be confused with an acyloin-like carboxyl carbon-to-carboxyl carbon coupling. Products of the head-to-head mechanism have been identified in Gram-positive bacteria such as *Micrococcus luteus* (157, 158) and *Arthrobacter aurescens*

(51) and in Gram-negative bacteria such as *Stenotrophomonas maltophilia* (147). *Micrococcus* and *Arthrobacter* strains produce fatty acids that are methyl branched terminally and subterminally (24, 157, 158). The long-chain olefinic hydrocarbons from those strains similarly contain a mixture of terminal and subterminal methyl group branching (5, 51, 159).

Albro and Ditmar (3, 4) acquired direct evidence for the head-to-head mechanism occurring in microbial whole organisms and cell extracts. In cell extracts, it was shown that one of the fatty acid carboxyl groups is lost as carbon dioxide, with the remaining carbon atoms being retained in the resultant hydrocarbon (4). The hydrocarbons contain a double bond at the point of condensation. More recently, Beller *et al*. described the genes encoding head-to-head fatty acid condensation pathway enzymes from *Micrococcus luteus*, which are known as *ole* genes for the *ole*fin products formed (14). Three genes from *Micrococcus luteus* were shown to confer on *Escherichia coli* the ability to make long-chain olefinic hydrocarbons. Two recent patent applications by L. Friedman *et al*. (18 September 2008, WO2008/113041; 4 December 2008, WO2008/147781) also described a three- or four-gene cluster as being involved in head-to-head hydrocarbon biosynthesis to make olefins. The patent applications identified homologs to *ole* genes in different bacteria, including strains of *Shewanella*.

Bacteria of the genus *Shewanella* have been heavily studied over the last decade because they are widespread and have the ability to use a startling variety of electron acceptors for respiration (49). There are more than 20 completed genome sequences for *Shewanella* strains. The model system for studying *Shewanella* is *S. oneidensis* MR-1.

The genome sequencing of *S. oneidensis* MR-1 was reported in 2002 (67), and the organism has been shown to be highly amenable to genetic manipulation (49).

The present study used *Shewanella oneidensis* strain MR-1 as a model system to investigate hydrocarbon biosynthetic genes and the possible biological function of the proteins they encode. The hydrocarbon produced by the Ole proteins in *S. oneidensis* MR-1 was found to be very different from hydrocarbons previously identified as deriving from a head-to-head condensation mechanism (147, 157, 161). The product was identified here as 3,6,9,12,15,19,22,25,28-hentriacontanonaene by chemical modification studies, mass spectrometry, and nuclear magnetic resonance (NMR) spectroscopy. Previously, a similar polyolefin had been identified in many Antarctic bacteria (110). Cloning of a heterologous *oleA* gene into *S. oneidensis* MR-1 was found to produce a completely different set of products. A hydrocarbon deletion mutant showed a distinctly longer growth lag than wild-type cells when shifted to a lower temperature, suggesting that the *ole* genes in *S. oneidensis* MR-1 may aid the cells in adapting to a sudden drop in temperature.

3.3 Materials and Methods

3.3.1 Bacterial strains, culture conditions, and growth

A list of *Shewanella* strains used in this study can be found in Table 3.1. Cultures of *S. oneidensis* MR-1 were routinely grown in Luria-Bertani (LB) medium

under ideal conditions (aerobic, 30°C) unless stated otherwise. Cultures were grown to early stationary phase at 36°C, 22°C, 15°C, or 4°C for experiments in which the relative amount of hydrocarbon was determined (n = 6). In cold adaption experiments (n = 6), the *oleABCD* mutant and wild-type strains were first grown to a similar optical density (OD) on LB medium overnight at 30°C and then diluted by the same dilution factor into fresh medium at 4°C with a beginning OD of approximately 0.01. Aerobic growth was continued at 4°C, and optical densities were measured using a Beckman DU 7400 spectrophotometer. For each treatment (six flasks), three OD measurements were made and then averaged.

Table 3.1 Strains and plasmids used for Chapter 3

Strain or plasmid Strains	Genotype or relative characteristic(s)	Ref
Strains		
Shewanella oneidensis MR-1	Wildtype	
Shewanella oneidensis Δole	*S. oneidensis* MR-1, *Δole*; does not produce hydrocarbon	This study
Shewanella oneidensis ΔoleC	*S. oneidensis* MR-1, *ΔoleC*; does not produce hydrocarbon	This study
Shewanella oneidensis ΔpfaA	*S. oneidensis* MR-1, *Δpfa*A; does not produce hydrocarbon	This study
Escherichia coli UQ950	*E. coli* DH5α λ(pir) host for cloning; F-Δ(*argF-lac*)*169 Φ80dlacZ58*(ΔM15) *glnV44*(AS) *rfbD1 gyrA96*(NalR) *recA1 endA1 spoT1 thi-1 hsdR17 deoR λpir+*	133
Escherichia coli WM3064	Donor strain for conjugation: *thrB1*004 *pro thi rpsL hsdS lacZ*ΔM15 RP4-1360 Δ(*araBAD*)*567 ΔdapA*1341::[*erm pir*(wt)]	133
Plasmid		
pSMV3	9.5-kb vector; Km^r-only version of pSMV8; *lacZ*; *sacB*	133
pSMV3-*Δole*	2.3-kb fusion PCR fragment containing *Δole* cloned into the SpeI/SacI site of pSMV3; used to make the *S. oneidensis Δole* strain	This study
pSMV3-*ΔoleC*	2.2-kb fusion PCR fragment containing *ΔoleC* cloned into the SpeI/SacI site of pSMV3; used to make the *S. oneidensis ΔoleC* strain	This study
pSMV3-*ΔpfaA*	2.0-kb fusion PCR fragment containing *ΔpfaA* cloned into the SpeI/ApaI site of pSMV3; used to make the *S. oneidensis ΔpfaA* strain	This study
pBBR1MCS-2	5.1-kb broad-host range plasmid; *lacZ*; Km^r	89
pOleC	2.1-kb PCR fragment containing the *S. oneidensis oleC*, cloned into the SpeI/SacI site of pBBR1MCS-2	This study
pPfaA	7.6-kb PCR fragment containing the *S. oneidensis pfaA*, cloned into the ApaI/SpeI site of pBBR1MCS-2	This study
pOleA-S.m.	1.1-kb PCR fragment containing the *S. maltophilia oleA*, cloned into the SpeI/SacI site of pBBR1MCS-2	This study

For maintenance of plasmids in *S. oneidensis* strains, 50 μg/ml of kanamycin (Km) was added to the medium. For selection for recombinants (see "Mutagenesis," below), Km was added to a final concentration of 50 μg/ml while sucrose was added to a final concentration of 5% (wt/vol). *Escherichia coli* strains and their genotypes are listed in Table 3.1. All *E. coli* strains were grown aerobically at 37°C in LB. Where appropriate, Km was added to the growth medium at a final concentration of 50 μg/ml and diaminopimelic acid was added to a final concentration of 0.3 mM.

3.3.2 Hydrocarbon and ketone analysis

Hydrocarbons and ketones were analyzed by gas chromatography-mass spectrometry (GC-MS) as previously described (50). Early-stationary-phase cultures, cells and medium together, were extracted. The resulting evaporated residue was recovered in 1 ml of methyl-*tert*-butyl ether and applied to a 4.0-g silica gel column, eluted with 35 ml of hexanes, concentrated, and subjected to molecular distillation using a Bantamware sublimation apparatus. The hydrocarbon distillate was collected between 100 and 115°C (0.02 torr), and the ketone distillate was collected between 120 and 130°C (0.02 torr). The distillates were recovered in 1 ml of pentanes and subjected to GC-MS analysis using an HP6890 gas chromatograph connected to an HP5973 mass spectrometer (Hewlett Packard, Palo Alto, CA). GC conditions consisted of the following: helium gas at 1 ml/min; HP-1ms column (100% dimethylpolysiloxane capillary, 30 m by 0.25 mm by 0.25 μm); temperature ramp, 100 to 320°C, at 10°C/min, with a 5-min hold at 320°C. The mass spectrometer was run in electron impact mode at 70 eV and 35 μA.

The 3,6,9,12,15,19,22,25,28-hentriacontanonaene (compound I) produced by wild-type *S. oneidensis* MR-1 was purified and identified through GC-MS and NMR analyses. NMR was performed using a Varian INOVA 500 MHz NMR apparatus. Olefin hydrogenation used 5% palladium on carbon as the catalyst under hydrogen at 1 to 2 atm pressure. Chemical characterization: thin-layer chromatography (TLC; hexanes:dichloromethane at 80:20 [vol/vol]), $R_F = 0.13$; (hexanes:dichloromethane, 80:20 [vol/vol], silver nitrate), $R_F = 0.027$; ^{1}H-NMR (500 MHz, $CDCl_3$): 5.28 to 5.45 ppm (17.8 H), 2.76 to 2.92 (14.0 H), 2.14 to 2.22 (3.9 H), 2.00 to 2.12 (4.8 H), 0.94 to 1.02 (5.9 H); UV/vis: λ_{max} 208 nm; medium-resolution MS (*m*/*z*): $[M]^+$ calculated for $C_{31}H_{46}$: 418.7; found: 418.3.

3.3.3 Mutagenesis

Deletion of the *oleABCD* cluster and *oleC* from MR-1 was achieved utilizing homologous recombination between flanking regions of the target gene(s) cloned into a suicide vector (133). Briefly, upstream and downstream regions of the target deletion were cloned into the suicide vector pSMV3 in a compatible *E. coli* cloning strain UQ950. The suicide vector was transformed into an *E. coli* mating strain WM3064 and then conjugated into MR-1. The initial recombination event was selected for by resistance to Km. Cells containing the integrated suicide vector were grown in the absence of selection overnight at 30°C and then plated onto LB plates containing 5% sucrose (133). Cells retaining the suicide vector were unable to grow due to the activity of SacB, encoded on the vector, while cells that had undergone a second recombination event formed colonies. Colonies were then screened by PCR to determine strains

containing the deletion. For creation of the *oleABCD* cluster knockout strain, primers oleclusterUF, oleclusterUR, oleclusterDF, and oleclusterDR containing SpeI, BsaI, BsaI, and SacI restriction sites, respectively, were designed for the regions flanking the two ends of the *oleABCD* cluster (gi numbers 24373309, 24373310, 24373311, and 24373312, respectively; locus tags SO_1742, SO_1743, SO_1744, and SO1745, respectively). For creation of the *oleC* knockout strain, primers oleCUF, oleCUR, oleCDF, and oleCDR containing SpeI, BsaI, BsaI, and SacI restriction sites, respectively, were designed for the regions flanking the ends of *oleC* (gi 24373311; locus tag SO_1744). Finally, for the creation of the *pfaA* knockout strain, primers pfaA1F, pfaA1R, pfaA2F, and pfaA2R containing the SpeI, BamHI, BamHI, and ApaI restriction sites, respectively, were designed for the regions flanking the ends of *pfaA* (gi 24373171; locus tag SO_1602). Primer names and sequences are listed in Table 3.2.

TABLE 3.2 Primers used in Chapter 3

Primer	Sequence
oleclusterUF	TTACTAGTATCATGCCAACCCTTTTCGC
oleclusterUR	TTGGTCTCCATCGGATAATTGATGCC
oleclusterDF	TTGGTCTCTCGATAGAAGAGGGGATG
oleclusterDR	AAGAGCTCGCACTCGGTGTTGATACAAA
oleCUF	TTACTAGTTTTAACGAAGGTGCGCTAAGG
oleCUR	AAGGTCTCCTCGAACAGCGCATCATCCA
oleCDF	TTGGTCTCATCGAGCTTGATCAATCTTT
oleCDR	AAGAGCTCCAGCTTCAGCTTACCTAAAC
pfaA1F	ACTAGTGCACTCAAGTCGCAGATATTGTTCGCA
pfaA1R	GGATCCACCAACGATGGCAATGGGCAT
pfaA2F	GGATCCAGTAAGACGCTTAACCAAGCAT
pfaA2R	GGGCCCGGTCAATGAATCAATCAGTTGCAACAAC
SO1744Fcomp	ACTAGTGATTACCCATATCAAGCACTTTATGACTGAGA
SO1744Rcomp	GAGCTCTTGAATGCAATGGGATAATGTTTCATCCC
pfaAcomplementF	GGGCCCATGAGCCATACCCCTTCACAGCCT
pfaAcomplementR	ACTAGTTAATGCGGCATGTGCGATTGGGTTGAGTG
SmclusterCompF	ACTAGTCCCCCTTTTGCCTGAGCCTTGGCGC
SmthiolaseCompR	GAGCTCGAAGATCATCGCTGTCCGTCGCGAGC

3.3.4 Mutant complementation and heterologous gene expression.

Complementation of the *oleC* and *pfaA* mutants was performed using the pBBR1MCS-2 expression vector (89) and the endogenous *lac* promoter (which is constitutive in MR-1 due to the absence of *lacI*). Primers SO1744Fcomp and SO1744Rcomp containing SpeI and SacI restriction sites or pfaAcomplementF and pfaAcomplementR containing ApaI and SpeI restriction sites were designed for the regions flanking the ends of *oleC* (gi 24373311; locus tag SO_1744) or *pfaA* (gi 24373171; locus tag SO_1602), respectively. The *Stenotrophomonas maltophilia oleA* (gi 194363945; locus tag Smal_0167) was amplified using primers SmclusterCompF and SmthiolCompR containing the SpeI and SacI restriction sites. Resulting PCR products were ligated into the Strataclone cloning system (Agilent Technologies) followed by ligation of the product into the pBBR1MCS-2 expression vector. Constructs were introduced into *E. coli* WM3064 and conjugated into the *oleC* deletion, *pfaA* deletion, or wild-type *S. oneidensis* MR-1 strain. Appropriately oriented inserts were verified by PCR analysis. The expression of the cloned genes was verified by detection of product activity using GC-MS analysis.

3.3.5 Sequence analysis.

Sequence comparisons were made using the National Center for Biotechnology Information BLAST (bl2seq) tool. Ole protein sequences from *S. oneidensis* MR-1 and *M. luteus* were compared. The gi numbers and sequences were obtained from the GenBank database.

3.4 Results and Discussion

3.4.1 A long-chain hydrocarbon is present in *S. oneidensis* cells at all growth phases.

The hydrocarbon was identified in the nonpolar fraction following solvent extraction from the cultures. Gas chromatography-mass spectrometry showed a single sharp peak at 20.2 min that had a parent ion at 418 mass units (Figure 3.1A). Reduction of the product with hydrogen yielded a single product with a slightly longer retention time and a parent ion of 436 mass units (Figure 3.1). The reduced product behaved identically to the C_{31} *n*-alkane hentriacosane. This indicated that the biological product was a hentriacontanonaene, but the positions of the nine double bonds could not be deduced from mass spectrometry. The compound had no appreciable UV absorbance above 230 nm, suggesting that the double bonds were not in conjugation. The proton NMR was decisive (Figure 3.2) and consistent with one nearly centrosymmetric structure only, specifically, 3,6,9,12,15,19,22,25,28-hentriacontanonaene (compound I). The absolute stereochemistry at the double bonds remains to be determined but is shown in the figure as all-*cis* because of further data on its biosynthetic origin (see below). The structure of compound I was consistent with it being derived from a head-to-head condensation between two fatty acyl chains to produce long-chain olefins containing a double bond between the central and an adjacent carbon atom in the chain.

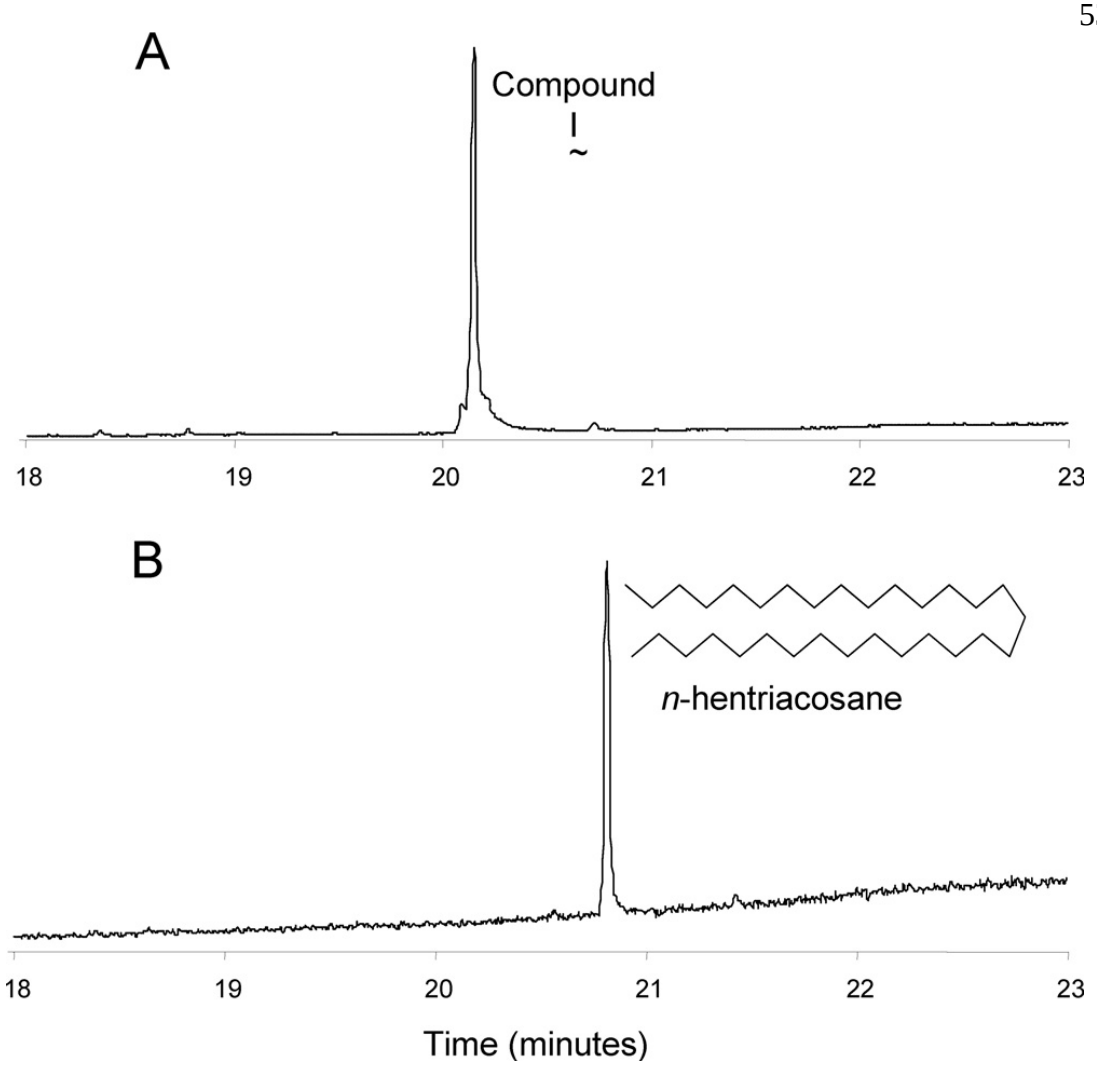


Figure 3.1. Gas chromatograph of the *S. oneidensis* hydrocarbon compound I (20.2 min) (A) and the product of its hydrogenation (20.8 min) that comigrates with and has an identical mass spectrum to *n*-hentriacosane (B).

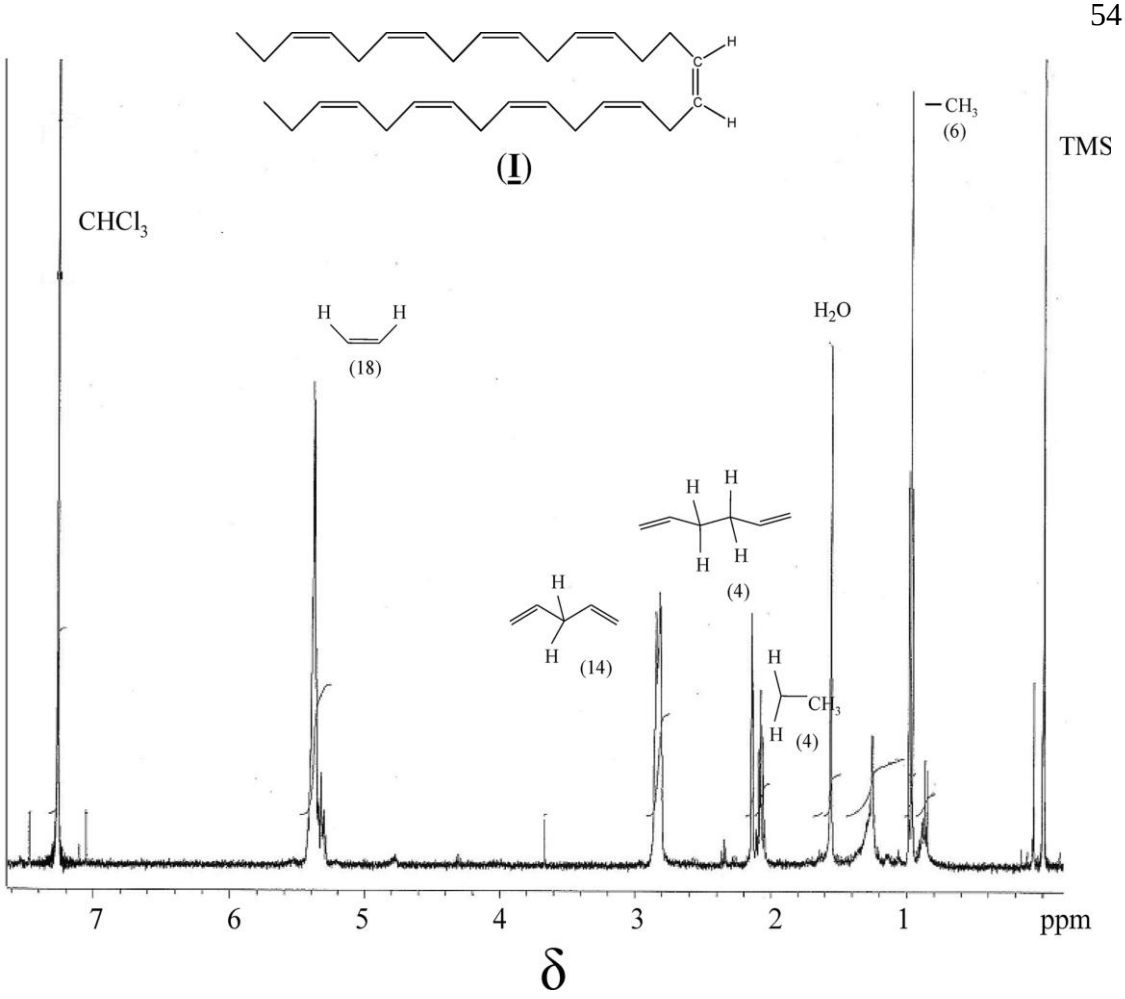


Figure 3.2 NMR spectrum of the hydrocarbon compound I produced by *S. oneidensis* strain MR-1 in deuterated chloroform ($CHCl_3$) with tetramethylsilane (TMS) as the reference standard. The fragment representing each resonance and the number of the protons on integration are indicated. The structure of the compound represented by the spectrum is shown at the top.

3.4.2 Origin of the fatty acids undergoing head-to-head condensation.

The structure of the hydrocarbon (compound I) produced by *S. oneidensis* MR-1 would require the condensation of two molecules of hexadeca-4,7,10,13-tetraenoic acid or an acyl equivalent of this, for example, the acyl-coenzyme A (CoA) derivative. This specific acyl derivative is known to be an intermediate in the biosynthesis of long-chain polyunsaturated fatty acids (PUFAs) (102). PUFAs such as eicosapentaenoic acid are known to be produced by various *Shewanella* species (17). Moreover, PUFA biosynthetic genes from *Shewanella* have been identified by heterologous expression (73) and in *S. oneidensis* strain MR-1 via genome annotation (67).

To confirm the involvement of the PUFA pathway genes in the biosynthesis of compound I, a *pfaA* (annotated as a multidomain β-keto acyl synthase; gi 24373171, locus tag SO_1602) deletion mutant was constructed. When this mutant was tested for hydrocarbon biosynthesis, neither compound I nor any hydrocarbon product could be detected. Hydrocarbon biosynthesis was restored by the presence of the plasmid-encoded *pfaA* (Figure D.1 in Appendix D).

3.4.3 Genetic analysis of *ole* gene homologs.

We next sought to study the genes responsible for the condensation of a PUFA intermediate leading to the formation of compound I. A cluster of genes in *Shewanella oneidensis* MR-1 was observed to be homologous to genes (*ole*) previously implicated in head-to-head hydrocarbon biosynthesis (Friedman and Rude, WO2008/113041; Friedman and Da Costa, WO2008/147781). These were *Shewanella* proteins (gi

24373309, 24373310, 24373311, and 24373312), which were annotated in the GenBank database as a 3-oxoacyl-(acyl carrier protein) synthase III, an α/β-fold family hydrolase, a peptide hydrolase, and a 3-hydroxysteroid dehydrogenase/isomerase family protein, respectively. The first protein (gi 24373309) had 31% sequence identity to the Mlut_13230 protein identified by Beller *et al*. to be involved in a head-to-head condensation pathway in *M. luteus* (14). The two proteins gi 243733310 and gi 24373311 from *S. oneidensis* MR-1 resembled the N terminus and carboxy terminus, respectively, of the protein Mlut_13240 in *M. luteus*. Protein 4 (gi 24373312) showed 31% sequence identity to the Mlut_13250 protein of *M. luteus*. The bioinformatics data suggested that *S. oneidensis* MR-1 proteins gi 24373309 through 24373312 were, like the *M. luteus* proteins, involved in a head-to-head condensation reaction. This was investigated genetically to both confirm the genes' involvement and to investigate the effect of gene alteration on product formation.

The choice of S. *oneidensis* strain MR-1 allowed us to use well-established gene deletion methods to test if the *oleABCD* genes were involved in olefin biosynthesis (Figure 3.3a). In-frame deletions of the entire *ole* cluster, and of *oleC* individually, were generated. The gene deletion was verified using PCR. A 1.7-kb band corresponding to the *oleC*-containing gene cluster in the wild type became a 0.3-kb fragment in Δ*oleC*, resulting from deletion of the 1.5-kb *oleC* (Figure 3.3b). The complement showed both 0.3- and 1.7-kb bands representing the deleted gene region plus the full *oleC* present on the pOleC plasmid. Figure 3.3c shows the gas chromatograph of the region where compound I, produced by wild-type *S. oneidensis*, eluted at approximately 20.2 min.

The *oleC* mutant showed no detectable peak in this region. The complemented strain showed a restoration of the 20.2-min peak. The identity of the compound eluting at 20.2 min was confirmed by mass spectrometry. GC experiments were performed in triplicate. Similarly, the *oleABCD* deletion strain did not produce compound I (see Figure D.2 in Appendix D).

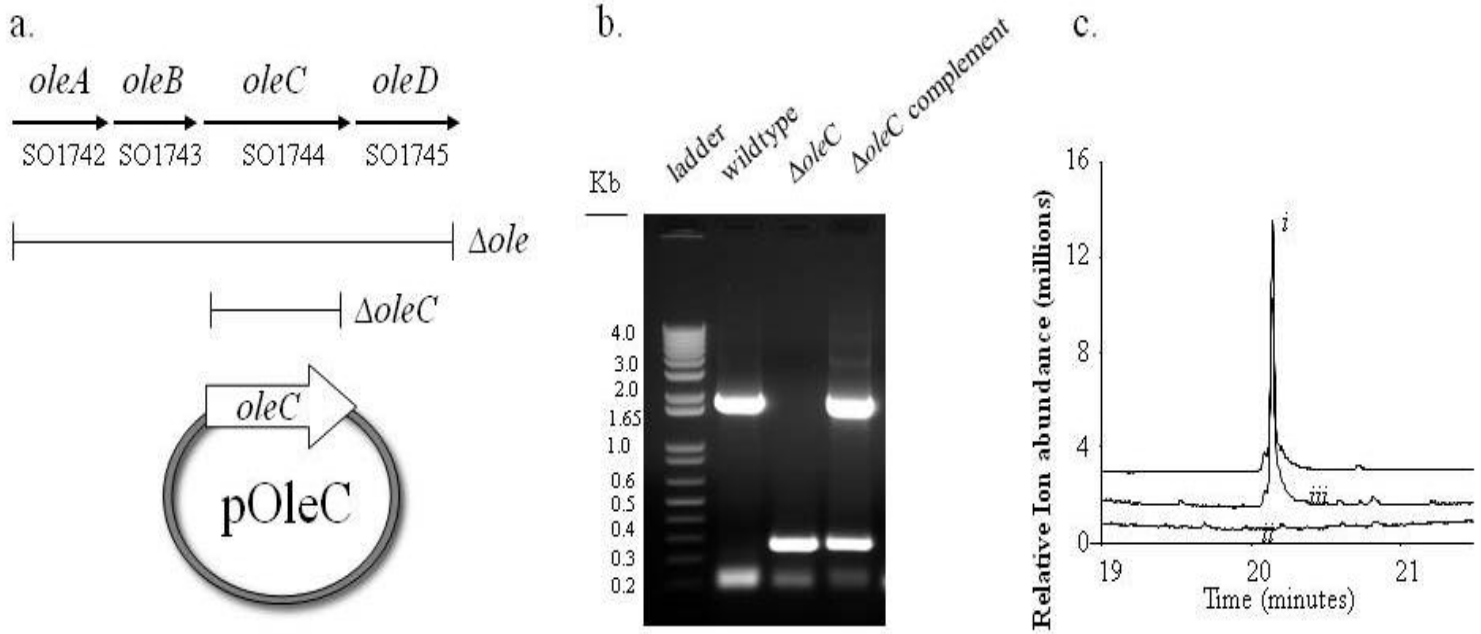

a.
oleA
oleB
oleC
oleD
SO1742
SO1743
SO1744
SO1745
Δole
ΔoleC
oleC
pOleC
b.
Kb
ladder
wildtype
ΔoleC
ΔoleC complement
4.0
3.0
2.0
1.65
1.0
0.6
0.5
0.4
0.3
0.2
c.
Relative Ion abundance (millions)
16
12
8
4
0
i
iii
ii
19
20
21
Time (minutes)

Figure 3.3 The *oleABCD* genes are required for long-chain olefin production by *S. oneidensis.* (a) Illustration of the *oleABCD* and *oleC* regions deleted and plasmid pOleC containing the *oleC* gene that complemented the *oleC* deletion. (b) DNA gel confirming gene deletion and complementation (primers used for analysis were SO1744CompF and SO1744CompR). (c) Gas chromatograph of solvent extracts from *S. oneidensis* wild-type (*i*), the *oleC* deletion mutant (*ii*), and the *oleC* mutant complemented with the pOleC plasmid (*iii*).

3.4.4 Formation of ketones and implications for the function of OleA.

The *S. oneidensis* MR-1 *oleC* deletion mutant did not produce a hydrocarbon, but it made another compound that was purified from a different distillation fraction than the hydrocarbon. The mass spectrum of the compound, compound III, had a parent ion of *m*/*z* 434. These data were consistent with a symmetrical molecule with eight double bonds and having the carbonyl functionality at the center of the hydrocarbon chain. Compound III was hydrogenated to produce a molecule with *m*/*z* 450 and showed an ion fragment of *m*/*z* 239. This confirmed the structure of compound III to be 3,6,9,12,19,22,25,28-hentriacontaoctaene-16-one. Compound III was not found in the *S. oneidensis* MR-1 *oleABCD* mutant.

Ketone products were also observed in an additional experiment involving heterologous *oleA* gene expression in *S. oneidensis* MR-1. The *oleA* gene homolog from *S. maltophilia* strain R551-3 was cloned into *S. oneidensis* strain MR-1. The

heterologous strain grew normally but produced a much wider range of nonpolar extractable products (Figure 3.4). The endogenous compound I was present and readily identified by GC retention time and mass spectrum and is shown in Figure 3.4 with an asterisk and the chemical formula $C_{31}H_{46}$. The recombinant *Shewanella* strain produced at least 17 additional long-chain compounds, of which 13 were monoketones (Figure 3.4). The chemical formulas are shown, indicating the degree of unsaturation of the hydrocarbon chains. All of the compounds are significantly more saturated than the endogenous $C_{31}H_{46}$ hydrocarbon, suggesting that the *Stenotrophomonas* OleA protein, unlike the *Shewanella* OleA protein, condenses fatty acids not derived from the polyunsaturated fatty acid pathway. The ketones were identified from their characteristic mass spectra; both the parent ions and ion fragments were consistent with these assignments. Moreover, the observation of a single major carbonyl ion, or two such ions of similar molecular weight, is consistent with the carbonyl functional group being present at the median carbon for odd-numbered chain lengths. This observation is consistent with these products arising from a head-to-head fatty acid condensation mechanism.

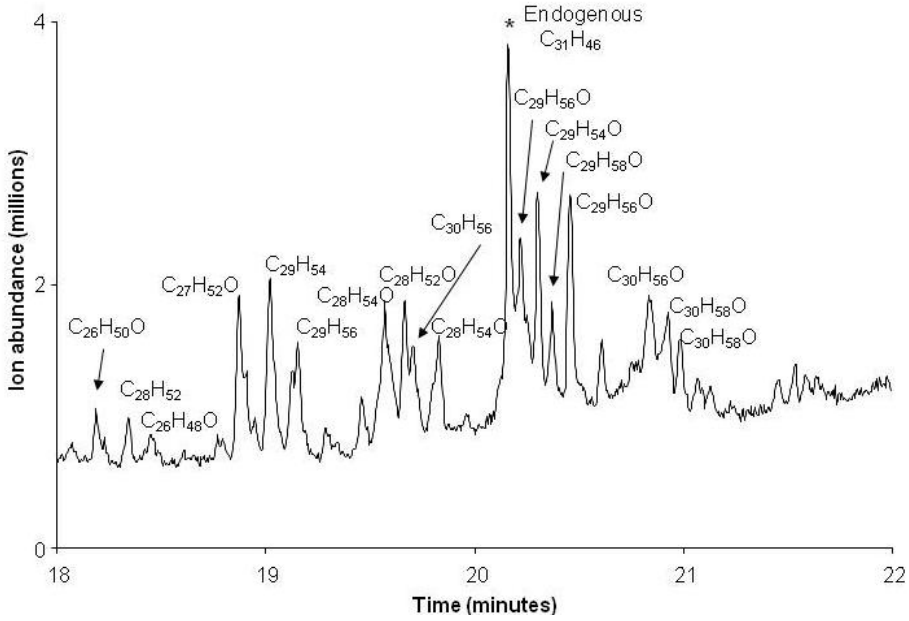


Figure 3.4 GC results for a solvent extract from recombinant *S. oneidensis* expressing the heterologous *S. maltophilia* OleA protein. Compounds were identified as hydrocarbons or ketones by mass spectrometry as described in the text and are designated by the molecular formula next to each major GC peak. The asterisk indicates compound I, which is endogenously produced by wild-type *S. oneidensis* MR-1.

The data shown in Figure 3.4 were striking because the native *Shewanella* only made a single endogenous $C_{31}H_{46}$ hydrocarbon, compound I. In contrast, *S. maltophilia* is known to produce a large number of different hydrocarbons with chain lengths of C_{26} to C_{30} (147), and the *S. maltophilia oleA* gene alone directed the formation of a much wider range of products in *Shewanella*. The observation here of diverse hydrocarbons and ketones has implications for the production of molecules for fuel or specialty

chemical applications via the heterologous expression of different *oleA* genes in *Shewanella*.

Ketone formation could potentially result from the OleA protein alone, and this would be consistent with the data presented here. OleA is in the thiolase superfamily, which catalyzes both decarboxylative and nondecarboxylative acyl group condensation reactions (62, 64). A nondecarboxylative thiolytic condensation would produce an intermediate that could give rise to ketones (Figure 3.5). Figure 3.5 shows the structure of the natively produced polyolefin, compound I. Hydrocarbons and ketones could both be derived from an intermediate generated by OleA, and that is consistent with reactions catalyzed by thiolase superfamily members, of which OleA is a member. Thioester cleavage could occur by the action of (i) OleA, (ii) a thioesterase, or (iii) spontaneous hydrolysis (50) to generate a β-keto acid (Figure 3.5C, compound II). β-Keto acids are known to be unstable and decarboxylate spontaneously (117). Spontaneous decarboxylation of β-keto acids in biological systems is well known and underlies the production of ketone bodies in mammalian liver (69). In the case of the *S. oneidensis oleC* mutant, intermediate compound II would be generated and decarboxylate to generate compound III, the observed ketone. When the OleA from *Stenotrophomonas* was expressed in *Shewanella*, a narrower specificity for the *Shewanella* enzymes could lead to the buildup of different intermediates that undergo hydrolysis and decarboxylation to yield the ketones. An alternative mechanism for the OleA-catalyzed condensation reaction has been proposed in the literature (14). Further studies will be required to discern between that and the role for OleA proposed here.

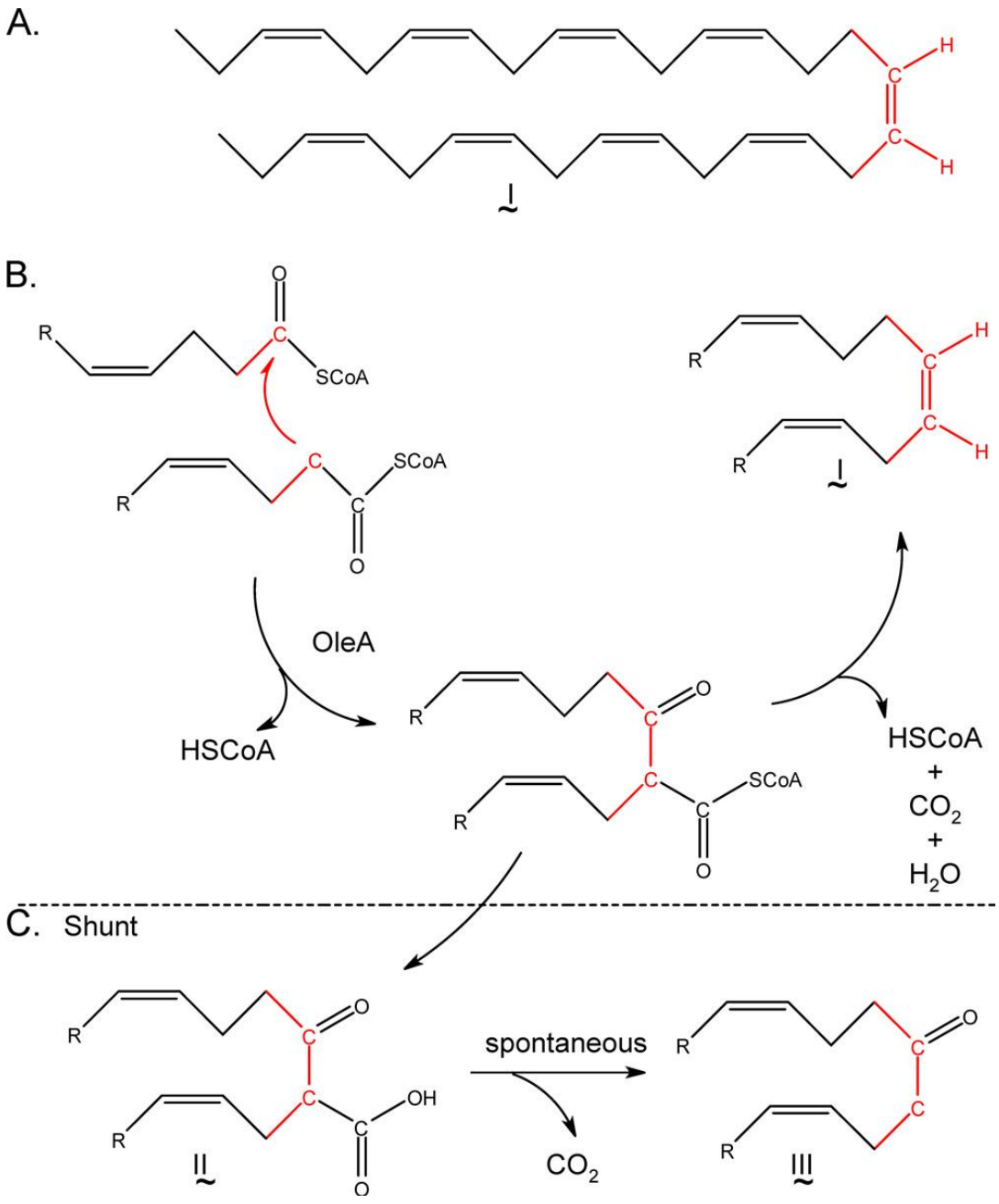


Figure 3.5 Product structures and proposed pathways in *S. oneidensis* MR-1 wild-type and mutant strains for head-to-head hydrocarbon and ketone formation, respectively. (A) Structure of compound I, identified as

described in the text. (B) Proposed role of OleA in the head-to-head biosynthetic pathway. (C) A proposed pathway to ketones in the presence of the OleA protein alone.

3.4.5 Potential role of an *ole* gene product(s) in cold adaption.

A hydrocarbon that appears to be identical to compound I was previously identified in a significant number of Antarctic bacterial isolates (110). The hypothesis that long-chain olefins might contribute to cold adaption was tested directly with *S. oneidensis* strain MR-1, which grows within the temperature range of 4 to 37°C (optimal growth at 30°C). The first observation in this study supporting the cold adaption hypothesis was that decreasing the growth temperature led to significant increases in the amount of compound I and compound III present in cells (Figure 3.6a).

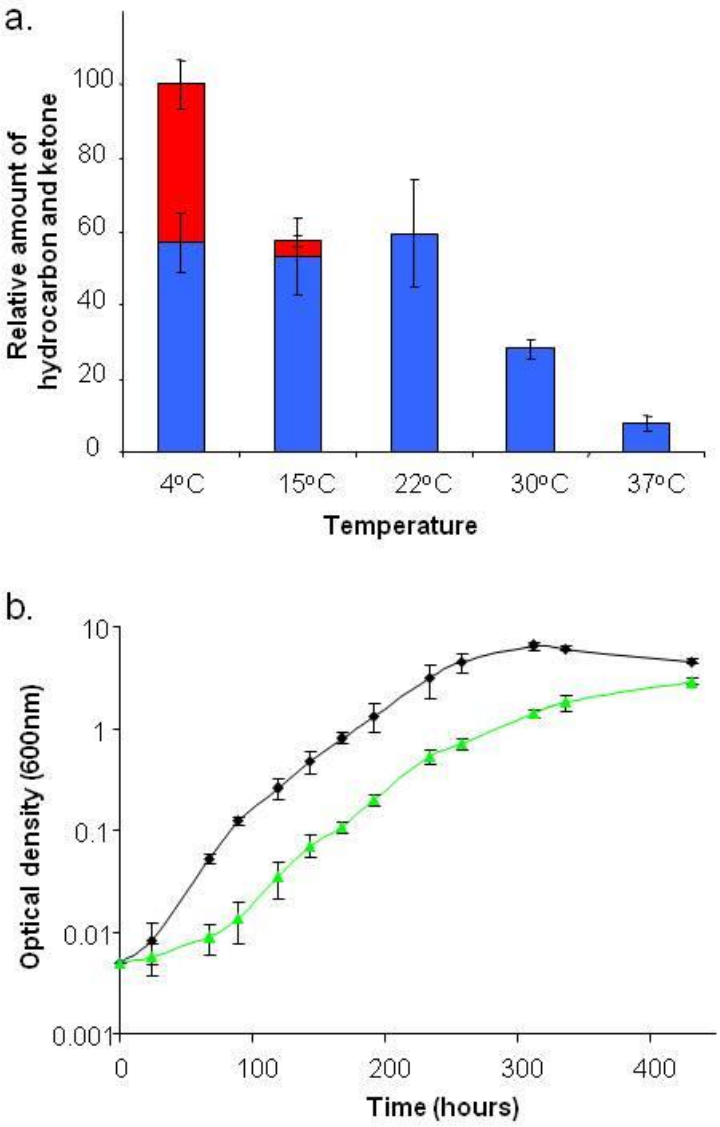


Figure 3.6 Long-chain polyunsaturated compounds as a function of growth temperature in *S. oneidensis* MR-1 wild type and an *oleABCD* deletion mutant. (a) Hydrocarbon (blue) and ketone (red) contents at different temperatures relative to the maximum observed (at 4°C). (b) Wild-type MR-1 (black) and the corresponding *oleABCD*-deficient mutant (green) were downshifted from 30°C to 4°C, and the cold temperature growth

curves are shown. Experimental points are average triplicate samplings from six treatments. Variation is shown as the standard deviation. See Appendix D Figure D.3 for growth curves of gene deletions and complements.

In other experiments, wild-type and olefin-deficient strains were grown at 30°C and then inoculated into medium at 4°C (Figure 3.6b). Although there was not much difference in the growth rate during exponential phase, the olefin-deficient *oleABCD* mutant strain showed a significantly longer lag phase prior to exponential growth (Figure 3.6b). When the *oleABCD* mutant was pregrown at 4°C, this lag in growth following transfer was not observed. These data suggested at least one role for long-chain olefins in facilitating growth following a shift to colder temperatures. We expect that the polyolefin would increase membrane fluidity and contribute to a maintenance of proper membrane function following a sudden decrease in temperature.

Structurally analogous long-chain alkadienes and alkatrienes are prominent in the lipids of marine photosynthetic eukaryotes, such as *Isochrysis galbana*, that grow at cold oceanic temperatures (126). They are also present, along with long-chain alkenones, in the lipid fractions of *Emiliania huxleyi* (126), a photosynthetic eukaryote which is so common that oceanic algal blooms of this organism are observable by satellite photographs (20). The mechanism of hydrocarbon formation in these eukaryotes remains open, but our findings here, coupled with ongoing genome sequencing of these organisms, may help provide insight. It is interesting that the

amount and degree of unsaturation of the long-chain hydrocarbons and alkenones increase with decreasing temperature (121). This suggests that long-chain hydrocarbons and ketones could be involved in cold adaption in both bacteria and eukaryotes.

While our paper was under review, we became aware of a new paper describing a $C_{31:9}$ hydrocarbon in a marine bacterium tentatively identified as a *Shewanella* sp. (S. Sugihara, R. Hori, H. Nakanowatari, Y. Takada, I. Yumoto, N. Morita, Y. Yano, K. Watanabe, and H. Okuyama, Lipids **45:**167-177, 2010). This hydrocarbon appears to be identical to the one in *S. oneidensis* strain MR-1 we describe here.

3.5 Addendum

3.5.1 OleA expression in *S. oneidensis*

With the knowledge that the *S. maltophilia oleA* gene could be expressed in various *S. oneidensis* backgrounds, we attempted to discover whether other *oleA* genes could be heterologously expressed in a *S. oneidensis oleABCD* deletion strain.

The *oleA* genes introduced into the *S. oneidensis ΔoleABCD* strain are listed in Table 3.3. The wildtype *Xanthomonas campestris, Colwellia psychrerythraea, Shewanella oneidensis, Shewanella amazonensis, Micrococcus luteus, Arthrobacter aurescens, Kocuria rhizophila,* and *Chloroflexus aurantiacus oleA* genes were amplified using primers and restriction sites listed in Table 3.4. The synthesized *oleA* genes of *Xanthomonas campestris, Congoribacter litoralis, gamma-proteobacteria*

NOR-5, *Xylella fastidiosa*, and *Pleseosystus pacifica* were optimized for *Shewanella oneidensis* codon usage using the Graphical Codon Usage Analyzer program (http://gcua.schoedl.de/sequential_v2.html) prior to being synthesized with SpeI and SacI restriction sites upstream and downstream of the sequence (for synthesized sequences, see appendix C). Transcription of *oleA* genes were identified using reverse transcriptase polymerase chain reaction (RT-PCR) analysis. Total RNA was extracted from actively growing cells using the Qiagen total RNA kit. Transcripts were identified using the Qiagen OneStep RT-PCR Kit. Primers used were RTPCRF1 (CTTAGAACCATCTACTGC) and RTPCRR1 (GTATCAGTAACCATACGATC).

Table 3.3: *oleA*s introduced into *S. oneidensis ΔoleABCD*

Organism origin	ketone(s) produced
Native organism *oleA*	
Shewanella oneidensis	C31:8
Shewanella amazonensis	C31:8
Colwellia psychrerythraea	C31:8
Stenotrophomonas maltophilia	C24-31
Chloroflexus aurantiacus	C28-31
Xanthomonas campestris	Not detected
Arthrobacter aurescens	Not detected
Micrococcus luteus	Not detected
Kocuria rhizophila	Not detected
Synthesized *oleA* (*S. oneidensis* codon specific)	
Xanthomonas campestris	Not detected
Congoribacter litoralis	Not detected
gamma proteobacteria NOR-5	Not detected
Plesiocystis pacifica	Not detected
Xylella fastidiosa	Not detected

We found that only five of the 14 *oleA* genes introduced into the *S. oneidensis* cluster-mutant background could produce ketones (Table 3.3). The *Shewanella*, C*olwellia*, and *Stenotrophomonas* OleA proteins all condensed fatty-acyl CoAs to produce ketones similar to the alkenes produced in their natural hosts. The ketones

derived from the *Chloroflexus* OleA protein differed from the *C. aurantiacus* natural product. This result is discussed later in Chapter 5.

Table 3.4: Primers used for Chapter 3.6

Primer	Sequence	Restriction site
XanoleAF	TTTAAAGGGCCC CCTGACGCCATGGCGCTACTTCGATGC	ApaI
XanoleAR	AAATTTGGATCCTCCCGGGATGGCCTGGCGCCTTCACCAA	BamHI
C.p.oleAF	ACTAGTATGAAATATTCCCGCGTCTTTATTA	SpeI
C.p.oleAR	GAGCTCTTACCACTTAAGCCCCAACATCATACAATTC	SacI
S.o.oleAF	ACTAGTTCATGAAATATTCCCGCGTATTTATTAATAG	SpeI
S.o.oleAR	GAGCTCGTTAAAGCATCGGCTAAGGCAGATAACAAC	SacI
S.a.oleAF	ACTAGTATGAAATATTCCCGCGTTTTTATCA	SpeI
S.a.oleAR	GAGCTCTTACCAGCGAAGCCCCAACATCATGCAG	SacI
C.a.oleAF	GGGCCCGCATGGTGATACATACCGAAGTACCGGCA	ApaI
C.a.oleAR	ACTAGTTGCACTGCATCACTCAAGGATCGCT	SpeI
M.l.oleAF	GGGCCCGTGACGAACGTGTCCGGCAACGCCAGCT	ApaI
M.l.oleAR	GTCGACTCACCATTCGATCTCGAGCATCGCGGT	SalI
A.a.oleAF	CTCGAGATGGCAGGGAATGCGACCTTCCGGCAC	XhoI
A.a.oleAR	GTCGACTCCGGGCCAGTTGGCGGCTACCAA	SalI
K.r.oleAF	GGGCCCTTGACCGGTAACTCCACCATCAGAAAC	ApaI
K.r.oleAR	GTCGACTCACCAGGCGATCTCCATGAGCGTG	SalI

While some OleA proteins were able to condense fatty-acyl CoAs successfully in their new background, others were not (Table 3.3). To help identify a cause for the lack of fatty-acyl condensation, we attempted to identify whether the *oleA* genes were transcribed. Total RNA was extracted from various *oleA*-constructs, reverse transcribed, and PCR amplified using primers RTPCRF1 and RTPCRR1. The resulting products were run on an agarose gel. As shown in figure 3.7, transcription of the *oleA* genes occurred in both the ketone producing and the ketone-lacking bacterial constructs. Control reactions containing the un-reverse transcribed PCR products revealed no plasmid or genomic DNA contamination in the mRNA samples (gel not shown). Protein gels failed to identify whether translation was occurring in the ketone-negative strains due to the fact that the strains do not appear to be over-expressing the proteins and we do not have an antibody for OleA protein to utilize in western blot analysis.

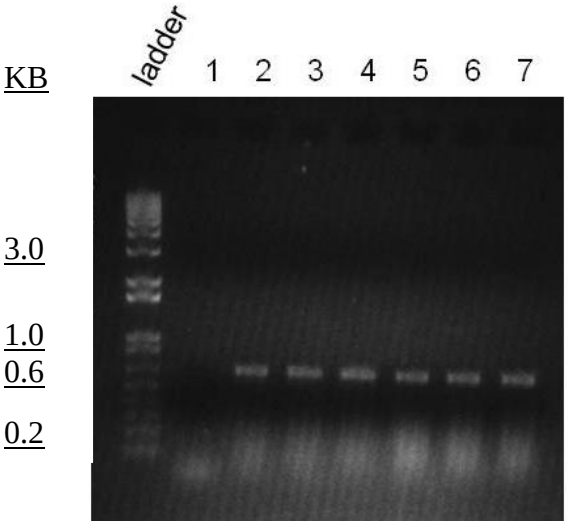


Figure 3.7 RT PCR reaction of *S. oneidensis ΔoleABCD* containing the pBBR1MCS2 with different *oleA* genes. Gel of the RT PCR reaction using primers RTPCRF1 and RTPCRR1. Lanes (1) *Shewanella ΔoleABCD*, (2) *Shewanella ΔoleABCD* with pOleA-*S.m.*, (3) *Shewanella ΔoleABCD* with pBBR1MCS2-*X. campestris oleA*, (4) *Shewanella ΔoleABCD* with pBBR1MCS2-*C. aurantiacus oleA*, (5) *Shewanella ΔoleABCD* with pBBR1MCS2-*X. fastidiosa oleA*, (6) *Shewanella ΔoleABCD* with pBBR1MCS2-*C. litoralis oleA*, and (7) *Shewanella ΔoleABCD* with pBBR1MCS2-*P. pacifica oleA.*

It remains unclear why certain strains produce hydrocarbons while others do not, though there are a few hypotheses. One hypothesis relates to protein miss-folding. J. Frias in the Wackett Laboratory has shown that the *Xanthomonas campestris* OleA condenses C10-C16 fatty acyl groups (personal communication). *Shewanella*

oneidensis does produce these fatty acyl compounds (167). Potentially the OleA proteins are not active because they are not properly folded in the *Shewanella* background.

Widespread head-to-head hydrocarbon biosynthesis in bacteria and role of OleA.

Sukovich, D.J., J.L. Seffernick, J.E. Richman, J.A. Gralnick, and L.P. Wackett. Applied and Environmental Microbiology. June 2010, volume 76, p. 3850-3862.

0099-2240/10/$12.00 doi: 10.1128/AEM.00436-10

Text and figures in Chapter 4 were previously published (except for Chapter 4.6) and are reproduced here with permission of the authors and the publisher.

Contributions to the writing of this chapter were made by each author. David Sukovich cultured, extracted, and analyzed each bacterial strain studied in the manuscript. David Sukovich also constructed each *Shewanella oneidensis* mutant and complement studied in this manuscript, as well as assisted Jennifer Seffernick in her bioinformatics studies. Jennifer Seffernick performed the majority of the bioinformatics experiments, as well as the comprehensive comparisons of the OleA protein sequences. Jack Richman was involved in hydrocarbon analysis and compound identification.

Chapter 4: Widespread Head-to-Head Hydrocarbon Biosynthesis in Bacteria and Role of OleA

4.1 Overview

Previous studies identified the *oleABCD* genes involved in head-to-head olefinic hydrocarbon biosynthesis. The present study more fully defined the OleABCD protein families within the thiolase, α/β-hydrolase, AMP-dependent ligase/synthase, and short-chain dehydrogenase superfamilies, respectively. Only 0.1 to 1% of each superfamily represents likely Ole proteins. Sequence analysis based on structural alignments and gene context was used to identify highly likely *ole* genes. Selected microorganisms from the phyla *Verucomicrobia*, *Planctomyces*, *Chloroflexi*, *Proteobacteria*, and *Actinobacteria* were tested experimentally and shown to produce long-chain olefinic hydrocarbons. However, different species from the same genera sometimes lack the *ole* genes and fail to produce olefinic hydrocarbons. Overall, only 1.9% of 3,558 genomes analyzed showed clear evidence for containing *ole* genes. The type of olefins produced by different bacteria differed greatly with respect to the number of carbon-carbon double bonds. The greatest number of organisms surveyed biosynthesized a single long-chain olefin, 3,6,9,12,15,19,22,25,28-hentriacontanonaene, that contains nine double bonds. *Xanthomonas campestris* produced the greatest number of distinct olefin products, 15 compounds ranging in length from C_{28} to C_{31} and containing one to three double bonds. The type of long-chain product formed was shown to be dependent on the *oleA* gene in experiments with *Shewanella oneidensis* MR-1 *ole* gene deletion mutants containing

native or heterologous *oleA* genes expressed in *trans*. A strain deleted in *oleABCD* and containing *oleA* in *trans* produced only ketones. Based on these observations, it was proposed that OleA catalyzes a nondecarboxylative thiolytic condensation of fatty acyl chains to generate a β-ketoacyl intermediate that can decarboxylate spontaneously to generate ketones.

4.2 Introduction

There is currently great interest in elucidating the means by which microbes produce nongaseous hydrocarbons for use as specialty chemicals and fuels (14, 95). While many details remain to be revealed, there appear to be several different pathways by which microbes biosynthesize long-chain hydrocarbons. The most studied of the pathways (35) involves the condensation of isoprene units to generate hydrocarbons with a multiple of five carbon atoms (C_{10}, C_{15}, C_{20}, etc.). A more obscure biosynthetic route is a reported decarbonylation of fatty aldehydes to generate a C_{n-1} hydrocarbon chain (38). A third mechanism that has received some attention is what has been denoted head-to-head condensation of fatty acids (2-5, 14, 159). In this pathway, the hydrocarbons are described to arise from the formation of a carbon-to-carbon bond between the carboxyl carbon of one fatty acid and the α-carbon of another fatty acid (3). This condensation results in a particular type of hydrocarbon with chain lengths of C_{23} to C_{33} and containing one or more double bonds. One double bond involves the median carbon in the chain at the point of fatty acid condensation. An example of this overall biosynthetic pathway leading to the formation of specific C_{29} olefinic hydrocarbon

isomers from fatty acid precursors has been demonstrated *in vivo* (154, 159, 160) and *in vitro* (4, 14).

The condensation, elimination of carbon dioxide, and loss of the other carboxyl group oxygen atoms likely require multiple enzyme-catalyzed reactions. Recent patent applications by L. Friedman *et al*. describe a role for three or four proteins in this biosynthetic pathway (18 September 2008, WO2008/113041; 4 December 2008, WO2008/147781). Most recently, Beller *et al*. demonstrated the requirement for three genes from *Micrococcus luteus* in the biosynthesis of long-chain olefins (14). That study also demonstrated *in vitro* production of olefins by recombinant proteins in the presence of crude cell extracts from *Escherichia coli*. In another study, *Shewanella oneidensis* strain MR-1 was shown to produce a head-to-head hydrocarbon (150). A cluster of four genes, *oleABCD*, was shown to be involved in olefin biosynthesis by that organism.

While genetic and biochemical data have provided evidence for Ole proteins producing long-chain olefins in *M. luteus* and *S. oneidensis*, there are many outstanding details of the biosynthesis that remain to be elucidated. Moreover, the extent to which microbial and other species produce head-to-head olefins is unclear. A recent patent application by Friedman and Rude (WO2008/113041) presented tables listing genes homologous to the *ole* genes described by Beller *et al*. (14). However, the homologs identified included genes from mouse and tree frog, organisms not known to produce head-to-head hydrocarbons. Additionally, hydrocarbon biosynthetic genes from *Arthrobacter* sp. FB24 were claimed, and that strain was later shown not to produce hydrocarbons under identical conditions for which other *Arthrobacter* strains did (51).

In that context, the present study closely examined the protein sequence families of Ole proteins and the configurations of putative *ole* genes within genomes to identify those most likely to be involved in head-to-head hydrocarbon biosynthesis. This was followed by experimental testing for the presence of long-chain head-to-head hydrocarbons in representative bacteria from diverse phyla. This study also found that, of closely related bacteria, some produce head-to-head hydrocarbons and others do not.

A previous publication investigated *in vitro* olefin biosynthesis from myristyl-coenzyme A (CoA) (14). That study showed ketone and olefin biosynthesis *in vitro* and proposed a mechanism requiring the participation of ancillary proteins not encoded in the *oleABCD* gene cluster. The mechanism proposed fatty acyl oxidation to generate a β-keto acid that is the substrate for the OleA protein.

In fact, different mechanisms have been suggested previously for the biosynthesis of head-to-head olefins ((3); Friedman and Rude, WO2008/113041; Friedman and Da Costa, WO2008/147781), and different roles for the OleA protein have been proposed ((14); Friedman and Rude, WO2008/113041; Friedman and Da Costa, WO2008/147781). It is not possible to deduce the olefinic biosynthetic pathway or individual reaction types based on protein sequence alignments alone, because this pathway is unique, differing markedly from isoprenoid or decarbonylation hydrocarbon biosynthesis pathways. Moreover, the individual Ole proteins are each homologous to proteins that collectively catalyze diverse reactions. In that context, we (i) initiated a more detailed study of the Ole protein superfamilies, (ii) identified likely olefin (*ole*) biosynthetic genes out of thousands of homologs, (iii) experimentally tested bacteria

from different phyla for long-chain olefins, (iv) developed insights into the role of OleA in head-to-head olefin biosynthesis, and (v) propose an alternative mechanism for head-to-head condensation of fatty acyl groups.

4.3 Methods and Materials

4.3.1 Strains and culture conditions.

Wild-type and recombinant bacteria used in this study are listed in Table 4.1. All organisms, including recombinant strains, were grown aerobically in 50-ml culture flasks on a rotary shaker at 225 rpm, except for *Geobacter* strains, which were grown in a 100-ml anaerobic culture flask flushed for 30 min with a nitrogen/carbon dioxide gas mix prior to culture inoculation (129). All organisms were grown at 30°C (12, 22, 85, 90, 129) except for *Shewanella amazonensis* (35°C) (166), *Shewanella frigidimarina* (22°C) (167), *Opitutaceae bacterium* TAV2 (22°C) (146), *Brevibacterium fuscum* (22°C) (85), *Colwellia psychrerythraea* (4°C) (176), *Chloroflexus aurantiacus* (55°C) (120), and all *Escherichia coli* strains (37°C) (133). The organisms were allowed to achieve stationary phase prior to hydrocarbon extraction and analysis. All organisms were grown in Luria broth (Difco) (85, 90, 166, 167) except for *S. frigidimarina* (17), *C. psychrerythraea* (176), and *Planctomyces maris* (Marine broth; Difco) (171), *Geobacter* species (Geobacter medium; DSMZ) (129), *C. aurantiacus* (*Chloroflexus* medium; DSMZ) (120), *Opitutaceae bacterium* TAV2 (R2A medium; Difco) (T. D. Schmidt, personal communication), and *X. campestris* (nutrient broth; Difco) (22).

Table 4.1. Organisms, plasmids, and primers used in Chapter 4

Organisms, plasmid, or primer	Genotype, relevant characteristic(s), or sequence	Source or reference
Genetically modified organisms		
S. oneidensis oleA	*S. oneidensis MR-1, oleA;* hydrocarbon minus	This study
S. oneidensis ole	*S. oneidensis MR-1, oleABCD;* hydrocarbon minus	150
E. coli UQ950	*E. coli DH5 (pir) host for cloning, F(argF-lac)169 80dlacZ58(M15) glnV44(AS)rfbD1 gyrA96 (NalR) recA1 endA1 spoT1 thi-1 hsdR17 deoR pir*	133
E. coli WM3064	Donor strain for conjugation, *thrB1004 pro thi rpsL hsdSlacZM15 RP4-1360 (araBAD)567 dapA1341::erm pir (wt)*	133
Plasmids		
pSMV3	9.5-kb vector; Kmr version of pSMV8; *lacZ sacB*	10
pSMV3-*oleA*	0.9-kb fusion PCR fragment containing *oleA* cloned Into SpeI/BamHI site of pSMV3; used to make *S. oneidensis oleA* strain	This study
pBBR1MCS-2	5.1-kb broad-host-range plasmid; *lacZ; Kmr*	89
pOleA-S.m.	1.1-kb PCR fragment containing *S. maltophilia oleA,* cloned into SpeI/SacI site of pBBR1MCS-2	150
pOleA	1.9-kb PCR fragment containing *S. oneidensis oleA,* cloned into SpeI/SacI site of pBBR1MCS-2	This study
Primers		
oleASoF1	ACTAGTTACATGTGCGTTTATTGCAACTGGCC	
oleASoR1	CCAGAGATATAGAGGCGCGAGGCGAGATTC	
oleASoF2	GGTCTCATGGCACACGATCAAGGCTTTTTAC	
oleASoR2	GGATCCCCAACAAATCAGTGTCGGCACC	
SooleACompF	ACTAGTTACATGTGCGTTTATTGCAACTGGCC	
SooleACompR	GAGCTCGTTAAAGCATCGGCTAAGGCAGATAACAA	

a The following wild-type organisms were also used in this study (reference numbers) shown in parentheses: *Shewanella oneidensis* MR-1 (49, 167), *Shewanella putrefaciens* CN-32 (78), *Shewanella baltica* OS185 (177) *Shewanella frigidimarina* NCIMB 400 (17), *Shewanella amazonensis* SB2B (166), *Shewanella denitrificans* OS217 (19), *Colwellia psychrerythraea* 34H (37), *Geobacter bemidjiensis* Bem (109), *Geobacter sulfurreducens* PCA (23), *Opitutaceae bacterium* TAV2 (146), *Planctomyces maris* DSM8797 (12), *Chloroflexus aurantiacus* J-10-fl (120), *Kocuria rhizophila* DC2201 (90), *Brevibacterium fuscum* ATCC 15993 (132), *Xanthomonas campestris* (165), *Vibrio furnissii* M1 (116), *and E. coli* K-12 (lab supply).

4.3.2 Hydrocarbon and ketone extraction, chromatography, and characterization.

Early-stationary-phase cultures were extracted as previously described (170). Briefly, both cells and medium from a 50-ml bacterial culture that had reached stationary phase were extracted using a mixture of spectrophotometric-grade methanol (Sigma-Aldrich), high-performance liquid chromatography (HPLC)-grade chloroform (Sigma-Aldrich), and distilled water in a 5:5:4 ratio. The resulting nonpolar phase was collected and dried under vacuum. Evaporated residue was recovered in 1 ml of methyl-*tert*-butyl ether (MTBE), applied to a 4.0-g silica gel column (Sigma-Aldrich), and eluted with 35 ml of HPLC-grade hexanes (Fischer Scientific), followed by 35 ml of MTBE and 25 ml of HPLC-grade ethyl acetate (Sigma). Each solvent fraction was concentrated and subjected to GC-mass spectrometry (MS) analysis using an HP6890 gas chromatograph connected to an HP5973 mass spectrometer (Hewlett Packard, Palo Alto, CA). GC conditions consisted of the following: helium gas, 1 ml/min; HP-1ms column (100% dimethylpolysiloxane capillary, 30 m by 0.25 mm by 0.25 μm); temperature ramp, 100 to 320°C, 10°C/min, with a 5-min hold at 320°C. The mass spectrometer was run in electron impact mode at 70 eV and 35 μA. Alkene and ketone products were identified from the parent ions and corresponding fragmentation patterns. Major compounds were further analyzed by hydrogenation over palladium on carbon (Sigma-Aldrich) and observation of the corresponding increase in mass to confirm the number of double bonds present.

4.3.3 Gene deletion and *oleA* gene complementation.

All deletion strains, plasmids, and primers used are listed in Table 4.1. Gene deletions were made using homologous recombination between flanking regions of *oleA* cloned into a suicide vector, pSMV3 (133). Briefly, by using oleASoF1, oleASoR1, oleASoF2, and oleASoR1, the upstream and downstream regions surrounding the gene were cloned using the restriction sites SpeI and BamHI into the suicide vector in a compatible *E. coli* cloning strain (UQ950) (133). This plasmid was transformed into an *E. coli* mating strain (WM3064) (133) and then conjugated into MR-1. While *E. coli* was commonly grown at 37°C, when *S. oneidensis* was present cells were incubated at 30°C. The initial recombination event was selected for by resistance to kanamycin. Cells containing the integrated suicide vector grew in the absence of selection overnight at 30°C and then were plated onto LB plates containing 5% sucrose (133). Cells retaining the suicide vector were unable to grow due to the activity of SacB, encoded on the vector, while cells that underwent a second recombination event formed colonies. Colonies were then screened by PCR to determine strains containing the deletion. The *oleABCD* gene cluster deletion of *S. oneidensis* MR-1 was created as described previously (150).

Complementation of the *S. oneidensis oleA* mutant was performed using the pBBR1MCS-2 expression vector (89) and the endogenous *lac* promoter (which is constitutive in MR-1 due to the absence of *lacI*). Primers oleASoFcomp and oleASoRcomp containing SacI and SpeI restriction sites were designed for the regions flanking the ends of *oleA*. Resulting PCR products were ligated into the Strataclone

cloning system (Agilent Technologies), followed by digestion and ligation of the product into the pBBR1MCS-2 expression vector. The *Stenotrophomonas maltophilia oleA* gene was introduced into pBBR1MCS-2 as described previously (150). Constructs were introduced into *E. coli* WM3064 prior to conjugation with the *oleA* deletion, the *ole* cluster deletion, or wild-type MR-1 strains. All constructs were verified through PCR and sequencing analysis. Following conjugation, all constructs were maintained using 50 µg/ml kanamycin.

4.3.4 Identification of *oleABCD*-containing organisms.

The *oleABCD* genes in *S. oneidensis* MR-1 were used to find homologous gene clusters in the GenBank nonredundant database using the BLAST algorithm (6). Subsequently, the OleA homologs in *Stenotrophomonas maltophilia* strain R551-3 (gi 194346749), *Arthrobacter aurescens* TC1 (gi 119962129), and *Micrococcus luteus* NCTC 2665 (gi 239917824) were used as additional queries to the database. Other homologous thiolases were identified. The genome context of each of these thiolases was investigated and allowed for the assembly of a set of organisms with either a four- or three-gene cluster, encoding OleA, -B, -C, and -D protein domains. A lack of clustering did not preclude the existence of the pathway in an organism. Therefore, those organisms that lacked clustered genes were searched for *oleBCD* genes in other locations of their genome. Organisms with clustering of at least two identifiable *ole* homologs and which had all four genes in their genome were included as potential hydrocarbon producers and investigated experimentally.

4.3.5 Superfamily sequence identification and alignments.

The PSI-BLAST algorithm with default conditions (134) was used with *S. oneidensis* MR-1 or *A. aurescens* TC1 Ole protein sequences as queries. Thousands of homologous sequences were found. The sequence and catalytic diversity within each superfamily were sufficiently broad that standard sequence alignment tools did not align amino acid residues that are known to comprise the active sites in proteins for which X-ray structures are available (34, 60, 62, 64, 75, 79, 107, 122, 145, 154, 168, 174). Thus, to properly align Ole protein sequences with other proteins in their respective superfamilies, it was necessary to generate structure-based alignments. For each OleABCD alignment, 6 to 10 homologous proteins from previously described high-resolution X-ray structures were structurally superimposed, using the Match command in Chimera (100). Conserved residues within each superfamily of homologs were derived from the literature (34, 60, 62, 64, 75, 79, 107, 122, 145, 154, 168, 174), and their locations were plotted onto the protein backbone to confirm alignments. Sequence alignments based on the structure alignments were exported. Multiple sequence alignments of each of the OleABCD families were made with 41 to 55 sequences, using ClustalW (30). In the case of the OleA alignments, 14 OleA homologs with genes that did not cluster with *oleBCD* genes were also included for sequence comparison purposes. A profile-profile alignment between the structural superfamily alignments and the family sequence alignments was produced, using ClustalW (30). These superfamily Ole sequence alignments were viewed in Chimera with the overlaid superfamily crystal structures linked to the alignments so that the positions of residues in the alignment

could be viewed (100). For OleBC fusion proteins, the individual domains were used for alignments with the appropriate superfamilies.

4.3.6 Analysis of protein superfamilies.

The Superfamily database (56) was searched with each of the *S. oneidensis* MR-1 Ole protein sequences. The superfamilies identified by these searches confirmed assignments made independently as described above. The number of distinct proteins in each superfamily was kindly provided from the Superfamily database (Derek Wilson, personal communication). The relevant superfamily categories in the Superfamily database are thiolase-like, α/β-hydrolases, acetyl-CoA synthetase-like, and NAD(P) Rossman fold domains. It should be noted that the NAD(P) Rossman fold domains superfamily, as listed in the Superfamily database, consists of a number of families in which the proteins share the ability to bind NAD(P), and it contained a total of 136,722 proteins as of 1 February 2010. These proteins have a second domain involved in substrate binding and which confers the catalytic residues. These differentiations are made in the Superfamily database at what is denoted as the family level. The OleD proteins belong to the tyrosine-dependent oxidoreductase domain family. This set was used for our analysis and was equivalent to the set given superfamily status by Jornvall *et al*. and described as the short-chain dehydrogenase/reductase superfamily (79).

4.3.7 Network clustering of OleABCD proteins.

Network clustering of each of the OleABCD proteins was analyzed using previously described procedures (106, 137). This method was used to make an all-by-all

BLASTp library for each of the OleABCD proteins using sequences from 15 organisms. The sequences used were (i) *S. oneidensis* MR-1, OleA gi24373309, OleB gi24373310, OleC gi24373311, and OleD gi24373312; (ii) *Shewanella amazonensis SB2B*, OleA gi119774319, OleB gi119774320, OleC gi119774321, and OleD gi119774322; (iii) *Shewanella baltica* OS185, OleA gi153000075, OleB gi153000076, OleC gi153000077, and OleD gi153000078; (iv) *Shewanella denitrificans* OS217, OleA gi91792727, OleB gi91792728, OleC gi91792728, and OleD gi91792730; (v) *Shewanella frigidimarina* NCIMB 400, OleA gi114562543, OleB gi114562544, OleC gi114562545, and OleD gi114562546; (vi) *Shewanella putrefaciens* CN-32, OleA gi146292545, OleB gi 146292546, OleC gi146292547, and OleD gi146292548; (vii) *Colwellia psychrerythraea* 34H, OleA gi71279747, OleB gi71279056, OleC gi71281286, and OleD gi71280771; (viii) *Geobacter bemidjiensis* Bem, OleA gi197118484, OleB gi197118483, OleC gi197118482, and OleD gi197118481; (ix) *Planctomyces maris* DSM 8797, OleA gi149174448, OleB gi149178001, OleC gi149178707, and OleD gi149178706; (x) *Opitutaceae bacterium* TAV2, OleA gi225164858, OleB gi225164858, OleC gi225155590, and OleD (no cluster); (xi) *Stenotrophomonas maltophilia* R551-3, OleA gi194363945, OleB gi194363946, OleC gi194363948, and OleD gi194363949; (xii) *Xanthomonas campestris* pv. *campestris* strain B100, OleA gi188989629, OleB gi188989631, OleC gi188989633, and OleD gi188989637; (xiii) *Chloroflexus aurantiacus* J-10-fl, OleA gi163849058, OleB gi163849062, OleC gi163849060, and OleD gi163849059; (xiv) *Arthrobacter aurescens* TC1, OleA gi119962129, OleB gi119960515 (residues 1 to 310), OleC gi119960515 (residues 389 to 921), and OleD gi119962242; (xv) *Arthrobacter chlorophenolicus* A6, OleA

gi220911225, OleB domain gi220911226 (residues 1 to 296), OleC gi220911226 (residues 370 to 927), and OleD gi220911227; (xvi) *Kocuria rhizophila* DC2201, OleA gi184200698, OleB gi184200697 (residues 1 to 312), OleC gi184200697 (residues 392 to 909), and OleD gi184200696; (xvii) *Micrococcus luteus* NCTC 2665, OleA gi239917824, OleB gi239917825 (residues 1 to 330), OleC gi239917825 (residues 439 to 978), and OleD gi239917826. From these sequences, a network diagram was created. The nodes represent protein sequences and the edges represent a BLAST linkage that connects the two proteins. A shorter edge represents a lower e-score (greater relatedness). Expectation values from e^{-2} to e^{-200} were analyzed for connectivity and divergence, respectively, of OleA, -B, -C, and -D protein sequence clusters.

4.4 Results and Discussion

4.4.1 Ole protein superfamily analysis.

Thousands of sequences were identified as being homologous to each of the OleA, OleB, OleC, and OleD sequences from *S. oneidensis* MR-1 (Table 4.2). OleA is homologous to members of the thiolase superfamily, also known as the condensing enzyme superfamily. The sequence relatedness between different OleA proteins and FabH, a thiolase superfamily member, has been noted previously even though sequence identities of OleA to FabH and other superfamily members are generally low, in the range of 20 to 30% ((14); Friedman and Rude, WO2008/113041; Friedman and Da Costa, WO2008/147781). OleB is a member of the α/β-hydrolase superfamily. OleC is a

member of the AMP-dependent ligase/synthase superfamily, also known as the acetyl-CoA synthetase-like superfamily. OleD is a member of the short-chain dehydrogenase/reductase superfamily.

Table 4.2. Ole proteins superfamilies, homolog characteristics, number of Ole proteins and homologs

Proteins in this study	Superfamily name (alternative names)	Enzymatic activities and biological functions in the superfamily	# of homologs in superfamily	# of Ole proteins identified
OleA	Thiolase (condensing enzymes)	Acyl-ACP synthase, thiolase (degradative), thiolase (biosynthetic), 3-hydroxyl-3-methylglutaryl-CoA synthase, fatty acid elongase, stage V sporulation protein, 6-methylsalicylate synthase, *Rhizobium* nodulation protein NodE, chalcone synthase, stilbene synthase, naringenin synthase, β-ketosynthase domains of polyketide synthase	13,586	69
OleB	α/β–Hydrolase	Esterase, haloalkane dehalogenase, protease, lipase, haloperoxidase, lyase, epoxide hydrolase, enoyl CoA hydratase/isomerase, MhpC C-C hydrolase (carbon-carbon bond cleavage)	67,923	69
OleC	AMP-dependent ligase/synthase (LuxE; acyl-adenylate/thioester forming, Acetyl-CoA synthetase-like)	Firefly luciferase, nonribosomal peptide synthase, acyl-CoA synthase (AMP forming), 4-chlorobenzoate:CoA ligase, acetyl-CoA synthetase, o-succinylbenzoic acid-CoA ligase, fatty acyl ligase, acetyl-CoA synthetase, 2-acyl-glycerophospho-ethanolamine acyl transferase, enterobactin synthase, amino acid adenylation domain, dicarboxylate-CoA ligase, crotonobetaine/carnitine-CoA ligase	19,660	69
OleD	Short-chain dehydrogenase/ reductase	Nucleoside-diphosphate sugar epimerase/ dehydratase/reductase, aromatic diol dehydrogenase, steroid dehydrogenase/ isomerase, sugar dehydrogenase, acetoacetyl-CoA reductase, 3-oxoacyl-ACP reductase, alcohol dehydrogenase, carbonyl reductase, 4-α-carboxysterol-C3-dehydrogenase/C4-decarboxylase, flavonol reductase, cinnamoyl CoA reductase, NAD(P) dependent cholesterol dehydrogenase,	25,454	69

Figure 4.1 shows conserved regions of a structure-based multiple sequence alignment for each of the OleA, -B, -C, and -D proteins with three of their respective superfamily members. Figure 4.1 focuses on regions containing catalytically important residues that are highly conserved among the homologous proteins. A more detailed set of alignments is available in the Figure E.1 in Appendix E. The superfamily members shown in Figure 4.1 were selected to represent proteins serving quite different biological functions. So, while OleABCD are clearly seen to contain critical catalytic residues of each respective superfamily, a precise prediction of the biochemical reaction catalyzed

is difficult due to the enormous functional diversity found within each Ole protein's superfamily.

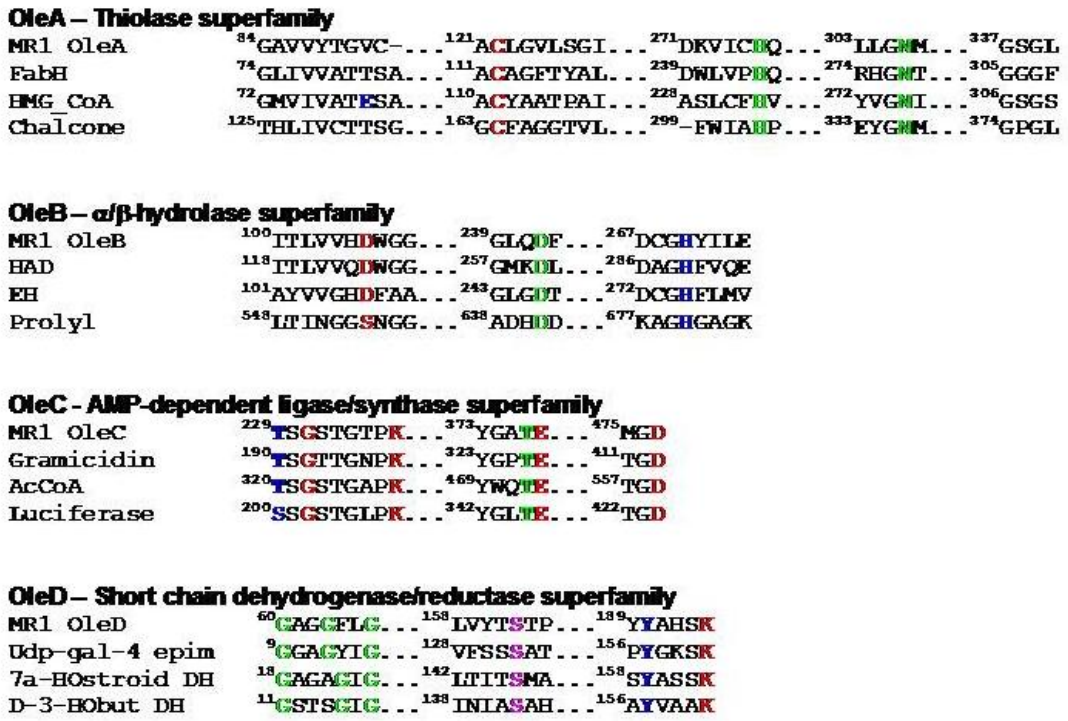


Figure 4.1. Structure-based amino acid sequence alignments of OleA, OleB, OleC, and OleD from *S. oneidensis* MR-1 (denoted MR1 in the figure) with highly conserved regions of proteins catalyzing divergent reactions from each respective superfamily. Accession numbers or PDB identifiers of the proteins used are as follows. OleA thiolase superfamily: OleA from *Shewanella oneidensis* MR-1 (gi 24373309), β-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli* (1EBL), 3-hydroxy-3-methylglutaryl-CoA synthase (HMG_CoA) from *Staphylococcus aureus* (1XPK), and chalcone synthase (Chalcone) from *Medicago sativa* (1CGZ); blue residues indicate the glutamate that abstracts a proton to

produce a carbanion for the nondecarboxylative condensation reaction, red indicates the absolutely conserved cysteine of the superfamily that forms a covalent bond with the substrate, and green residues are involved in formation of an oxyanion hole. OleB α/β-hydrolase superfamily: OleB from *Shewanella oneidensis* MR-1 (gi 24373310), haloalkane dehalogenase (HAD) from *Xanthobacter autotrophicus* GJ10 (1B6G), epoxide hydrolase (EH) from *Agrobacterium radiobacter AD1* (1EHY), and prolyloligopeptidase (prolyl) from porcine brain (1H2W); red residues indicate the catalytic nucleotide (Ser, Asp, or Cys in the whole superfamily), green indicates the general acid, and blue indicates the conserved histidine that activates water. OleC AMP-dependent ligase/synthetase superfamily: OleC from *Shewanella oneidensis* MR-1 (gi 24373311), gramicidin synthetase (gramicidin) from *Brevibacillus brevis* (1AMU), acetyl-CoA synthetase (AcCoA) from *Saccharomyces cervisiae* (1RY2), and luciferase from the Japanese firefly (2D1Q); red indicates absolute conservation in the three consensus regions identified by Conti *et al*. ([STG]-[STG]-G-[ST]-[TSE]-[GS]-x-[PALIVM]-K, [YFW]-[GASW]-x-[TSA]-E, [STA]-[GRK]-D) (34); blue and green indicate Thr/Ser residues thought to be involved in binding the phosphoryl group in ATP and AMP. OleD short chain dehydrogenase/reductase superfamily: OleD from *Shewanella oneidensis* MR-1 (gi 24373312), UDP-galactose-4-epimerase (Udp-gal-4 epim) from humans (1EK6), 7-α-hydroxysteroid dehydrogenase (7a-HOstroid

DH) from *Escherichia coli* (1AHH), and D-3-hydroxybutyrate dehydrogenase (D-3-HObut DH) from *Pseudomonas fragi* (3ZTL); blue identifies the tyrosine anion that abstracts the proton from the substrate, red is a lysine that stabilizes the tyrosine anion, green is a glycine-rich region involved in cofactor $NAD(P)^+$ binding, and pink is the serine that orients the substrate or stabilizes intermediates.

The superfamilies to which Ole proteins belong each consist of between 10^4 and 10^5 curated protein members that have been identified for inclusion in the Superfamily database (Table 4.2). The present study suggested that only 0.1% to 1% of the proteins in each superfamily represents Ole proteins that participate in head-to-head hydrocarbon biosynthesis. The identification of these Ole proteins in the sequenced genomes of microorganism is discussed below.

4.4.2 Protein relatedness and gene organization used to identify *ole* genes among thousands of homologs.

Only a limited number of bacteria to date have been found to produce long-chain olefinic hydrocarbons. For example, among 10 *Arthrobacter* strains tested, 6 produced long-chain olefinic hydrocarbons and 4 did not (51). Of three closely related *Arthrobacter* strains for which genome sequences were available, two (*A. aurescens* TC1 and *A. chlorophenolicus* A6) were shown to produce hydrocarbons and one (*Arthrobacter* sp. FB24) was devoid of long-chain olefinic hydrocarbons. The FB24 strain that did not produce hydrocarbons contained *ole* gene homologs, but the percent identity was much lower and the genes were distributed within the genome differently.

By examining such divergences, a strategy for identifying highly likely *ole* genes was developed as described in the Materials and Methods section. The example below is illustrative.

A putative *oleA* gene region was identified in *Geobacter bemidjiensis* Bem that, after translation, showed 58% amino acid sequence identity to the OleA protein in *S. oneidensis* MR-1. Directly downstream from the *G. bemidjiensis* Bem *oleA* gene, *oleBCD* gene homologs were present in a configuration that mirrored that of *S. oneidensis* MR-1 (Figure 4.2). An OleA homolog was also identified in *Geobacter sulfurreducens* PCA. It showed significantly lower amino acid sequence identity, 28%, to the OleA from *S. oneidensis* MR-1. It lacked flanking *ole* gene neighbors. Closer examination of the two genomes revealed that the OleA homolog in *G. sulfurreducens* PCA was encoded by a gene region that matched a gene region with identical synteny in *G. bemidjiensis* Bem. This same gene region was also identified in *S. oneidensis* MR-1. From this analysis, it was concluded that the OleA homolog in *G. sulfurreducens* PCA was not involved in a head-to-head condensation reaction and it was suggested that this organism was genetically incapable of making head-to-head olefins. Cells of *G. sulfurreducens* PCA were tested experimentally for the presence of long-chain olefinic hydrocarbons. Hydrocarbons were absent under identical growth conditions in which they were present in *G. bemidjiensis* Bem (see the discussion on hydrocarbon identification below).

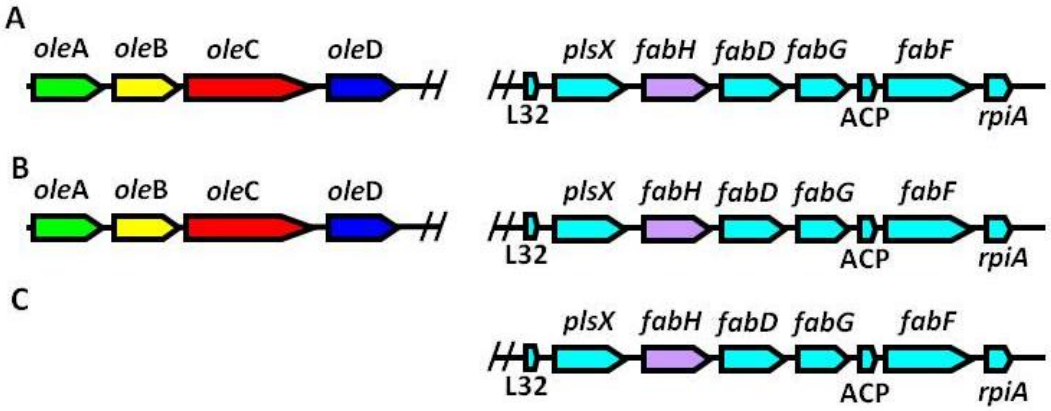
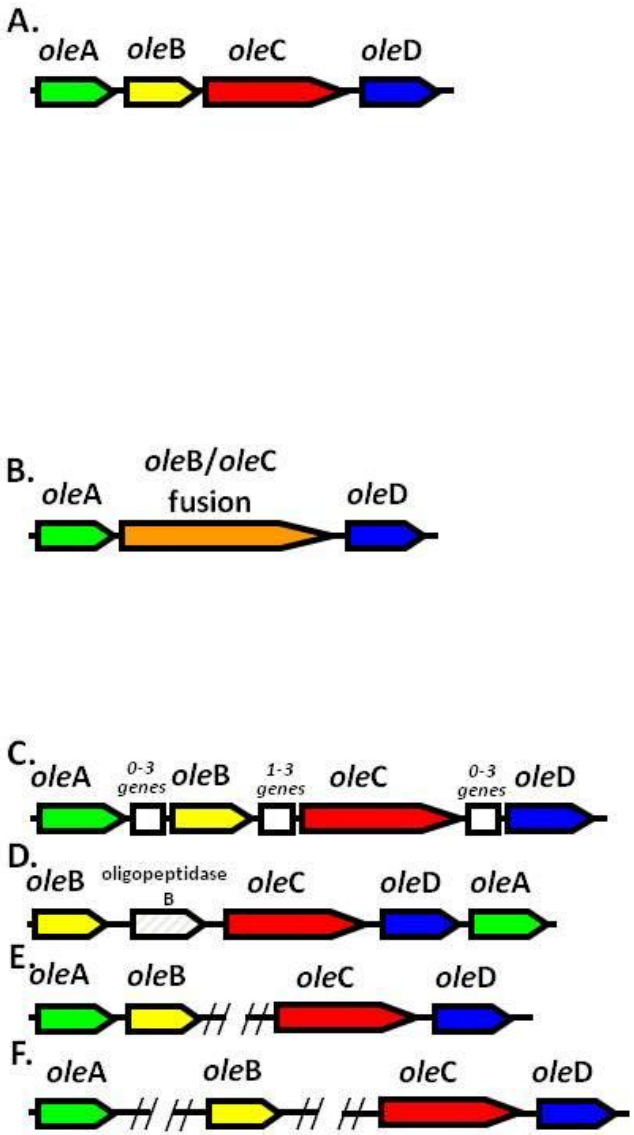


Figure 4.2. Analysis of the gene regions of *S. oneidensis* MR-1 (A), *G. bemidijiensis* Bem (B), and *G. sulfurreducens* PCA (C). Genes denoted *oleA* and *fabH* are homologs to the *oleA* from *S. oneidensis* MR-1. A predicted *oleA* gene region is shown for *S. oneidensis* (A) and *G. bemidijiensis* Bem (B), clustering with *oleBCD* genes. The *fabH* gene, which is an *oleA* homolog with highest percent identity with *G. sulfurreducens* PCA, fails to cluster with *oleBCD* homologs.

A collection of 3,558 genomes were examined using the described methods, leading to the identification of several different *ole* gene arrangements (Figure 4.3). One major distinction in *ole* gene organization had been recognized previously by Friedman *et al*. (WO2008/113041 and WO2008/147781); a significant number of organisms contained either three or four separate *ole* genes. Of those characterized in this study, the largest set contained four contiguous *oleABCD* genes. However, some bacteria of the class *Actinobacteria* contained three *ole* genes, with the *oleB* and *oleC* gene regions fused into one gene (Fig. 4.3A and B). In 61 organisms either the four- or three-gene

cluster was readily identifiable (Fig. 4.1A to D). Genomes that had a clear clustering of homologs of at least two of these genes were included as potential clusters. At least one sample organism from each of the gene clusterings in Fig. 4.3A to F was obtained, and the phenotype was confirmed experimentally by the presence of long-chain olefinic hydrocarbons in solvent extracts of growing cells (see below). As of 20 July 2009, highly likely *ole* genes were identified in 69 genomes. This was out of 3,558 total genomes. Thus, only 1.9% of the genomes examined contained evidence for *ole* genes based on the methods described here. Of the bacterial genomes, 69 out of 1,331, or 5.2%, showed bioinformatic evidence for *ole* genes. The genome analysis included 2,143 *Eukaryota* and 84 *Archaea*, none of which showed clear evidence of containing an *ole* gene cluster. This analysis does not rule out that the head-to-head hydrocarbon genes and pathway will be shown to be present in *Archaea* or *Eukaryota*, but only that our analysis could not identify them with confidence.



Colwellia psychrerythraea 34H
Desulfococcus oleovorans Hxd3
Desulfuromonas acetoxidans DSM 684
Gamma proteobacterium NOR5-3
Geobacter sp. FRC-32, sp. M21, G. bemidjiensis Bem, G. lovleyi SZ, and G. uraniireducens Rf4
Moritella sp. PE36
Photobacterium profundum SS9 and 3TCK
Plesiocystis pacifica SIR-1
Psychromonas ingrahamii 37
Psychromonas sp. CNPT3
Shewanella sp. ANA-3, sp. MR-4, sp. MR-7, sp. W3-18-1, S. amazonensis SB2B, S. baltica OS155, S. baltica OS223, S. baltica OS185, S. baltica OS195, S. denitrificans OS217, S. frigidimarina NCIMB 400, S. halifaxensis HAW-EB4, S. loihica PV-4, S. oneidensis MR-1, S. pealeana ATCC 700345, S. piezotolerans WP3, S. putrefaciens 200, S. putrefaciens CN-32, S. sediminis HAW-EB3, and S. woodyi ATCC 51908
Streptomyces ambofaciens DSM40697 and S. ghanaensis ATCC 14672

Arthrobacter aurescens TC1 and A. chlorophenolicus A6
Brachybacterium faecium DSM 4810
Brevibacterium linens BL2
Clavibacter michiganensis subsp. sepedonicus
Congregibacter litoralis KT71
Geodermatophilus obscurus DSM 43160
Kineococcus radiotolerans SRS30216
Kocuria rhizophila DC2201
Kytococcus sedentarius DSM 20547
Micrococcus luteus NCTC 2665
Nakamurella multipartita DSM 44233

Stenotrophomonas maltophilia R551-3 and *S. maltophilia* K279a
Xanthomonas axonopodis pv. citri str. 306
Xanthomonas campestris pv. campestris str. ATCC 33913, str. 8004, and str. B100
Xanthomonas oryzae pv. oryzicola BLS256, X. oryzae KACC10331, X. oryzae MAFF 311018 and X. oryzae PXO99A

Chloroflexus aggregans DSM 9485, C. aurantiacus J-10-fl, sp. Y-400-fl

Opitutus terrae PB90-1
Pelobacter propionicus DSM 2379
Shewanella benthica KT99

Planctomyces maris DSM 8797

Figure 4.3. *ole* gene regions of different bacteria. The gene region configuration is shown on the left, and the bacteria containing each are listed at the right. The double slashes in panels E and F indicate that the genes on either side are not contiguous. Green, *oleA*; yellow, *oleB*; red, *oleC*; blue, *oleD*; orange, *oleBC* fusion; white, other genes not currently identified as being involved in hydrocarbon biosynthesis. The different parts represent the most common contiguous four-gene configuration (A), the three-gene cluster in which the *oleB* and *oleC* genes are in a single gene, *oleBC* fusion (B), and gene organization with various insertions between the identified *ole* genes (C). The white boxes indicate multiple genes that may be encoded in the same or opposite directions to the *ole* genes. In particular, various *Xanthomonas* spp. strains have different numbers of genes identified in the indicated locations. (D) *Chloroflexi* species that have an oligopeptidase inserted between *oleB* and *oleC*. Also, the *oleA* homolog is located after the other genes. (E) Configuration in which pairs of genes are in different parts of the genome. (F) A configuration in which *oleA* and *oleB* are located in different parts of the genome but *oleC* and *oleD* are clustered. Note: hydrocarbon production was confirmed in at least one organism in each class, A to F. Identifiers for each of the genes are listed in Table E.1 in Appendix E.

4.4.3 Hydrocarbon identification.

It could not be inferred from sequence analysis alone whether all of the gene configurations would give rise to hydrocarbon products. In this context, at least one organism from each class (A to F) of Figure 4.3 was tested directly for long-chain olefin biosynthesis. In previous studies (2, 14, 51, 150, 157), olefins were produced under all growth conditions for all of the organisms tested; olefin production appears to be constitutive. In this context, each strain was grown under optimum conditions for the strain, as described in Table 4.1 and Materials and Methods. From each organism, nonpolar material was extracted with solvent and analyzed by chromatography and mass spectrometry. Controls were conducted with solvent blanks and organisms previously described not to produce head-to-head hydrocarbons (170) to exclude that olefins were derived from solvents or work-up procedures. This study showed that bacteria from the different types of gene clusterings shown in Fig 4.3 produced hydrocarbons in direct experimental tests (Fig 4.4 and Table 4.3). Different hydrocarbons were produced, but all were long chain ($>C_{23}$) and contained at least one double bond, consistent with their formation by a head-to-head coupling of fatty acyl groups.

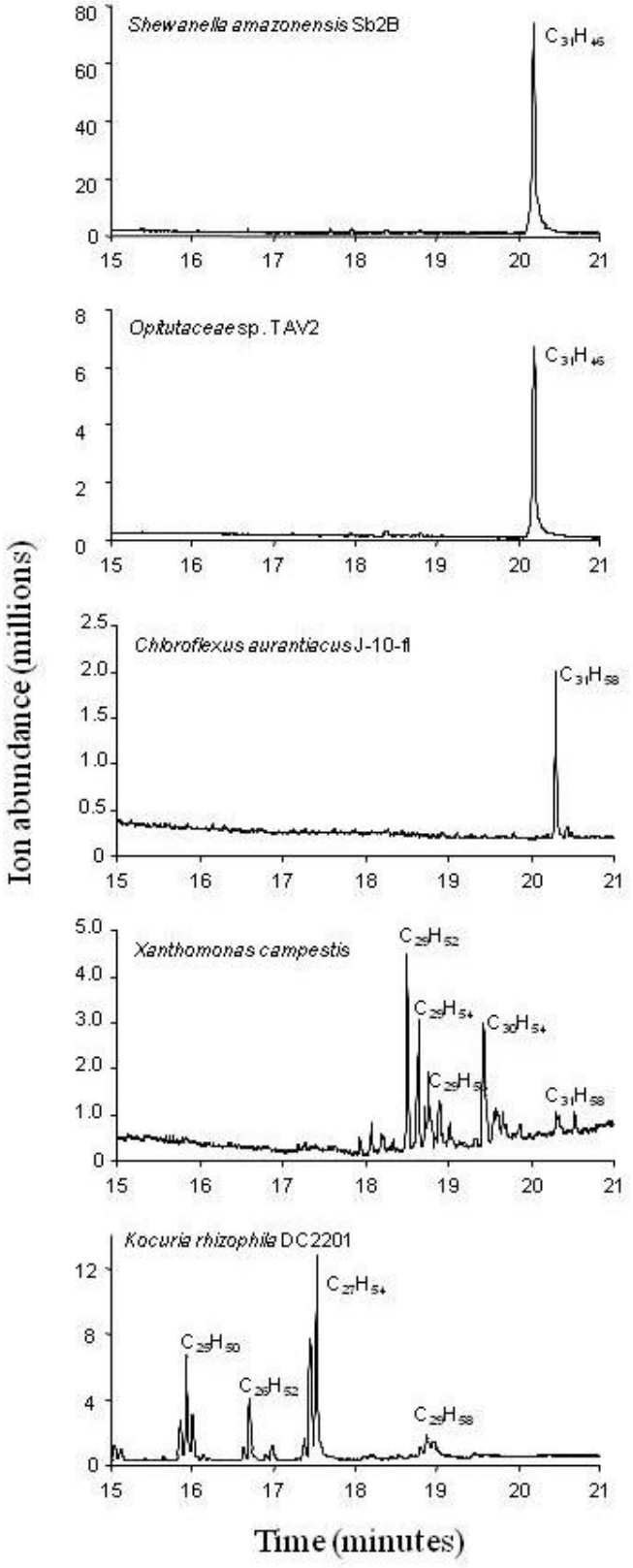

Shewanella amazonensis Sb2B
C31H46
Optutaceae sp. TAV2
C31H46
Chloroflexus aurantiacus J-10-fl
C31H58
Xanthomonas campestis
C29H52
C29H54
C30H54
C29H56
C31H58
Kocuria rhizophila DC2201
C27H54
C29H50
C28H52
C29H58
Ion abundance (millions)
Time (minutes)

Figure 4.4. Gas chromatograms of extracts from different bacteria containing *ole* genes as identified by bioinformatics. Bacteria were extracted, and extracts were analyzed by GC-MS as described in Materials and Methods. The major products are labeled with their chemical formulas. No hydrocarbon peaks were identified beyond the elution range shown.

Table 4.3. Compilation of head-to-head olefins produced by different bacteria

Microorganism	Total hydrocarbons		Predominant hydrocarbon	
	# hydrocarbons detected	Size range	Mass spectrum[1] (*m/z*)	Chemical formula
Chloroflexus aurantiacus J-10-fl	1	$C_{31}H_{58}$	430, 303	$C_{31}H_{58}$
Kocuria rhizophila DC2201	12	$C_{24}H_{48}$-$C_{29}H_{58}$	478, 348	$C_{27}H_{54}$
Brevibacterium fuscum ATCC 15993	9	$C_{27}H_{54}$-$C_{29}H_{58}$	406, 376	$C_{29}H_{58}$
Xanthomonas campestris pv. *campestris*	15[2]	$C_{28}H_{56}$-$C_{31}H_{58}$	402, 303	$C_{29}H_{54}$
Shewanella oneidensis MR-1	1	$C_{31}H_{46}$	418, 281	$C_{31}H_{46}$
Shewanella putrefaciens CN-32	1	$C_{31}H_{46}$	418, 281	$C_{31}H_{46}$
Shewanella baltica OS185	1	$C_{31}H_{46}$	418, 281	$C_{31}H_{46}$
Shewanella frigidimarina NCIMB 400	1	$C_{31}H_{46}$	418, 281	$C_{31}H_{46}$
Shewanella amazonensis SB2B	1	$C_{31}H_{46}$	418, 281	$C_{31}H_{46}$
Shewanella denitrificans OS217	1	$C_{31}H_{46}$	418, 281	$C_{31}H_{46}$
Colwellia psychrerythraea 34H	1	$C_{31}H_{46}$	418, 281	$C_{31}H_{46}$
Geobacter bemidjiensis Bem	1	$C_{31}H_{46}$	418, 281	$C_{31}H_{46}$
Opitutaceae bacterium TAV2	1	$C_{31}H_{46}$	418, 281	$C_{31}H_{46}$
Planctomyces maris DSM 8797	1	$C_{31}H_{46}$	418, 281	$C_{31}H_{46}$

[1] Identifying ions for the predominant long-chain olefin
[2] Number readily identifiable by gas chromatography-mass spectrometry

Shewanella amazonensis SB2B, isolated from the Amazon River Delta off the coast of Brazil (166), contains recognizable *ole* genes. It produced a single product with a carbon chain length of 31 and with nine double bonds ($C_{31}H_{46}$). The GC retention time and the mass spectrum indicated that the compound was identical to that produced by *S. oneidensis* strain MR-1, which had been described previously (150). The hydrocarbon in *S. oneidensis* MR-1 is the C_{31} polyolefin 3,6,9,12,15,19,22,25,28-hentriacontanonaene. Additional *Shewanella* strains were tested in this study, and all produced the C_{31} polyolefin as the only discernible hydrocarbon (Table 4.3).

Colwellia psychrerythraea is an obligate psychrophile that grows at temperatures below 0°C, *Geobacter bemidjiensis* Bem was isolated from a petroleum-contaminated aquifer sediment, *Opitutaceae* TAV2 is a member of the phylum *Verrucomicrobia* but is not well studied (136), and *Planctomyces maris* DSM8797 is in the phylum *Planctomycetes* and was isolated from the open ocean (12). Despite the great phylogenetic and ecological diversities of these bacteria, they all produced a single hydrocarbon product with the same retention time (20.2 min) and mass spectrum, consistent with its identity as 3,6,9,12,15,19,22,25,28-hentriacontanonaene (Figure 4.4 and Table 4.3).

A closely migrating, but clearly distinct, hydrocarbon product was produced by *Chloroflexus aurantiacus* strain J-10-fl (Fig. 4.4), a bacterium isolated from hot springs and that grows optimally at 55°C (120). The *Chloroflexus* hydrocarbon migrated more slowly on the GC column (20.4 min), and the mass spectrum indicated a chemical formula of $C_{31}H_{58}$, consistent with a hydrocarbon containing 31 carbon atoms and three

double bonds. These data are consistent with previous reports that identified hentriaconta-9,15,22-triene ($C_{31}H_{58}$) growing in microbial mats (164) and being formed by *Chloroflexus* spp. in pure culture (163).

Kocuria rhizophila strain DC2201 was isolated for its ability to withstand organic solvents (52), and its complete genome sequence was reported in 2008 (152). Here it was shown to produce multiple olefinic hydrocarbon products that ranged from 24 to 29 carbon atoms (Table 4.3). Each identified compound contained one double bond. The clusters of compounds eluting at approximately 16 min, 16.8 min, 17.5 min, and 19 min (Figure 4.4) represent isomeric clusters of C_{25}, C_{26}, C_{27}, and C_{29} chain lengths, respectively, based on mass spectrometry. This type of hydrocarbon cluster resembled, but was not identical to, those found in species of *Arthrobacter* (51) and *Micrococcus* (2, 14, 160) that have been studied previously. The major compounds in *Kocuria* analyzed here contained 25 and 27 carbon atoms. Another actinobacterial strain that had not yet been tested for the presence of head-to-head hydrocarbons, *Brevibacterium fuscum* ATCC 15993, similarly produced isomeric clusters of hydrocarbons but in the range of 27 to 29 carbon atoms (Table 4.3).

The most extensive array of hydrocarbon products from those organisms tested here was observed with *Xanthomonas campestris* (Figure 4.4 and Table 4.3), a bacterium that causes a range of plant diseases (22, 165). *X. campestris* produced hydrocarbons with chain lengths of C_{28}, C_{29}, C_{30}, and C_{31}. Based on the mass spectra, hydrocarbons containing one, two, or three double bonds could be identified. There was additional structural complexity that was likely due to isomerization, which could arise

from different types of methyl branching at the hydrocarbon termini. The complexity of the mixture precluded precise structural determinations, which would require the availability of synthetic standards.

Negative controls were run to rule out artifacts that could result, for example, from hydrocarbon contamination external to the cells (170). The most telling experimental results were obtained with *Geobacter sulfurreducens* PCA, an organism closely related to *G. bemidjiensis* Bem but suggested from bioinformatics analysis, in this study, to contain an *oleA* homolog with a different function (Figure 4.2). Olefinic hydrocarbons were not detected in *G. sulfurreducens.* Additionally, long-chain olefinic hydrocarbons were not detected in cultures of *E. coli* K-12 or *Vibrio furnissii* M1, both of which were determined not to contain *ole* genes based on the bioinformatics analysis described here.

Most previous studies had investigated bacterial head-to-head hydrocarbon biosynthesis in members of the *Actinobacteria*, including *Micrococcus* (2, 3, 14) and *Arthrobacter* (51). Long-chain olefinic hydrocarbons had also been demonstrated in *Stenotrophomonas maltophilia* (147), a member of the phylum *Proteobacteria*. The present study showed additional *Actinobacteria* (*Brevibacterium*) and *Proteobacteria* (*Geobacter* sp.) produce head-to-head hydrocarbons. In addition, members of the phyla *Verucomicrobia*, *Planctomyces*, and *Chloroflexi* were shown to contain bona fide *ole* genes and to produce olefinic hydrocarbons. This greatly expanded the phylogenetic diversity demonstrated experimentally to produce head-to-head olefinic hydrocarbons and revealed the type(s) of hydrocarbon produced. The latter could not be discerned

from the *ole* gene sequences alone based on previous studies. The present study is a start in establishing a link between one of the Ole protein sequences with the hydrocarbon(s) produced as discussed in the section below.

4.4.4 OleA has a major role in determining the type of head-to-head products formed.

The different long-chain olefinic hydrocarbons identified in this and other studies show variable chain lengths and degrees of unsaturation. These findings could be determined largely by the fatty acid composition within the cell, by the substrate specificity of the Ole proteins, by other proteins, or by some combination of these factors. To begin to investigate this, *S. oneidensis* MR-1 strains with different *oleA* gene contents were grown identically and tested for hydrocarbon content. The *S. oneidensis* MR-1 strains contained, respectively (A) the native *Shewanella oleA* gene only, (B) the native *Shewanella oleA* gene plus a *Stenotrophomonas oleA* gene, (C) no *oleA* gene, (D) the *Stenotrophomonas oleA* gene in a *Shewanella oleA* deletion strain, or (E) the *Stenotrophomonas oleA* gene in a *Shewanella oleABCD* deletion strain. Each strain (A to E) was grown under the same conditions of medium, temperature, and aeration. Each strain was harvested and extracted the same way. Each extract was subjected to the same chromatographic procedures.

The chromatograms shown in Figure 4.5 suggested that the product composition is strongly influenced by the *oleA* gene. In the same cell type, with cells grown under the same conditions and therefore likely having the same fatty acid precursor pools, the

product distribution was completely different when *oleA* genes from different organisms were present. When *oleA* genes native to *Shewanella* and *Stenotrophomonas* were expressed in the same cell, the products were additive to what was found with either alone (Figure 4.5B). Moreover, the *Stenotrophomonas oleA* gene, in the absence of the native *oleBCD* genes, was sufficient to make products of fatty acid head-to-head condensation (Figure 4.5E). This has implications for the mechanism of olefin biosynthesis and will be discussed in more detail below.

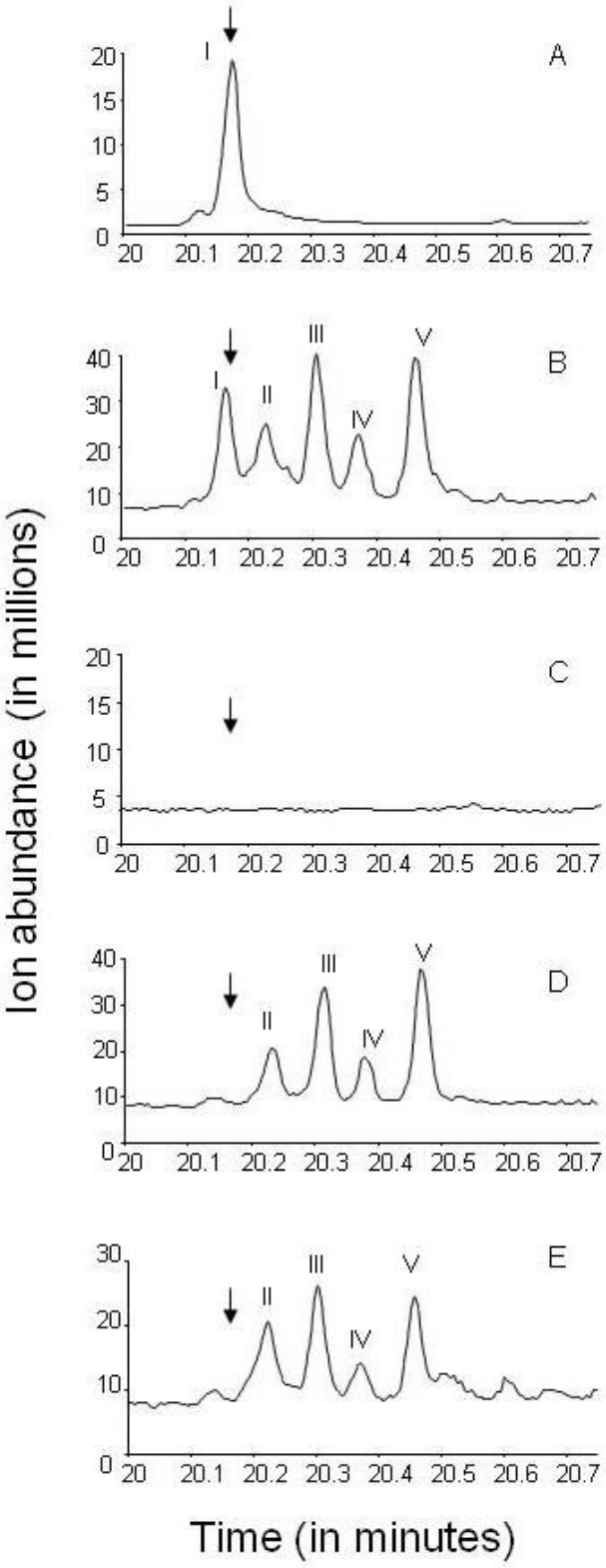

A
B
C
D
E
I
II
III
IV
V
Ion abundance (in millions)
Time (in minutes)
20
20.1
20.2
20.3
20.4
20.5
20.6
20.7

Figure 4.5. Gas chromatograms of extracts from wild-type and mutant *S. oneidensis* strains with and without the *oleA* gene from *S. maltophilia*. Extracts are from the following strains: (A) *S. oneidensis* MR-1 wild type; (B) *S. oneidensis* MR-1 wild type with *S. maltophilia oleA*; (C) *S. oneidensis* Δ*oleA*; (D) *S. oneidensis* Δ*oleA* with *S. maltophilia oleA*; (E) *S. oneidensis* Δ*oleABCD* with *S. maltophilia oleA*. Products were identified as described in the text.

When the *Shewanella oleA* gene was present, the cells made compound I (Figure 4.5A and B), which had been previously identified as a polyolefin containing nine double bonds derived from an intermediate in the polyunsaturated fatty acid biosynthetic pathway (150). The presence of the *Stenotrophomonas oleA* gene led to the formation of new products of fatty acid condensation. All of the later-eluting compounds, labeled II to V in Figure 4.5, were ketones. This was apparent from mass spectrometry based on (i) the parent ions, (ii) prominent fragment ions, and (iii) comparison to an authentic long-chain ketone standard. From known fragmentation of alkyl ketones and the observed fragmentation with standard 14-heptacosanone, the major fragments expected were R-CH_2-C=O. In the case of 14-heptacosanone, the carbonyl group is directly in the middle and fragmentation at either side on the carbonyl functionality yields a fragment of *m/z* 211, and this was observed experimentally using GC-MS. Compound II (Figure 4.5) showed fragments of *m/z* 223 and 225 and a parent ion of *m/z* 420, consistent with a compound containing a carbonyl functionality directly

in the middle of a C_{29} chain with 14 saturated carbon atoms on one side and a C_{14} chain with one double bond on the other. Compound III showed a fragment with *m/z* 223 and a parent ion of *m/z* 418. This mass spectrum is consistent with a compound containing a carbonyl functionality directly in the middle of a C_{29} chain flanked by two C_{14} chains, each containing one double bond. Compound IV showed a fragment with *m/z* 225 and a parent ion of *m/z* 422. This mass spectrum is consistent with a compound containing a carbonyl functionality directly in the middle of a C_{29} chain flanked by two saturated C_{14} chains. Compound V had a very similar mass spectrum as compound II. This suggested that it is a positional isomer of compound II and consists of hydrocarbon chains with one double bond and a saturated chain, respectively, linked together by a carbonyl functionality.

The data above are consistent with a fatty acid condensation between specific saturated and monounsaturated fatty acids. In separate experiments in which the *Shewanella oleABCD* deletion mutant was complemented with the *Shewanella oleA* gene, a compound with *m/z* 434 was obtained. This mass is consistent with a C_{31} compound containing one ketone functionality and eight carbon-carbon double bonds. The structure was confirmed by chemical modification. After hydrogenation, the compound had a parent ion of *m/z* 450 with a major fragment ion of *m/z* 239. This had the expected parent ion and major ion fragment for 16-hentriacontanone. Like the results shown in Figure 4.5A to E, this result was consistent with an *oleA* gene product causing specific condensation of two fatty acids. The *Shewanella* OleA showed selectivity for

polyunsaturated fatty acids, while the *Stenotrophomonas* OleA showed selectivity for saturated or mono- or di-unsaturated fatty acids.

A mechanism to explain the formation of ketones in the presence of *oleA* genes alone is proposed below. In total, these data highlight a potentially strong selectivity difference between OleA proteins from *Shewanella* and *Stenotrophomonas*. The observations here, showing that different *oleA* genes exert a strong influence on fatty acid condensation, have implications for the potential use of different *ole* genes to produce targeted hydrocarbon products commercially. Certain hydrocarbon products may be more desirable for industrial applications. In this context, a knowledge of OleA protein specificity would be critical in efforts to control product structure.

4.4.5 Olefin type in divergent bacteria tracks most closely with OleA sequence.

Very different types of olefin products were observed in wild-type bacteria, containing a range of from one to nine double bonds. Most bacteria in this study made exclusively the nonaene polyolefin previously identified in *Shewanella*. Data were presented in a previous study that indicated that the C_{31} nonaene compound was derived from polyunsaturated fatty acid precursors (150). However, polyunsaturated fatty acids account for 10% or less of the total fatty acids produced by *Shewanella* and other bacterial strains (1, 65, 71, 96). This strongly suggested that Ole enzymes must show selectivity in condensing certain fatty acids and not others. In light of the observations with *oleA* genes from *Shewanella* and *Stenotrophomonas* (Figure 4.5), the OleABCD

protein sequences were analyzed to see if, among the diverse bacteria analyzed here, Ole protein sequence relatedness correlated with the type of olefin produced by the cell.

Network clustering software was used to visualize the multidimensional relatedness of different sequences, as this method has been shown to be superior to trees for visualizing protein sequence relatedness (100, 106). The method makes an all-by-all BLASTp library of a sequence set. From these data, a network diagram is created in which the nodes represent protein sequences and the edges represent a BLAST linkage that connects the two proteins. A shorter edge represents a lower e-score (greater relatedness). For example, an e-value cutoff of e^{-73} was used in Fig[ure] 4.6. If the e-value of any pairwise comparison is lower (more related) than e^{-73}, then the sequences (Figure 4.6, circles/nodes) are connected by a line. Nodes that are not connected, or connected to fewer other nodes, are more divergent sequences. In this way, the network representation allows visualizations of connectivity more fully than protein tree analyses.

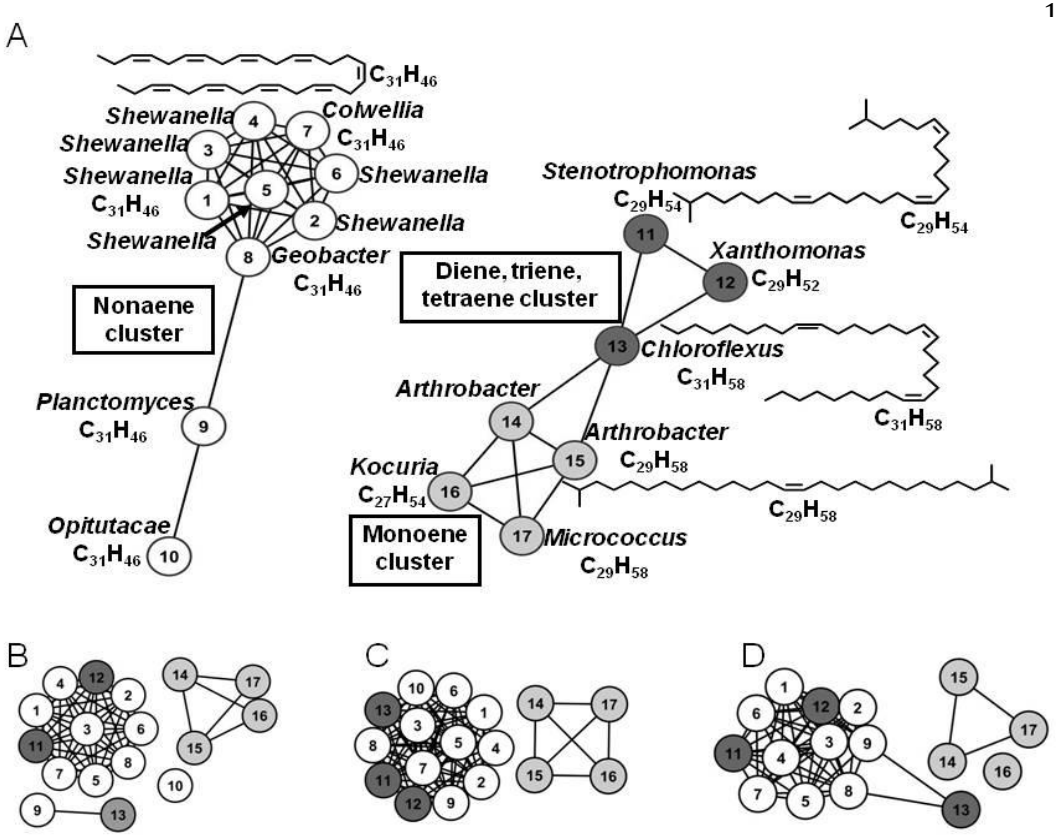


Figure 4.6. Network protein sequence clusters for OleA (A), OleB (B), OleC (C), and OleD (D). The nodes represent protein sequences, and the edges represent a BLAST linkage that connects the two proteins with an e-score better than e^{-73}. Other methods are described in the text. The nodes are numbered to identify the organism from which each Ole protein was derived. The organism names and number identifiers for each sequence are listed in Materials and Methods. The nodes are colored to reflect the type of hydrocarbon produced by that organism: white, a $C_{31}H_{46}$ nonaene product; dark gray, diene, triene, or tetraene product; light gray, monoene product. Additional network diagrams that depict divergence of the clusters can be found in Figure E.2 in Appendix E.

The network sequence analysis conducted for 17 each OleA, OleB, OleC, and OleD sequences is shown in Figure 4.6 (see Figure E.2 in Appendix E for results of more detailed clustering experiments). Those 17 were selected because all had been experimentally tested and shown to produce olefinic hydrocarbons, and the hydrocarbon products were identified. The top left side of Figure 4.6 readily shows that 10 of the OleA proteins cluster together (having all pairwise comparisons with e-values less than e^{-73}) and all produce the single polyolefinic hydrocarbon that is derived from polyunsaturated fatty acids. One explanation for this, which we favor based on the other data presented, is that the OleA proteins in *Shewanella*, *Geobacter*, *Planctomyces*, and *Opitutacae* specifically condense polyunsaturated fatty acids but do not condense the larger pool of more highly saturated fatty acids found in these classes of bacteria (31, 65, 71, 91).

Figure 4.6A also shows that the OleA proteins that make moderately saturated head-to-head olefins cluster differently than the OleA found in bacteria that produce the polyolefinic hydrocarbon. For example, *Chloroflexus aurantiacus* is known to make a C_{31} triene hydrocarbon (164), and that was confirmed in this study. A C_{31} triene would derive from the head-to-head condensation of two monounsaturated fatty acids. *Chloroflexus aurantiacus* makes predominantly C_{16} and C_{18} saturated fatty acids (163). The most obvious explanation is that the head-to-head biosynthetic pathway shows selectivity for only certain fatty acids within *Chloroflexus*.

Since the Ole-mediated head-to-head condensation process shows selectivity, it was investigated which Ole protein sequence networks clustered most strongly with the

type of head-to-head olefin formed. Figure 4.6A, B, C, and D represent the clustering networks of OleA, OleB, OleC, and OleD, respectively. For OleA (Figure 4.6A), sequence relatedness tracks with the type of olefinic hydrocarbon produced. For OleB, -C, and -D (Figure 4.6B, C, and D), the sequences cluster differently and are less reflective of the olefinic hydrocarbon structure. This is perhaps most apparent with OleB (Figure 4.6B).

With the cluster represented by the OleABCD sequences from the actinobacterial genera *Arthrobacter*, *Kocuria*, and *Micrococcus*, it was not possible to discern selectivity. The olefinic hydrocarbons produced are methyl branched, and the major fatty acids in *Arthrobacter* and *Micrococcus* are methyl branched (158, 162). The OleA proteins in the actinobacterial branch may be nonselective, or the proteins may have evolved selectivity that mirrors the major fatty acid types produced by the cell.

4.4.6 Potential mechanisms of OleA.

The observation that *Shewanella* OleA (Figure 4.5), *Stenotrophomonas* OleA (Figure 4.5), and other OleA proteins (Figure 4.6) confer fatty acid substrate selectivity is consistent with OleA catalyzing the first reaction in head-to-head hydrocarbon formation. An alternative proposal has been advanced in which several β-oxidation steps precede the OleA-catalyzed condensation reaction, and the reaction is coincident with the decarboxylation step (14). That mechanism was supported by two observations, the requirement for *E. coli* cell extract to support *in vitro* olefin synthesis and sequence alignments of the *Micrococcus luteus* OleA with *E. coli* FabH. The latter enzyme

catalyzes a decarboxylative fatty acyl (Claisen) condensation reaction. OleA proteins show the highest percent sequence identity with thiolase superfamily members, like FabH, that catalyze decarboxylative Claisen condensations.

The present study offers an alternative mechanism. As illustrated in Table 4.2, the thiolase superfamily contains several members that catalyze nondecarboxylative fatty acyl condensation reactions, for example, the biosynthetic thiolase involved in polyhydroxybutyrate (PHB) biosynthesis (36) and 3-hydroxyl-3-methylglutaryl-CoA synthase (HMG-CoA synthase) (145). The latter enzyme and other nondecarboxylative thiolase superfamily enzymes share the same highly conserved residues with those of OleA and FabH (Figure 4.1). The decarboxylative and nondecarboxylative thiolase superfamily proteins use these residues in an analogous manner to acylate a cysteine and then attack the bound acyl group with an enzyme-generated carbanion (62). The differences in mechanisms are subtle. Thus, sequence arguments cannot rule in or out decarboxylative versus nondecarboxylative mechanisms for OleA proteins.

Moreover, the mechanism proposed by Beller *et al*. for OleA is not analogous to that catalyzed by FabH. FabH acts on condensing a fatty acyl group containing an α-carboxy group, and this activation mechanism is not shown in the proposed mechanism (14). Those authors propose a series of steps catalyzed by unidentified enzymes to generate a β-ketoacyl chain that then reacts in condensation with release of coenzyme A and carbon dioxide.

An alternative mechanism would be for OleA to catalyze a nondecarboxylative Claisen condensation directly analogous to the reaction catalyzed by biosynthetic thiolases that function in PHB (36) and steroid (62) synthesis. Both biosynthetic and catabolic thiolases show free reversibility, and dozens of enzymes in the thiolase superfamily are already known to catalyze this general reaction. While the equilibrium constant for the biosynthetic direction is typically unfavorable, subsequent steps can affect the equilibrium, as occurs in PHB and steroid biosyntheses.

The product data are also suggestive that OleA catalyzes the first step in head-to-head hydrocarbon biosynthesis. The product selectivity shown in this study to arise from the *oleA* gene would be unusual if the OleA protein was in the middle of the biosynthetic pathway, as proposed by Beller *et al*. (14). Biosynthetic pathways are typically controlled at the first committed step in the pathway (61, 97). The mechanism proposed by Beller *et al*. requires additional enzymes to generate the 1,3-diketone that is proposed to undergo OleA-catalyzed condensation with a second fatty acyl chain. Those putative genes were searched for in the present study. The genes would need to be present in organisms producing head-to-head hydrocarbons and they might be expected to be contiguous, at least in some organisms, to the other genes encoding enzymes in the same metabolic pathway. However, we could not identify genes contiguous to the *oleABCD* gene clusters encoding enzymes that act to oxidize an acyl chain and generate a β-ketoacyl chain. This suggests that the OleABCD proteins may be sufficient for ketone and olefin biosynthesis.

Unlike the previous study (14), a nondecarboxylative, thiolytic type of fatty acyl condensation is proposed here. The nondecarboxylative type of mechanism would explain the observed formation of ketones with OleA *in vivo* and *in vitro* ((2, 14, 150); Friedman and Rude, WO2008/113041; Friedman and Da Costa, WO2008/147781; this study) and that the proposed 1,3-dione intermediate (14) has not been observed to date. Ketone formation following a direct OleA-catalyzed nondecarboxylative coupling of fatty acyl chains is chemically plausible and biochemically precedented (Fig. 4.7). This is reminiscent of the formation of acetone in humans via acetoacetyl-CoA (69). Acetoacetyl-CoA is a β-ketoacyl compound, as is the thiolytic product of the OleA reaction that we propose (Fig. 4.7). In human liver, excess acetoacetyl-CoA can give rise to acetoacetate, which is known to undergo spontaneous decarboxylation to acetone. The spontaneous decarboxylation of β-keto acids of this type has been known for more than 80 years and is quite facile (117). In a similar manner, we propose that the product(s) of the OleA reaction, if not acted upon by OleBCD, could undergo thioester hydrolysis either spontaneously (50) or enzymatically and then decarboxylation to generate a ketone(s). Note that the acyl-CoA compounds shown in Figure 4.7 are directly analogous. They can both arise from thiolytic condensation of either acetyl-CoA or longer-chain acyl-CoAs, respectively. The thioester could undergo enzyme-catalyzed hydrolysis. Alternatively, spontaneous thioester hydrolysis is known to be an important step in the mammalian blood clotting cascade (50). Thioester hydrolysis and facile β-ketoacid decarboxylation offer a plausible explanation as to why monoketones have been observed whenever the *oleA* gene, by itself, is cloned into a heterologous host ((14, 150); Friedman and Rude, WO2008/113041; Friedman and Da Costa,

WO2008/147781; this study). Moreover, ketones were observed in this study when exogenous *oleA* genes were placed into the *S. oneidensis* MR-1 background.

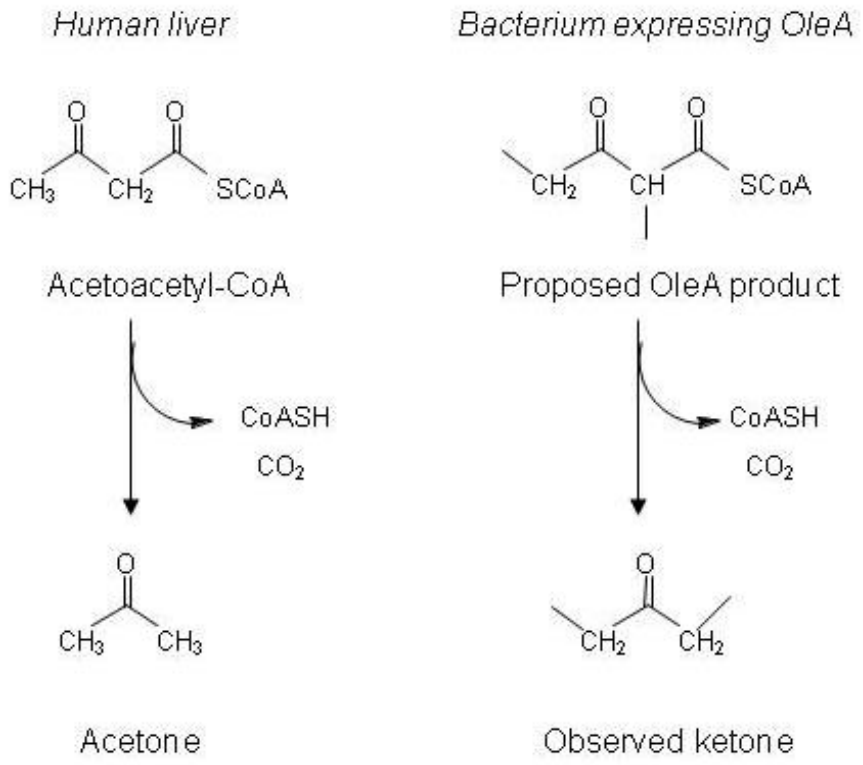


Figure 4.7. Parallel biological reaction sequences. At left is the known reaction sequence leading to ketones in humans. At right is the proposed reaction sequence leading to ketones in bacteria expressing OleA

At this time, there are two alternative mechanisms proposed for the initiation of head-to-head hydrocarbon biosynthesis and the specific role of OleA. Further studies will be required to discern between the mechanism proposed previously (14) and the nondecarboxylative Claisen condensation favored here.

4.5 Addendum

4.5.1 *Clavibacter michiganensis*

Clavibacter michiganensis subspecies also have the *ole* gene cluster with *oleA*, *oleD*, and a fused *oleBC*; three genes overall. *C. michiganensis* species are plant pathogens known to infect many plant types (103). Numerous strains of this organism were obtained (strains tested are listed in Table 4.4) , grown for 2 days on NBY Broth (8.0g Nutrient Broth, 2.0g Yeast Extract, 2.5g Glucose, 2.0g K_2HPO_4, 0.5g KH_2PO_4, and 1ml 1M $MgSO_4$ x $7H_2O$) and extracted. Resulting extracts were placed on the GC-MS for analysis.

Table 4.4. *C. michiganensis* strains tested

ID	Species
CIC14	*Clavibacter michiganensis* subsp. *michiganensis*
CIC17	*Clavibacter michiganensis* subsp. *nebraskensis*
CIC21	*Clavibacter michiganensis* subsp. *tessellarius*
CIC23	*Clavibacter michiganensis* subsp. *michiganensis*
CIC250	*Clavibacter michiganensis* subsp. *sepedonicus*
CMI R1-1	*Clavibacter michiganensis* subsp. *insidiosis*

All strains of cultured *C. michiganens*is were found to produce one identical olefinic hydrocarbon product that contained 29 carbon atoms that eluted at 18.83 minutes from the GC. Figure 4.8 is a representative GC-MS chromatogram of the *C. michiganensis* subsp. *sepedonicus* extract (ID ClC250). This compound contains one double bond and has a parent ion of m/z 406. Mass spectra of the compound was found to be identical to a synthesized *cis*-3,25-dimethyl-13-heptacosaene standard. A co-

injection of the *cis*-3,25-dimethyl-13-heptacosaene standard was performed with the *C. michiganensis* subsp. *sepedonicus* extract. The standard co-eluted with the hydrocarbon peak. No other hydrocarbons were identified, including no *cis*-2,26-dimethyl-13-heptacosaene or mixtures of *cis*-2,25-dimethyl-13-heptacosaene and *cis*-3,26-dimethyl-13-heptacosaene.

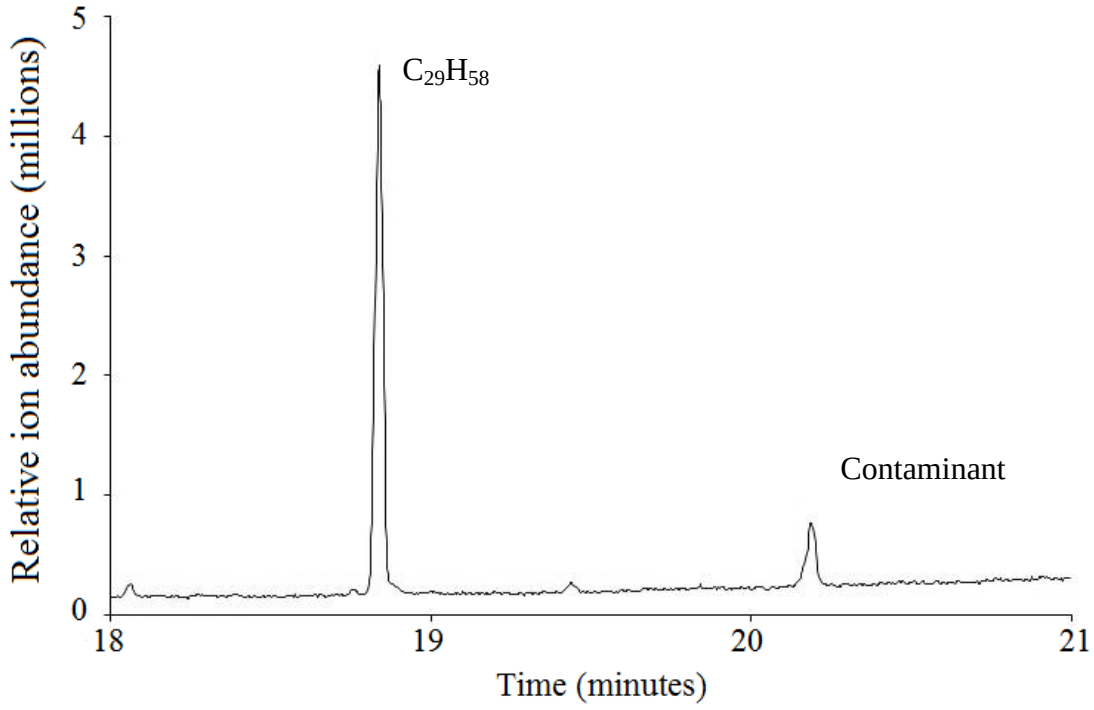


Figure 4.8. Representative gas chromatogram of extract from *C. michiganensis* subsp. *sepedonicus*. Product was identified as described in the text.

Chapter 5: OleA: fatty acyl CoA specificity versus availability

5.1 Overview

Previous studies identified the involvement of OleA in the condensation of two fatty acyl CoAs in hydrocarbon biosynthesis. The present study more fully defines the selectivity of OleA in its selection for fatty acyl CoA condensation. Strains of *S. oneidensis* and *R. eutropha* were constructed to heterologously express the *C. aurantiacus* OleA protein. The *C. aurantiacus* OleA appears to have a broader substrate specificity than was suggested by the in vivo data with *C. aurantiacus* that showed a single major product, 9, 15, 22-hentriacontatriene. When expressed heterologously in *E. coli* or *R. eutropha*, a wider range of products were observed. This suggests that in *vivo* olefin formation by *Chloroflexus* OleA may be determined by the types of fatty acyl CoAs available in each specific bacterial strain.

5.2 Introduction

Chloroflexus aurantiacus is a thermophilic bacteria isolated from a Yellowstone hot springs. It is known to produce 9, 15, 22-hentriacontatriene (120, 149). Bioinformatics work showed that the OleA protein was most closely similar to the *Stenotrophomonas* and *Xanthomonas* OleA proteins (149), two organisms shown to produce numerous hydrocarbon compounds ranging in sizes from 24-31 carbons in length (149, 161).

Other organisms were found to naturally produce one C31-hydrocarbon, though the hydrocarbon product by these organisms was 3, 6, 9, 12, 15, 19, 22, 25, 28-

hentriacontanonaene (110, 149, 150). When the *S. amazonensis, S. oneidensis,* or *C. psychrerythraea oleA* genes were cloned into a *S. oneidensis* MR-1 *oleA* gene deletion strain, the resulting constructs only produced the C31-nonaene (Chapter 3.6, this thesis), implying that the OleA proteins of the hentriacontanonaene organisms had high substrate specificity; specificity was dependent upon the OleA protein, not substrate availability. In contrast, when the *S. maltophilia oleA* gene was cloned into the *S. oneidensis oleA* gene deletion mutant, the resulting construct produced a similar range of hydrocarbon products as did *S. maltophilia* (150), implying the *S. maltophilia* OleA protein had less substrate specificity than the C31-nonaene-specific OleA proteins.

In this chapter we tested the *C. aurantiacus* OleA *in vivo* in different bacterial backgrounds to identify whether it was more like a broad specificity OleA.

5.3 Methods

Chloroflexus aurantiacus was routinely grown in *Chloroflexus* media at 55^{o}C (120). *Shewanella* strains were grown in Luria broth at 30^{o}C. *Ralstonia eutropha* was grown in Luria broth at room temperature (72). Primers used in this study are listed in Table 5.1. The *Chloroflexus aurantiacus oleA* gene was amplified using primers ClthiolCompF and ClthiolCompR containing the SpeI and SacI restriction sites. PCR products were ligated into the Strataclone cloning system (Agilent Technologies) followed by ligation of the product into the pBBR1MCS2 expression vector. The *oleA* gene was also amplified using primers RethiolF and RethiolR containing the ClaI and XhoI restriction sites. Resulting PCR products were ligated into the Strataclone system followed by excision and ligation of the insert into the pBBR vector provided by the

Srienc Laboratory (University of Minnesota). Vector constructs were introduced into *E. coli* WM3064 and conjugated into the *oleABCD* gene deletion of *S. oneidensis* (pBBR1MCS2 vector) or *R. eutropha* (pBBR vector). Appropriately oriented inserts were verified by PCR analysis.

For hydrocarbon analysis, cultures were extracted using the Bligh and Dyer technique as previously described (15) prior to GC-MS analysis. Hexadecane spikes were routinely added to extracts for hydrocarbon quantification. Bacterial constructs were routinely grown at 30°C unless stated otherwise.

Table 5.1 Strains, vectors, and primers used in Chapter 5

Strains	Notes	Ref
Chloroflexus aurantiacus sp. J-10-fl	wildtype strain	120
Shewanella oneidensis ΔoleABCD	*oleABCD* gene deletion	150
Ralstonia eutropha	wildtype strain	72
Plamsids		
pBBR1MCS2	5.1KB broad-host range plasmid, *lacZ*, Km^r	89
pChloro	pBBR1MCS2 containing 1.1KB fragment of *C. aurantiacus oleA*	Thesis Chapter 3
pBBR	6.0KB *Ralstonia* plasmid, Km^r	Srienc Lab
pBBRChloro	pBBR containing 1.1KB fragment of *C. aurantiacus oleA*	This study
Primers		
ChlorooleAClaF	CATATTATCGATATGCTATTCAGGCATGTCATGATCG	
ChlorooleAXhoR	CAATATCTCGAGTCACCACGTCACACTCATCATTGAAC	

5.4 Results and Discussion

When *Chloroflexus aurantiacus* was grown under optimal conditions for three weeks, the organism produced one hydrocarbon (Figure 5.1). This compound contained a parent ion of m/z 430, consistent with the hydrocarbon 9, 15, 25-hentriacontatriene.

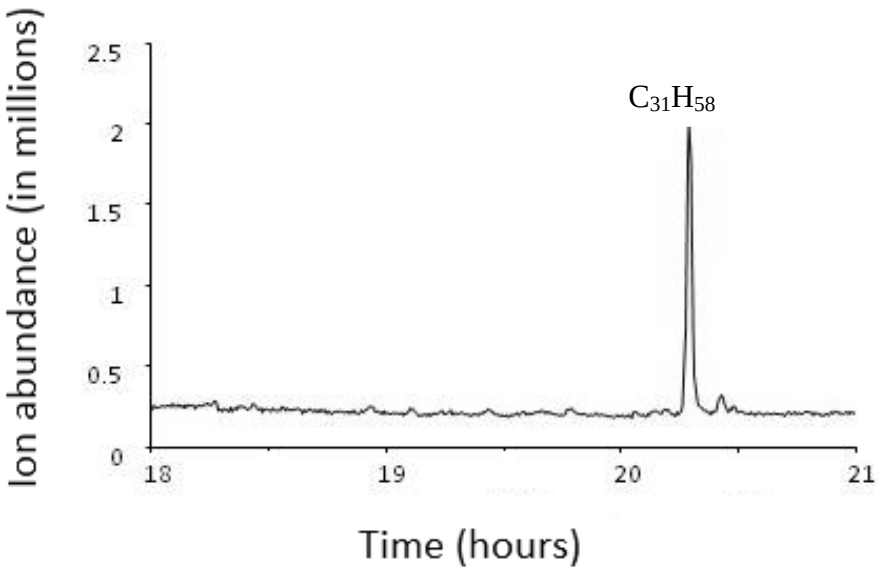


Figure 5.1. GCMS chromatogram of a solvent extract from a wildtype *C. aurantiacus* culture. Compounds were identified as hydrocarbons or ketones by mass spectrometry as described in the text.

Previous studies found that OleA proteins condensed two fatty acyl CoAs to produce a fatty acyl compound. The fatty acyl compound is spontaneously decarboxylated to produce a ketone when the OleBCD proteins are not present (150). When the *C. aurantiacus oleA* gene was introduced into the *S. oneidensis oleABCD* deletion strain, the organism not only produced the expected C31-ketodiene, but also numerous other ketones ranging from 28 to 31 carbons in length (Figure 5.2). The predominant ketone produced contained a fragment with *m/z* 223 and a parent ion of *m/z* 418. This mass spectrum is consistent with a compound containing a carbonyl functionality directly in the middle of a C29 chain flanked by two C14 chains, each containing one double bond. Another predominant compound showed fragments of *m/z*

223 and 225 and a parent ion of *m/z* 420, consistent with a compound containing a carbonyl functionality directly in the middle of a C29 chain with 14 saturated carbon atoms on one side and a C14 chain with one double bond on the other. Positional isomers were also noted in the gas chromatogram. A similar profile of compounds were seen for ketones of 28, 30, and 31 carbons in length, but in lesser quantities.

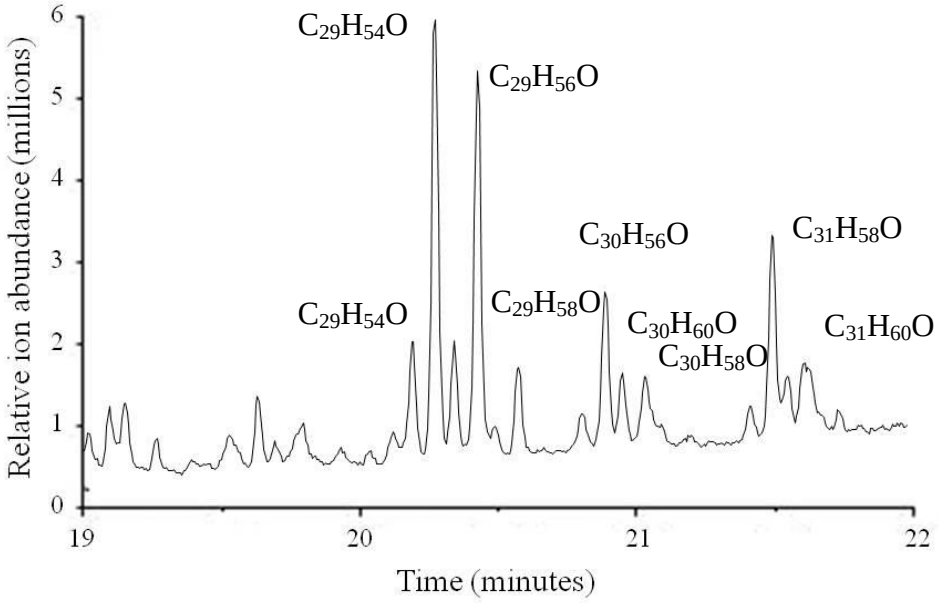


Figure 5.2. GC results for a solvent extract from a recombinant *S. oneidensis oleABCD* deletion strain expressing the heterologous *C. aurantiacus* OleA protein. Compounds were identified as ketones by mass spectrometry as described in the text.

Various *oleA* genes were introduced into *Ralstonia eutropha* (see Appendix F for a list of *oleA* genes introduced, as well as primers used for cloning the genes into the pBBR vector in Table F.1) and though all *oleA* genes were transcribed as found by RT-

PCR (see Appendix F, Figure F.1), only constructs containing the *C. aurantiacus oleA* gene were able to produce identifiable products (Figure 5.3). The *R. eutropha* strain expressing the *C. aurantiacus oleA* gene produced predominantly a product with a parent ion of m/z 390 with a strong ion peak at m/z 209 as determined by gas chromatography-mass spectroscopy. This mass spectrum is consistent with a compound containing a carbonyl functionality directly in the middle of a C27 chain flanked by two C13 chains, each containing one double bond. The constructs also had a strong peak with a parent ion of m/z 392 with secondary ions of m/z 209 and 211, consistent with a compound containing a carbonyl functionality directly in the middle of a C27 chain with 13 saturated carbon atoms on one side and a C13 chain with one double bond on the other. A third peak identified as a saturated ketone of 27 carbons was also detected (m/z 394, 211). Other minor ketones were identified, all of which were 29 carbons in length. These included two isoforms of a ketone with two unsaturation locations (m/z 418, 223).

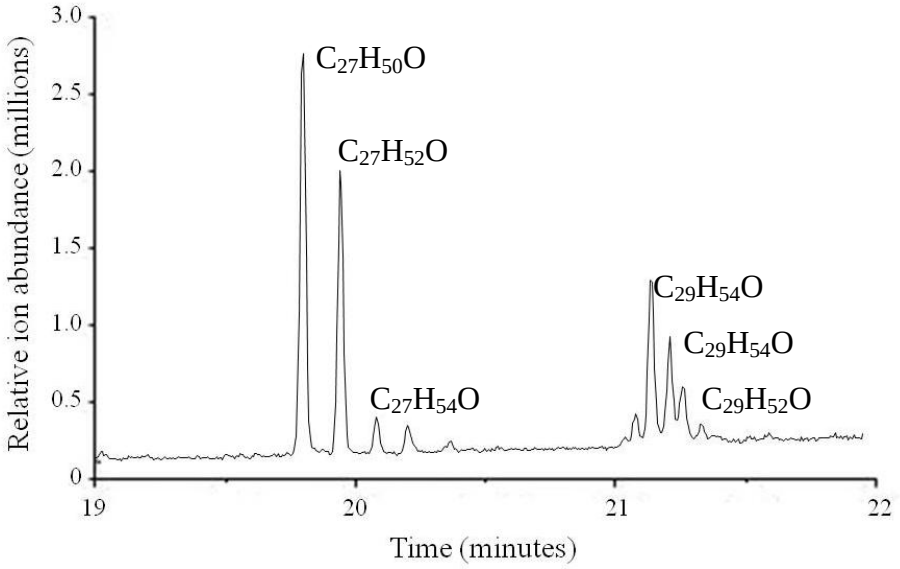


Figure 5.3. GC results for a solvent extract from a recombinant *R. eutropha* expressing the *C. aurantiacus* OleA protein. Compounds were identified as ketones by mass spectrometry as described in the text.

These findings were somewhat surprising. Though fatty acid compositions were not determined during the experiments described above, when grown under similar conditions Kenyon *et al.* (1974) and van der Meer *et al.* (1999) report that *Chloroflexus aurantiacus* produces predominantly 18:1 monenoic acid and 16:0 and 18:0 saturated acids (81, 163). The 16:1Δ9 fatty acid is produced, but to only approximately 2% of the total fatty acid content of the organism (81). We know that OleA condenses fatty acyl CoAs, not free fatty acids (J. Frias,University of Minnesota, personal communication). Given that the majority of the fatty acids produced by the organism are C16 and C18 in length, we expected to have large CoA pools of these compounds. With the understanding that *Chloroflexus aurantiacus* naturally produces a hydrocarbon derived from two 16:1Δ9 fatty acyl CoA chains, we assumed that the *C. aurantiacus* OleA

protein was highly selective in the fatty acyl CoAs it condensed. The results found when the *C. aurantiacus oleA* gene was expressed in the other organisms contradicts this hypothesis. *S. oneidensis* produces predominantly C15 fatty acids (167) while *R. eutropha* had been previously shown to produce predominantly C14 saturated and monounsaturated fatty acids (54). As observed in our *Shewanella* and *Ralstonia* constructs, if the related fatty acyl CoA chains were condensed, the resulting ketone would be 29 or 27 carbons in length respectively and contain parent ions of m/z 420-424 and 392-396. The reasons why *C. aurantiacus* does not produce a wider range of hydrocarbons naturally is unclear. The *C. aurantiacus* may condense palmityl CoA followed by desaturation of the resulting hydrocarbon product. There is also the possibility that in *C. aurantiacus* there is a larger pool of palmitoleoyl CoA than expected. References pertaining to fatty acyl CoA availability in *C. aurantiacus* were not found.

While the *C. aurantiacus* OleA protein condenses two palmitoleoyl CoA chains in its original environment, it is a more promiscuous protein than originally thought. Further study will be necessary to determine how much condensation is dependent upon the OleA protein specificity and how much is dependent upon fatty acyl CoA availability.

Chapter 6: Increases in ketone production using pBBAD22K vector

6.1 Overview

A previous study found that the OleA protein from *Stenotrophomonas maltophilia* can be heterologously expressed in *Shewanella oneidensis,* resulting in the production of ketones ranging from 27 to 31 carbons in length. The present study documents the attempt to increase ketone production using the *S. maltophilia oleA* gene. By changing the vector backbone and promoter driving the expression of the *oleA* gene from a constitutively active Lac-promoter to an L-arabinose-inducible promoter, a 5-fold increase in ketone production was observed. The resulting bacterial strains grew slower and to a lower optical density. Based on these observations, it was proposed that next steps toward producing more ketones in the *S. oneidensis oleABCD* deletion strain should include the attempt to increase fatty acyl CoA precursors through genomic manipulation.

6.2 Introduction

The OleA protein condenses two fatty acyl CoA chains as the initiation step in hydrocarbon production in heterotrophic bacteria (150, personal communication with J. Frias, University of Minnesota). Pools of fatty acyl CoA in bacteria are finite. Ways to increase hydrocarbon production include deletion of the genes encoding part of the fatty acid oxidation pathway to prevent fatty acyl CoA breakdown (53), increase the proteins used to produce the precursors for hydrocarbon formation (53),and increase the proteins directly involved in the hydrocarbon biosynthesis pathway. There are positives and

negative consequences of one manipulating the central metabolism in bacteria. When one overexpresses a protein, the bacterium has the ability to increase production of a given substance. The resulting constructs are often sickly, resulting from the siphoning of energy and raw materials away from other important functions (141). Also, overproduced proteins may form inclusion bodies—or miss-folded protein masses—in the cell (13).

When thinking about overexpressing a given protein, there are a number of factors that need to be considered. For example, various promoters are stronger than others in a bacterial platform. Induction of a strong promoter often increases protein accumulation (141). Like promoters, not all vectors are the same. The number of vectors per cell differ depending on the vector's origin of replication, allowing the cell to have greater potential for gene expression (141).

In this study, we attempted to increase hydrocarbon production by manipulating the quantities of OleA protein produced through use of pBBAD22K, a broad host range vector tightly controlled through L-arabinose induction (148).

6.3 Methods

All bacterial strains, plasmids, and primers used in this study are listed in Table 6.1. The *S. maltophilia oleA* gene was also amplified using primers pBSmoleAF and pBSmoleR containing the EcoRI and SacI restriction sites. Resulting PCR products were ligated into the Strataclone system followed by ligation of the PCR product into the pBBAD22K vector (Figure 6.1) (148). Vector constructs were introduced into *E. coli* WM3064 and conjugated into the *oleABCD* deletion strain of *S. oneidensis*.

Bacteria were commonly grown aerobically at 30^{o}C for two days in Luria broth prior to ketone extraction. When needed, cultures were supplemented with Kanamycin (50 µg/ml).

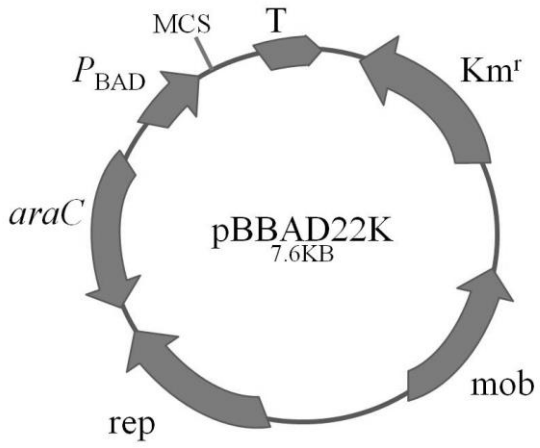


Figure 6.1 Vector map of pBBAD22K. Plasmid contains pBAD promoter with a multiple cloning site downstream, a kanamycin resistance cassette, origin of replication, and an origin of transfer.

Ketones were extracted using the Bligh and Dyer technique (51). Briefly, 7.5 ml of chloroform and methanol were added to the 3ml of culture followed by 3 ml of water and vortexed vigorously. The chloroform phase was separated and allowed to evaporate. Residue was solubilized in MTBE prior to analysis on the GC-MS. An internal standard of hexadecane was added to each extraction to allow for quantification and intra-chromatograph comparison.

Table 6.1 Strains and plasmids used in Chapter 6

Strain or plasmid Strains	Genotype or relative characteristic(s)	Ref
Strains		
Shewanella oneidensis ΔoleABCD	*S. oneidensis* MR-1, *Δole*; does not produce hydrocarbon	150
Escherichia coli UQ950	*E. coli* DH5α λ(pir) host for cloning; F-Δ(*argF-lac*)*169 Φ80dlacZ58*(ΔM15) *glnV44*(AS) *rfbD1 gyrA96*(NalR) *recA1 endA1 spoT1 thi-1 hsdR17 deoR λpir+*	133
Escherichia coli WM3064	Donor strain for conjugation: *thrB1*004 *pro thi rpsL hsdS lacZ*ΔM15 RP4-1360 Δ(*araBAD*)*567 ΔdapA*1341::[*erm pir*(wt)]	133
Plasmid		
pBBR1MCS2	5.1-kb broad-host range plasmid; *lacZ*; Km^r	89
pOleA-*S.m.*	1.1-kb PCR fragment containing the *S. maltophilia oleA*, cloned into the SpeI/SacI site of pBBR1MCS-2	150
pBBAD22K	7.6-kb broad-host range plasmid	148
pBBAD-*S.m.*	1.1-kb frament containing the *S. maltophilia oleA* cloned into the EcoRI/SacI site of pBBAD	This study
Primers		
pBSMoleAF	CAATCGAATTCGCCTTGGCGCTAACCTCGATGCTCTTC	
pBSMoleAR	GATTAGAGCTCCACCTACCCATTTCCTCTTCCGGTGG	

6.4 Results and Discussion

In *Escherichia coli*, the pBBR1MCS2 is a low copy number vector that is constitutively expressed (43). The pBBAD22K is also a low copy number vector, but is under the tight control of the arabinose promoter (148). In an attempt to increase ketone production in *Shewanella oneidensis ΔoleABCD*, the *oleA* gene from *Stenotrophomonas maltophilia* was introduced into the pBBAD22K vector. When this construct was induced with 0.1% L-arabinose for two days, it was found that there was an increase in ketone production as compared to the pBBR1MCS2 vector construct (Figures 6.1 and 6.2). Greater quantities of ketone were produced across the profile of compounds.

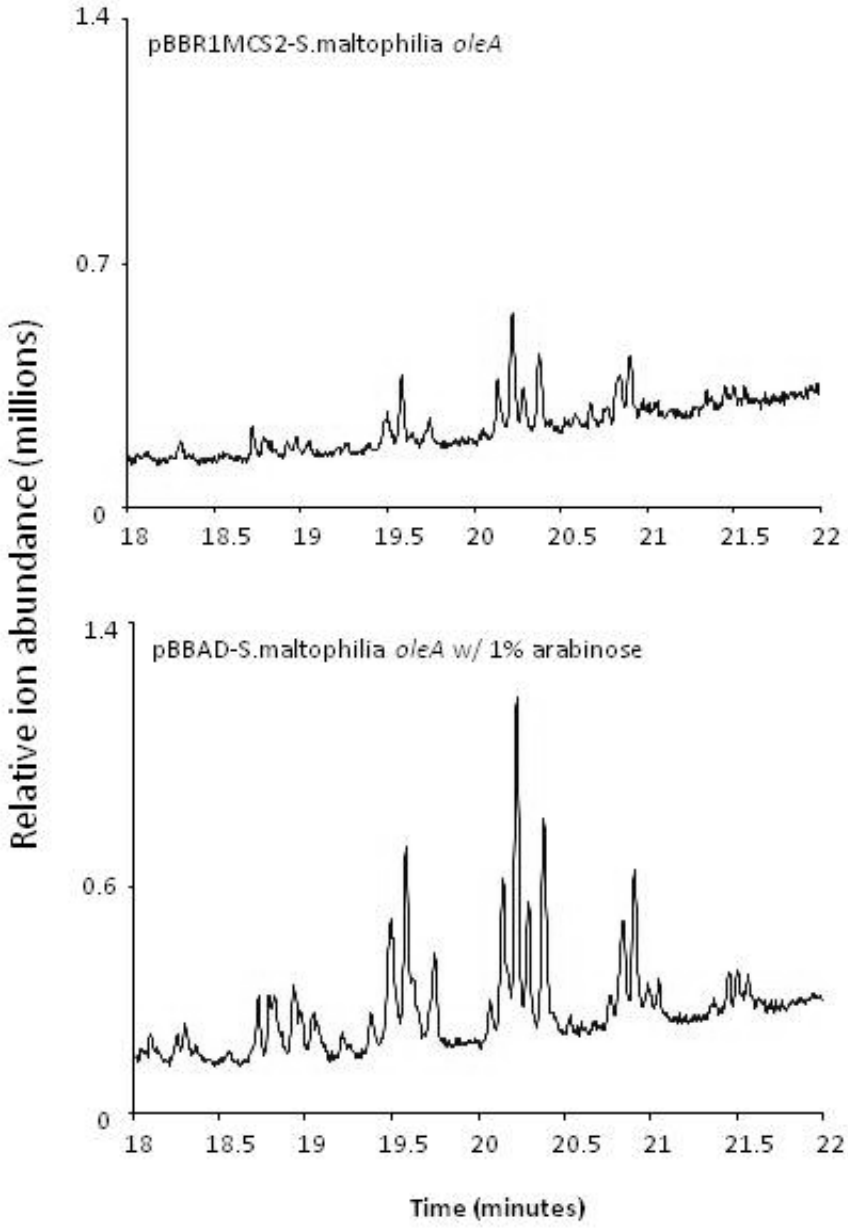


Figure 6.2. GC results for a solvent extract from recombinant *S. oneidensis ΔoleABCD* expressing the heterologous *S. maltophilia oleA* gene on (a) the pBBR1MCS2 plasmid and (b) the pBBAD22K plasmid. Compounds were identified as ketones by mass spectrometry as described in the text. An internal standard of hexadecane was used to allow GC comparisons.

Cultures of *S. oneidensis ΔoleABCD* containing the pBBAD-S.m. vector were subjected to various degrees of L-arabinose saturation to identify the greatest amount of ketone that could be produced using this construct. L-arabinose ranging from 0.01% to 1.0% (weight to volume) was added to cultures and allowed to incubate for two days. As depicted in Figure 6.2, the 1.0% arabinose-induced cultures (n = 9) produced approximately 5x more ketone than the pOleA-S.m. control cultures (n = 9), while the 0.1% induced cultures produced approximately 4x more ketone (n = 9). In contrast, the 0.01% induced cultures (n = 8) produced very little ketone as compared to the pOleA-S.m. control cultures. After three days of induction, optical densities of cultures of *S. oneidensis ΔoleABCD* containing the pBBAD-S.m. vector induced with 1.0% arabinose were lower than the pOleA-S.m. control cultures. One explanation for this difference in culture phenotypes is that the OleA protein is siphoning away too much fatty acyl CoA, resulting in the disruption of cellular function leading to cell death.

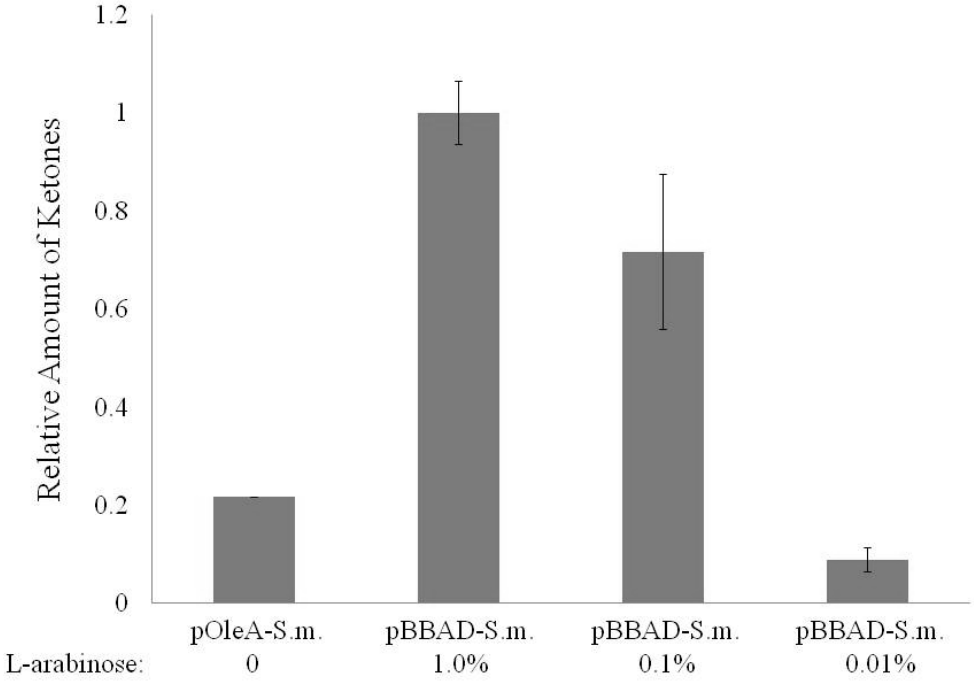


Figure 6.3 Relative amounts of ketone produced by *S. oneidensis ΔoleABCD* containing the pBBR1MCS2 and pBBAD22K vectors with the *S. maltophilia oleA* gene. Cultures containing the pBBAD-S.m. vector were induced with 1.0%, 0.1%, and 0.01% L-arabinose for 2 days. Cultures were extracted as described in the text. All extractions were normalized to culture OD and an internal standard of hexadecane. Variation is shown asstandard deviations between the replicates.

We were able to dramatically increase ketone production under the conditions described above. While it is interesting to note the ketone production was greater than the previously constructed organisms, the total quantity of ketone produced was still low. It appears that we have achieved the maximum amount of ketone production using the described *S. oneidensis ΔoleABCD* containing the pBBAD-S.m. vector. Not only

are more ketones not being produced with an increase in L-arabinose, but the health of the cultures are compromised: constructs do not reach the same optical density and decrease sooner than the wildtype *S. oneidensis* or the pBBR1MCS2 strains containing the *S. maltophilia oleA* gene. I hypothesize that, under this higher producing scenario, too many fatty acyl CoAs are being shunted away from normal metabolic pathways by the OleA protein for the bacteria to function and remain viable. Therefore I propose that the next step in creating an organism to produce more ketone is to increase fatty acyl CoA precursors through gene knockouts and overexpression.

Chapter 7: Thesis conclusions and future work

7.1 Conclusions

The proceeding chapters focused on three aspects of bacterial hydrocarbon production: (1) how certain heterotrophic bacteria produce hydrocarbons, (2) whether we can manipulate bacteria to produce hydrocarbons, and (3) whether we can detect when bacteria produce hydrocarbons. A previous report by Park *et al.* stating that *Vibrio* species produce hydrocarbons was not reproduced in our laboratory. Heterotrophic bacterial hydrocarbon biosynthesis was traced to the *oleABCD* gene cluster. It was found that many bacteria contained this gene cluster and had the ability to produce hydrocarbons. These bacteria include members of the *Proteobacteria* Phylum as well as bacteria from the *Verrucomicrobia*, *Planctomycetes*, *Actinobacteria*, and *Chloroflexi* Phyla. Prior to these studies, bacteria from the *Verrucomicrobia* and *Planctomycetes* Phyla had not been reported to produce hydrocarbons. *Shewanella oneidensis* MR-1 was found to produce one hydrocarbon product, identified as 3,6,9,12,15,19,22,25,28-hentriacontanonaene. The hydrocarbon production was linked to temperature changes in the culture. We identified that ketones could be produced by bacteria through use of the OleA protein. When *oleA* genes were heterologously expressed in *Shewanella oneidensis, Ralstonia eutropha,* and *Escherichia coli*, ketones of diverse lengths and saturation were produced. The ketone characteristics were dependent upon two factors: (1) OleA protein specificity for different fatty acyl CoAs and (2) bacterial fatty acyl CoA pool availability. Expression of the OleA protein in cyanobacteria was attempted. Finally, exploratory experiments were performed in an attempt to find a compound that could be used to detect or select for ketone production.

7.2 Future Work

7.2.1 OleA protein Screening

There are over seventy sequenced genomes containing a homologous gene to the *Shewanella oneidensis* MR-1 *oleA*. In the proceeding chapters I only mentioned testing fourteen for activity in *Shewanella oneidensis*. To date, the majority of the ketone-producing bacterial constructs utilize the OleA protein of *S. maltophilia*, a protein that condenses numerous fatty acyl CoAs. A protein expression and ketone-production screen of the remaining *oleA* genes may find a more beneficial OleA protein to base hydrocarbon production in *Shewanella* upon. An unscreened OleA protein may allow an *S. oneidensis* construct to produce different ketones that are more useful for industrial or commercial use. The untested protein could also be more stable in the *Shewanella* background.

7.2.2 OleA protein optimization

Dr. Neissa Pinzon has found that Nile Red can be used to detect an increase of ketones in bacterial cultures. Currently transposon mutagenesis coupled to a Nile Red assay is being utilized to identify genes involved in ketone production in the *S. oneidensis* constructs. Other mutagenic techniques could be used to directly mutate OleA to produce different ketone products. One technique not currently being exploited is DNA shuffling. DNA shuffling allows one to propagate mutants in a directed evolution manner through mixing its DNA structure with other homologs. This technique may allow the creation of a new protein sequence that (1) has higher stability in the *S. oneidensis* background, (2) condenses a wider assortment of fatty acyl CoAs, or (3) condenses very specific fatty acyl CoA chains. The former two outcomes may

result in an increase in ketone accumulation in the bacterial constructs while the latter scenario may result in a decrease. Any scenario could result in better product(s) for industrial use.All three scenarios could be detected using a Nile Red assay coupled to gas chromatography-mass spectroscopy analysis.

7.2.3 *Ralstonia* expression

Ralstonia eutropha is a lithotrophic bacterium. It is able to use hydrogen as an electron donor and oxygen as an electron acceptor to grow, as well as fix carbon dioxide. In Chapter 5 I mentioned *R. eutropha* was used to help identify whether ketone production by the *Chloroflexus aurantiacus* OleA protein was dependent upon fatty acyl CoA availability or OleA protein specificity. While the *C. aurantiacus oleA* gene was successfully cloned into *R. eutropha,* ketone production was only observed when cells were grown under high nutrient conditions. Trouble-shooting this system for production of ketone under the hydrogen/oxygen/CO_2 environment may be beneficial.

References

1. **Abboud, R., R. Popa, V. Souza-Egipsy, C. S. Giometti, S. Tollaksen, J. J. Mosher, R. H. Findlay, and K. H. Nealson.** 2005. Low-temperature growth of *Shewanella oneidensis* MR-1. Applied and Environmental Microbiology **71:**811-816.

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

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

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

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

6. **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.

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

8. **Atsumi, S., W. Higashide, and J. C. Liao.** 2009. Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. Nature Biotechnology **27:**1177-U142.

9. **Bancroft, I., and R. J. Smith.** 1988. An analysis of restriction endonuclease sites in cyanophages infecting the heterocystous cyanobacteria *Anabaena* and *Nostoc*. Journal of General Virology **69:**739-743.

10. **Baron, D., E. LaBelle, D. Coursolle, J. A. Gralnick, and D. R. Bond.** 2009. Electrochemical measurement of electron transfer kinetics by *Shewanella oneidensis* MR-1. Journal of Biological Chemistry **284:**28865-28873.

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

12. **Bauld, J., and J. T. Staley.** 1976. *Planctomyces maris sp. nov.*, marine isolate of the Planctomyces-Blastocaulis group of budding bacteria. Journal of General Microbiology **97:**45-55.

13. **Belder, J. B., M. Nguyen-Disteche, N. Houba-Herin, J. M. Ghuysen, I. N. Maruyama, H. Hara, Y. Hirota, and M. Inouye.** 1988. Overexpression, solubilization and refolding of a genetically engineered derivative of the penicillin-binding protein 3 of *Escherichia coli* K12. Mol Microbiol **2:**519-25.

14. **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.

15. **Bligh, E. G., and W. J. Dyer.** 1950. A rapid method for total lipid extraction and purification. Can. J. Biochem. Physiol. **70:**517–525.

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

17. **Bowman, J. P., S. A. McCammon, D. S. Nichols, J. H. Skerratt, S. M. Rea, P. D. Nichols, and T. A. McMeekin.** 1997. *Shewanella gelidimarina sp. nov.* and *Shewanella frigidimarina sp. nov.*, novel Antarctic species with the ability to produce eicosapentaenoic acid (20:5 omega 3) and grow anaerobically by dissimilatory Fe(III) reduction. International Journal of Systematic Bacteriology **47:**1040-1047.

18. **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.

19. **Brettar, I., R. Christen, and M. G. Hofle.** 2002. *Shewanella denitrificans* sp nov., a vigorously denitrifying bacterium isolated from the oxic-anoxic interface of the Gotland Deep in the central Baltic Sea. International Journal of Systematic and Evolutionary Microbiology **52:**2211-2217.

20. **Brown, C. W., and J. A. Yoder.** 1994. *Coccolithophorid* blooms in the global ocean. Journal of Geophysical Research-Oceans **99:**7467-7482.

21. **Buikema, W. J., and R. Haselkorn.** 1991. Isolation and complementation of nitrogen fixation mutants of the cyanobacterium *Anabaena* sp. strain PCC7120. Journal of Bacteriology **173:**1879-1885.

22. **Burnes, T. A., R. A. Blanchette, and R. L. Farrell.** 2000. Bacterial biodegradation of extractives and patterns of bordered pit membrane attack in pine wood. Applied and Environmental Microbiology **66:**5201-5205.

23. **Caccavo, F., D. J. Lonergan, D. R. Lovley, M. Davis, J. F. Stolz, and M. J. McInerney.** 1994. *Geobacter sulfurreducens sp. nov*., a hydrogen-oxidizing and acetate-oxidizing dissimilatory metal-reducing microorganism. Applied and Environmental Microbiology **60:**3752-3759.

24. **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.

25. **Chang, Y.-J., S.-H. Song, S.-H. Park, and S.-U. Kim.** 2000. Amorpha-4,11-diene synthase of *Artemisia annua*: cDNA isolation and bacterial expression of a terpene synthase involved in Artemisinin biosynthesis. Archives of Biochemistry and Biophysics **383:**178-184.

26. **Channon, H. J., and A. C. Chibnall.** 1929. The ether-soluble substance of cabbage leaf cytoplasm. V. The isolation of *n*-nonacosane and di-*n*-tetradecylketone. Biochem. J. **23:**168-175.

27. **Chase, J., K. Touhara, G. Prestwich, C. Schal, and G. J. Blomquist.** 1992. Biosynthesis and endocrine regulation of the production of the German cockroach sex pheromone, 3,11-dimethylnonacosan-2-one. Proc. Natl. Acad. Sci. USA **89:**6050-6054.

28. **Chauvel, A., and G. Lefebvre.** 1989. Petrochemical Processes. Editions Technip, Paris, France.

29. **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.

30. **Chenna, R., H. Sugawara, T. Koike, R. Lopez, T. J. Gibson, D. G. Higgins, and J. D. Thompson.** 2003. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Research **31:**497-500.

31. **Choo, Y. J., K. Lee, J. Song, and J. C. Cho.** 2007. *Puniceicoccus vermicola gen. nov., sp nov.*, a novel marine bacterium, and description of *Puniceicoccaceae fam. nov., Puniceicoccales ord. nov., Opitutaceae fam. nov., Opitutales ord. nov* and *Opitutae classis nov* in the phylum 'Verrucomicrobia'. International Journal of Systematic and Evolutionary Microbiology **57:**532-537.

32. **Colby, S. M., W. R. Alonso, E. J. Katahira, D. J. McGarvey, and R. Croteau.** 1993. 4S-limonene synthase from the oil glands of spearmint (*Mentha spicata*). cDNA isolation, characterization, and bacterial expression of the catalytically active monoterpene cyclase. Journal of Biological Chemistry **268:**23016-23024.

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

34. **Conti, E., N. P. Franks, and P. Brick.** 1996. Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. Structure **4:**287-298.

35. **Daum, M., S. Herrmann, B. Wilkinson, and A. Bechthold.** 2009. Genes and enzymes involved in bacterial isoprenoid biosynthesis. Current Opinion in Chemical Biology **13:**180-188.

36. **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.

37. **Deming, J. W., L. K. Somers, W. L. Straube, D. G. Swartz, and M. T. Macdonell.** 1988. Isolation of an obligately barophilic bacterium and description of a new genus, *Colwellia gen. nov*. Systematic and Applied Microbiology **10:**152-160.

38. **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.

39. **Devarenne, T. P., D. H. Shin, K. Back, S. Yin, and J. Chappell.** 1998. Molecular characterization of tobacco squalene synthase and regulation in response to fungal elicitor. Archives of Biochemistry and Biophysics **349:**205-215.

40. **Ditty, J. L., S. R. Canales, B. E. Anderson, S. B. Williams, and S. S. Golden.** 2005. Stability of the *Synechococcus elongatus* PCC 7942 circadian clock under

directed anti-phase expression of the kai genes. Microbiology-Sgm **151:**2605-2613.

41. **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.

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

43. **Elzer, P. H., M. Kovach, R. W. Phillips, G. T. Robertson, K. M. Peterson, and R. M. Roop.** 1995. *In vivo* and *in vitro* stability of the broad-host-range cloning vector pBBR1MCS in six *Brucella* species. Plasmid **33:**51-57.

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. **Fehler, S. W., and R. J. Light.** 1972. Biosynthesis of methylheptadecanes in *Anabaena variabilis*. *In vitro* incorporation of S-(methyl- 14 C)adenosylmethionine. Biochemistry **11:**2411-6.

46. **Ferrari, M., U. M. Pagnoni, F. Pelizzoni, V. Lukes, and G. Ferrari.** 1971. Terpenoids from *Copaifera langsdorfii*. Phytochemistry **10:**905-907.

47. **Filling, C., K. D. Berndt, J. Benach, S. Knapp, T. Prozorovsk, 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.

48. **Filling, C., E. Nordling, J. Benach, K. D. Berndt, R. Ladenstein, H. Jornvall, and U. Oppermann.** 2001. Structural role of conserved Asn179 in the short-chain dehydrogenase/reductase scaffold. Biochem. Biophys. Res. Commun. **289:**712-717.

49. **Fredrickson, J. K., M. F. Romine, A. S. Beliaev, J. M. Auchtung, M. E. Driscoll, T. S. Gardner, K. H. Nealson, A. L. Osterman, G. Pinchuk, J. L. Reed, D. A. Rodionov, J. L. M. Rodrigues, D. A. Saffarini, M. H. Serres, A. M. Spormann, I. B. Zhulin, and J. M. Tiedje.** 2008. Towards environmental systems biology of *Shewanella*. Nature Reviews Microbiology **6:**592-603.

50. **Fredslund, F., L. Jenner, U. B. Husted, J. Nyborg, G. R. Andersen, and L. Sottrup-Jensen.** 2006. The structure of bovine complement component 3

reveals the basis for thioester function. Journal of Molecular Biology **361:**115-127.

51. **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.

52. **Fujita, K., T. Hagishita, S. Kurita, Y. Kawakura, Y. Kobayashi, A. Matsuyama, and H. Iwahashi.** 2006. The cell structural properties of *Kocuria rhizophila* for aliphatic alcohol exposure. Enzyme and Microbial Technology **39:**511-518.

53. **Fujita, Y., H. Matsuoka, and K. Hirooka.** 2007. Regulation of fatty acid metabolism in bacteria. Molecular Microbiology **66:**829-839.

54. **Galbraith, L., M. H. Jonsson, L. C. Rudhe, and S. G. Wilkinson.** 1999. Lipids and fatty acids of *Burkholderia* and *Ralstonia* species. Fems Microbiology Letters **173:**359-364.

55. **Gorenflo, V., A. Steinbuchel, S. Marose, M. Rieseberg, and T. Scheper.** 1999. Quantification of bacterial polyhydroxyalkanoic acids by Nile red staining. Applied Microbiology and Biotechnology **51:**765-772.

56. **Gough, J., K. Karplus, R. Hughey, and C. Chothia.** 2001. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. Journal of Molecular Biology **313:**903-919.

57. **Gralnick, J. A., H. Vali, D. P. Lies, and D. K. *Newman*.** 2006. Extracellular respiration of dimethyl sulfoxide by *Shewanella* oneidensis strain MR-1. Proceedings of the National Academy of Sciences of the United States of America **103:**4669-4674.

58. **Greene, M. J., and D. M. Gordon.** 2003. Social insects - Cuticular hydrocarbons inform task decisions. Nature **423:**32-32.

59. **Whited, G.M., F.J. Feher, D.A. Benko, M.A. Cervin, G.K. Chotani, J.C. McAuliffe, R.J. LaDuca, E.A. Ben-Shoshan, and K.J. Sanford.** 2010. Development of a gas-phase bioprocess for isoprene-monomer production using metabolic pathway engineering. Industrial Biotechnology **6:**152-163.

60. **Gulick, A. M., X. F. Lu, and D. Dunaway-Mariano.** 2004. Crystal structure of 4-chlorobenzoate : CoA ligase/synthetase in the unliganded and aryl substrate-bound states. Biochemistry **43:**8670-8679.

61. **Gunnarsson, N., A. Eliasson, and J. Nielsen.** 2004. Control of fluxes towards antibiotics and the role of primary metabolism in production of antibiotics, p. 137-178, Molecular Biotechnology of Fungal Beta-Lactam Antibiotics and Related Peptide Synthetases, vol. 88. Springer-Verlag Berlin, Berlin.

62. **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.

63. **Hau, H. H., A. Gilbert, D. Coursolle, and J. A. Gralnick.** 2008. Mechanism and consequences of anaerobic respiration of cobalt by *Shewanella oneidensis* Strain MR-1. Applied and Environmental Microbiology **74:**6880-6886.

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

65. **Hedrick, D. B., A. D. Peacock, D. R. Lovley, T. L. Woodard, K. P. Nevin, P. E. Long, and D. C. White.** 2009. Polar lipid fatty acids, LPS-hydroxy fatty acids, and respiratory quinones of three *Geobacter* strains, and variation with electron acceptor. Journal of Industrial Microbiology & Biotechnology **36:**205-209.

66. **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.

67. **Heidelberg, J. F., I. T. Paulsen, K. E. Nelson, E. J. Gaidos, W. C. Nelson, T. D. Read, J. A. Eisen, R. Seshadri, N. Ward, B. Methe, R. A. Clayton, T. Meyer, A. Tsapin, J. Scott, M. Beanan, L. Brinkac, S. Daugherty, R. T. DeBoy, R. J. Dodson, A. S. Durkin, D. H. Haft, J. F. Kolonay, R. Madupu, J. D. Peterson, L. A. Umayam, O. White, A. M. Wolf, J. Vamathevan, J. Weidman, M. Impraim, K. Lee, K. Berry, C. Lee, J. Mueller, H. Khouri, J. Gill, T. R. Utterback, L. A. McDonald, T. V. Feldblyum, H. O. Smith, J. C. Venter, K. H. Nealson, and C. M. Fraser.** 2002. Genome sequence of the dissimilatory metal ion-reducing bacterium *Shewanella oneidensis*. Nature Biotechnology **20:**1118-1123.

68. **Heldt, H.-W., and F. Heldt.** 2005. Plant biochemistry, 3rd edition ed. Elsevier Academic Press, Burlington, MA.

69. **Hird, F. J. R., and R. H. Symons.** 1962. The mechanism of ketone-body formation from butyrate in rat liver. Biochem. J. **84:**212-216.

70. **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.

71. **Ivanova, E. P., S. Flavier, and R. Christen.** 2004. Phylogenetic relationships among marine *Alteromonas*-like proteobacteria: emended description of the family *Alteromonadaceae* and proposal of *Pseudoalteromonadaceae fam. nov., Colwelliaceae fam. nov., Shewanellaceae fam. nov., Montellaceae fam. nov., Ferrimonadaceae fam. nov., Idiomarinaceae fam. nov* and *Psychromonadaceae fam. nov.* International Journal of Systematic and Evolutionary Microbiology **54:**1773-1788.

72. **Jackson, J. K.** 1998. Recombinant modulation of the *phb*CAB operon copy number in *Ralstonia eutropha* and modification of the precursor selectivity of the *Pseudomonas oleovorans* Polymerase I. University of Minnesota, St. Paul.

73. **Jeong, Y. S., S. K. Song, S. J. Lee, and B. K. Hur.** 2006. The growth and EPA synthesis of *Shewanella oneidensis* MR-1 and expectation of EPA biosynthetic pathway. Biotechnology and Bioprocess Engineering **11:**127-133.

74. **Jez, J. M., M. B. Austin, J. Ferrer, M. E. Bowman, J. Schroder, and J. P. Noel.** 2000. Structural control of polyketide formation in plant-specific polypeptide synthases. Chemistry and Biology **7:**919-930.

75. **Jiang, C. G., S. Y. Kim, and D. Y. Suh.** 2008. Divergent evolution of the thiolase superfamily and chalcone synthase family. Molecular Phylogenetics and Evolution **49:**691-701.

76. **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.

77. **Jones, K. M., and R. Haselkorn.** 2002. Newly identified cytochrome c oxidase operon in the nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC 7120 specifically induced in heterocysts. Journal of Bacteriology **184:**2491-2499.

78. **Jorgensen, B. R., and H. H. Huss.** 1989. Growth and activity of *Shewanella putrefaciens* isolated from spoiling fish. International Journal of Food Microbiology **9:**51-62.

79. **Jornvall, H., B. Persson, M. Krook, S. Atrian, R. Gonzalezduarte, J. Jeffery, and D. Ghosh.** 1995. Short-chain dehydrogenases reductases (SDR). Biochemistry **34:**6003-6013.

80. **Juarez, P., and G. J. Blomquist.** 1992. A microsomal fatty acid synthetase from the integument of *Blattella germanica* synthesizes methyl-branched fatty acids, precursors to hydrocarbon and contact sex pheromone. Arch Biochem Biophys. **293:**333-41.

81. **Kenyon, C. N., and A. M. Gray.** 1974. Preliminary analysis of lipids and fatty-acids of green bacteria and *Chloroflexus aurantiacus*. Journal of Bacteriology **120:**131-138.

82. **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.

83. **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.

84. **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.

85. **Kinoshita, S., K. Kadota, T. Inoue, H. Sawada, and H. Taguchi.** 1988. Purification and properties of 3 kinds of alpha-hydroxysteroid dehydrogenases from *Brevibacterium fuscum* DC33. Journal of Fermentation Technology **66:**145-152.

86. **Kissin, Y. V.** 1993. Catagenesis of light acyclic isoprenoids in petroleum. Organic Geochemistry **20:**1077-1090.

87. **Kissin, Y. V.** 1998. Catagenesis of light aromatic compounds in petroleum. Organic Geochemistry **29:**947-962.

88. **Kissin, Y. V.** 1990. Catagenesis of light cycloalkanes in petroleum. Organic Geochemistry **15:**575-594.

89. **Kovach, M. E., P. H. Elzer, D. S. Hill, G. T. Robertson, M. A. Farris, R. M. Roop, and K. M. Peterson.** 1995. 4 new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic casettes. Gene **166:**175-176.

90. **Kovacs, G., J. Burghardt, S. Pradella, P. Schumann, E. Stackebrandt, and K. Marialigeti.** 1999. *Kocuria palustris sp. nov. and Kocuria rhizophila sp. nov.*, isolated from the rhizoplane of the narrow-leaved cattail (*Typha angustifolia*). International Journal of Systematic Bacteriology **49:**167-173.

91. **Kulichevskaya, I. S., A. O. Ivanova, S. E. Belova, O. I. Baulina, P. L. E. Bodelier, W. I. C. Rijpstra, J. S. S. Damste, G. A. Zavarzin, and S. N. Dedysh.** 2007. *Schlesneria paludicola gen. nov., sp nov.*, the first acidophilic member of the order *Planctomycetales*, from *Sphagnum*-dominated boreal wetlands. International Journal of Systematic and Evolutionary Microbiology **57:**2680-2687.

92. **Kuzma, J., M. Nemecek-Marshall, W. Pollock, and R. Fall.** 1995. Bacteria produce the volatile hydrocarbon isoprene. Curr Microbiol. **30:**97-103.

93. **Lanzón, A., T. Albi, A. Cert, and J. Gracián.** 1994. The hydrocarbon fraction of virgin olive oil and changes resulting from refining. Journal of the American Oil Chemists' Society **71:**285-291.

94. **le Maire, M., P. Champeil, and J. V. Moller.** 2000. Interaction of membrane proteins and lipids with solubilizing detergents. Biochimica Et Biophysica Acta-Biomembranes **1508:**86-111.

95. **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.

96. **Lee, W. H., K. W. Cho, S. Y. Park, K. S. Shin, D. S. Lee, S. K. Hwang, S. J. Seo, J. M. Kim, S. Y. Ghim, B. H. Song, S. H. Lee, and J. G. Kim.** 2008. Identification of psychrophile *Shewanella* sp KMG427 as an eicosapentaenoic acid producer. Journal of Microbiology and Biotechnology **18:**1869-1873.

97. **Lehninger, A. L.** 1978. Biochemistry: the molecular basis of cell structure and function. Worth Publishers, New York, NY.

98. **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.

99. **Loreto, F., and V. Velikova.** 2001. Isoprene produced by leaves protects the photosynthetic apparatus against ozone damage, quenches ozone products, and reduces lipid peroxidation of cellular membranes. Plant Physiology **127:**1781-1787.

100. **Meng, E. C., E. F. Pettersen, G. S. Couch, C. C. Huang, and T. E. Ferrin.** 2006. Tools for integrated sequence-structure analysis with UCSF Chimera. Bmc Bioinformatics **7:**10.

101. **Meriläinen, G., V. Poikela, P. Kursula, and R. K. Wierenga.** 2009. The thiolase reaction mechanism: the importance of Asn316 and His348 for stabilizing the enolate intermediate of the Claisen condensation. Biochemistry **48:**11011-11025.

102. **Metz, J. G., P. Roessler, D. Facciotti, C. Levering, F. Dittrich, M. Lassner, R. Valentine, K. Lardizabal, F. Domergue, A. Yamada, K. Yazawa, V. Knauf, and J. Browse.** 2001. Production of polyunsaturated fatty acids by polyketide synthases in both prokaryotes and eukaryotes. Science **293:**290-293.

103. **Metzler, M. C., M. J. Laine, and S. H. DeBoer.** 1997. Status of molecular biological research on the plant pathogenic genus *Clavibacter*. Fems Microbiology Letters **150:**1-8.

104. **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.

105. **Miyamato, K.** 2001. Microorganisms and process for producing petroleum substitute oil by using these microorganisms. Japanese patent P2001-190323.

106. **Morris, J. H., C. C. Huang, P. C. Babbitt, and T. E. Ferrin.** 2007. StructureViz: linking Cytoscape and UCSF Chimera. Bioinformatics **23:**2345-2347.

107. **Nardini, M., and B. W. Dijkstra.** 1999. alpha/beta hydrolase fold enzymes: the family keeps growing. Current Opinion in Structural Biology **9:**732-737.

108. **Nelson, D. R., and G. J. Blomquist.** 1995. Insect Waxes. Oily Press, Dundee, Scotland.

109. **Nevin, K. P., D. E. Holmes, T. L. Woodard, E. S. Hinlein, D. W. Ostendorf, and D. R. Lovley.** 2005. *Geobacter bemidjiensis sp nov* and *Geobacter psychrophilus sp nov*., two novel Fe(III)-reducing subsurface isolates. International Journal of Systematic and Evolutionary Microbiology **55:**1667-1674.

110. **Nichols, D. S., P. D. Nichols, and T. A. McMeekin.** 1995. A new n-C31/9 polyene hydrocarbon from antarctic bacteria. Fems Microbiology Letters **125:**281-285.

111. **Niederholtmeyer, H., B. T. Wolfstadter, D. F. Savage, P. A. Silver, and J. C. Way.** 2010. Engineering cyanobacteria to synthesize and export hydrophilic products. Applied and Environmental Microbiology **76:**3462-3466.

112. **Nishida, T., R. Hori, N. Morita, and H. Okuyama.** 2010. Membrane eicosapentaenoic acid is involved in the hydrophobicity of bacterial cells and affects the entry of hydrophilic and hydrophobic compounds. Fems Microbiology Letters **306:**91-96.

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

114. **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.

115. **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.

116. **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.

117. **Pedersen, K. J.** 1929. The ketonic decomposition of beta-keto carboxylic acids. Journal of American Chemistry **51:**2098-2107.

118. **Penuelas, J., J. Llusia, D. Asensio, and S. Munne-Bosch.** 2005. Linking isoprene with plant thermotolerance, antioxidants and monoterpene emissions. Plant Cell and Environment **28:**278-286.

119. **Petrov, A. A.** 1987. Petroleum hydrocarbons. Springer-Verlag, New York, NY.

120. **Pierson, B. K., and R. W. Castenholz.** 1974. A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. Arch. Microbiol. **100:**5-24.

121. **Prahl, F. G., and S. G. Wakeham.** 1987. Calibration of unsaturation patterns in long-chain ketone compositions for paleotemperature assessment. Nature **330:**367-369.

122. **Qian, Z., C. J. Fields, Y. Yu, and S. Lutz.** 2007. Recent progress in engineering alpha/beta hydrolase-fold family members. Biotechnology Journal **2:**192-200.

123. **Reed, J., D. Quilici, G. Blomquist, and R. Reitz.** 1995. Proposed mechanism for the cytochrome-P450 catalyzed conversion of aldehydes to hydrocarbons in the housefly, *Musca domestica*. Biochemistry **34:**26221–27.

124. **Reed, J. R., D. Vanderwel, S. Choi, J. G. Pomonis, R. C. Reitz, and G. J. Blomquist.** 1994. Unusual mechanism of hydrocarbon formation in the housefly: cytochrome P450 converts aldehyde to the sex pheromone component (Z)-9-tricosene and CO2. Proceedings of the National Academy of Sciences of the United States of America **91:**10000-10004.

125. **Rezanka, T., J. Zahradnik, and M. Podojil.** 1982. Hydrocarbons in green and blue-green algae. Folia Microbiologica **27:**450-454.

126. **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.

127. **Risser, D. D., and S. M. Callahan.** 2007. Mutagenesis of hetR reveals amino acids necessary for HetR function in the heterocystous cyanobacterium *Anabaena* sp strain PCC 7120. Journal of Bacteriology **189:**2460-2467.

128. **Rohdich, F., S. Hecht, K. Gärtner, P. Adam, C. Krieger, S. Amslinger, D. Arigoni, A. Bacher, and W. Eisenreich.** 2002. Studies on the nonmevalonate terpene biosynthetic pathway: Metabolic role of IspH (LytB) protein. Proceedings of the National Academy of Sciences of the United States of America **99:**1158-1163.

129. **Rollefson, J. B., C. E. Levar, and D. R. Bond.** 2009. Identification of genes involved in biofilm formation and respiration via mini-himar transposon mutagenesis of *Geobacter sulfurreducens*. Journal of Bacteriology **191:**4207-4217.

130. **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.

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

132. **Saito, N., and S. Sugiyama.** 1964. Production of dihydroxyacetone and other ketoses by *Brevibacterium fuscum*. Agric. Biol. Chem. **28:**48-55.

133. **Saltikov, C. W., and D. K. Newman.** 2003. Genetic identification of a respiratory arsenate reductase. Proceedings of the National Academy of Sciences of the United States of America **100:**10983-10988.

134. **Sayers, E. W., T. Barrett, D. A. Benson, E. Bolton, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, S. Federhen, M. Feolo, L. Y. Geer, W. Helmberg, Y. Kapustin, D. Landsman, D. J. Lipman, Z. Y. Lu, T. L. Madden, T. Madej, D. R. Maglott, A. Marchler-Bauer, V. Miller, I. Mizrachi, J. Ostell, A. Panchenko, K. D. Pruitt, G. D. Schuler, E. Sequeira, S. T. Sherry, M. Shumway, K. Sirotkin, D. Slotta, A. Souvorov, G. Starchenko, T. A. Tatusova, L. Wagner, Y. L. Wang, W. J. Wilbur, E. Yaschenko, and J. Ye.** 2010. Database resources of the National Center for Biotechnology Information. Nucleic Acids Research **38:**D5-D16.

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

136. **Schmidt, T. D.** Accessed 23 November 2009. Genome Project: *Opitutaceae* bacterium TAV2. DOE Joint Genome Institute, Walnut Creek, CA. http://genome.jgi-psf.org/opiba/opiba.info.html.

137. **Seffernick, J. L., S. K. Samanta, T. M. Louie, L. P. Wackett, and M. Subramanian.** 2009. Investigative mining of sequence data for novel enzymes: a case study with nitrilases. Journal of Biotechnology **143:**7-26.

138. **Sharkey, T. D., S. Yeh, A. E. Wiberley, T. G. Falbel, D. M. Gong, and D. E. Fernandez.** 2005. Evolution of the isoprene biosynthetic pathway in *Kudzu*. Plant Physiology **137:**700-712.

139. **Shyu, J. B. H., D. P. Lies, and D. K. Newman.** 2002. Protective role of *tolC* in efflux of the electron shuttle anthraquinone-2,6-disulfonate. Journal of Bacteriology **184:**1806-1810.

140. **Simoneit, B. R. T., and P. F. Lonsdale.** 1982. Hydrothermal petroelum in mineralized mounds at the seabed of Guaymas Basin. Nature **295:**198-202.

141. **Snyder, L., and W. Champness.** 1997. Molecular Genetics of Bacteria. American Society for Microbiology, Washington, D.C.

142. **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.

143. **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.

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

145. **Steussy, C. N., A. D. Robison, A. M. Tetrick, J. T. Knight, V. W. Rodwell, C. V. Stauffacher, and A. L. Sutherlin.** 2006. A structural limitation on enzyme activity: The case of HMG-CoA synthase. Biochemistry **45:**14407-14414.

146. **Stevenson, B. S., S. A. Eichorst, J. T. Wertz, T. M. Schmidt, and J. A. Breznak.** 2004. New strategies for cultivation and detection of previously uncultured microbes. Applied and Environmental Microbiology **70:**4748-4755.

147. **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 **2:**337-348.

148. **Sukchawalit, R., P. Vattanaviboon, R. Sallabhan, and S. Mongkolsuk.** 1999. Construction and characterization of regulated L-arabinose-inducible broad host range expression vectors in *Xanthomonas*. Fems Microbiology Letters **181:**217-223.

149. **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.

150. **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.

151. **Szkopinska, A., E. Swiezewska, and F. Karst.** 2000. The regulation of activity of main mevalonic acid pathway enzymes: Farnesyl diphosphate synthase, 3-hydroxy-3-methylglutaryl-CoA reductase, and squalene synthase in yeast *Saccharomyces cerevisiae*. Biochemical and Biophysical Research Communications **267:**473-477.

152. **Takarada, H., M. Sekine, H. Kosugi, Y. Matsuo, T. Fujisawa, S. Omata, E. Kishi, A. Shimizu, N. Tsukatani, S. Tanikawa, N. Fujita, and S. Harayama.** 2008. Complete genome sequence of the soil actinomycete *Kocuria rhizophila*. Journal of Bacteriology **190:**4139-4146.

153. **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.

154. **Thoden, J. B., A. M. Gulick, and H. M. Holden.** 1997. Molecular structures of the S124A, S124T, and S124V site-directed mutants of UDP-galactose 4-epimerase from *Escherichia coli*. Biochemistry **36:**10685-10695.

155. **Tillman, J. A., S. J. Seybold, R. A. Jurenka, and G. J. Blomquist.** 1999. Insect pheromones - an overview of biosynthesis and endocrine regulation. Insect Biochemistry and Molecular Biology **29:**481-514.

156. **Tissot, B. P., and D. H. Welte.** 1978. Petroleum formation and occurrence: a new approach to oil and gas exploration. Springer-Verlag, New York, NY.

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

158. **Tornabene, T. G., E. Gelphi, 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.

159. **Tornabene, T. G., and S. P. Markey.** 1971. Characterization of branched monounsaturated hydrocarbons of *Sarcina lutea* and *Sarcina flava*. Lipids **3:**190-195.

160. **Tornabene, T. G., S. J. Morrison, and W. E. Kloos.** 1970. Aliphatic hydrocarbon contents of various members of the family *Micrococcaceae*. Lipids **5:**929-934.

161. **Tornabene, T. G., and S. L. Peterson.** 1978. *Pseudomonas maltophilia:* identification of the hydrocarbons, glycerides, and glycolipoproteins of cellular lipids. Canadian Journal of Microbiology **254:**525-532.

162. **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.

163. **van der Meer, M. T. J., S. Schouten, D. E. van Dongen, W. I. C. Rijpstra, G. Fuchs, J. S. S. Damste, J. W. de Leeuw, and D. M. Ward.** 2001. Biosynthetic controls an the C-13 contents of organic components in the photoautotrophic bacterium *Chloroflexus aurantiacus*. Journal of Biological Chemistry **276:**10971-10976.

164. **van der Meer, M. T. J., S. Schouten, D. M. Ward, J. A. J. Geenevasen, and J. S. S. Damste.** 1999. All-cis hentriaconta-9,15,22-triene in microbial mats formed by the phototrophic prokaryote Chloroflexus. Organic Geochemistry **30:**1585-1587.

165. **Vauterin, L., B. Hoste, K. Kersters, and J. Swings.** 1995. Reclassification of *Xanthomonas*. International Journal of Systematic Bacteriology **45:**472-489.

166. **Venkateswaran, K., M. E. Dollhopf, R. Aller, E. Stackebrandt, and K. H. Nealson.** 1998. *Shewanella amazonensis sp. nov*., a novel metal-reducing facultative anaerobe from Amazonian shelf muds. International Journal of Systematic Bacteriology **48:**965-972.

167. **Venkateswaran, K., D. P. Moser, M. E. Dollhopf, D. P. Lies, D. A. Saffarini, B. J. MacGregor, D. B. Ringelberg, D. C. White, M. Nishijima, H. Sano, J. Burghardt, E. Stackebrandt, and K. H. Nealson.** 1999. Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis sp. nov.* International Journal of Systematic Bacteriology **49:**705-724.

168. **Verschueren, K. H. G., F. Seljee, H. J. Rozeboom, K. H. Kalk, and B. W. Dijkstra.** 1993. Crystallographic analysis of the catalytic mechanism of haloalkane dehydrogenase. Nature **363:**693-698.

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

170. **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.

171. **WardRainey, N., F. A. Rainey, and E. Stackebrandt.** 1997. The presence of a dnaK (HSP70) multigene family in members of the orders *Planctomycetales* and *Verrucomicrobiales*. Journal of Bacteriology **179:**6360-6366.

172. **Withers, S. T., S. S. Gottlieb, B. Lieu, J. Newman, and J. D. Keasling.** 2007. Identification of isoprentenol biosynthetic genes from *Bacillus subtilis* by a screening method based on isoprenoid precursor toxicity. Applied and Environmental Microbiology **73:**6277-6283.

173. **Withers, S. T., and J. D. Keasling.** 2007. Biosynthesis and engineering of isoprenoid small molecules. Applied Microbiology and Biotechnology **73:**980-990.

174. **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.

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

176. **Yumoto, I., K. Kawasaki, H. Iwata, H. Matsuyama, and H. Okuyama.** 1998. Assignment of *Vibrio* sp. strain ABE-1 to *Colwellia maris sp. nov*., a new psychrophilic bacterium. International Journal of Systematic Bacteriology **48:**1357-1362.

177. **Ziemke, F., M. G. Hofle, J. Lalucat, and R. Rossello-Mora.** 1998. Reclassification of *Shewanella putrefaciens* Owen's genomic group II as *Shewanella baltica sp. nov*. International Journal of Systematic Bacteriology **48:**179-186.

Appendix A: Cyanobacteria as platforms for hydrocarbon production

A.1 Overview

Two separate attempts were made to construct a cyanobacterium strain expressing the *S. maltophilia oleA* gene by: (1) incorporation of the *oleA* directly into the genome and (2) expression in *trans* on a plasmid. These bacterial constructs were never successfully made. A truncated *oleA* was incorporated into the genome of *Synechococcus elongatus*, but a plasmid containing the *oleA* gene was never stably transferred into *Anabaena*.

A.2 Introduction

An ideal platform for hydrocarbon production would involve an organism(s) that can utilize carbon sources created from carbon dioxide to produce biomass and ketone. The Wackett Laboratory has hypothesized that this could be achieved through use of a co-culture of cyanobacteria and *Shewanella*. Using sunlight and CO_2, cyanobacteria would create a carbon source that *Shewanella* would then use to produce the compounds of interest. Recently, the Silver Laboratory of Harvard Medical engineered a *Synechococcus* 7002 strain that produced lactic acid (111), a preferred carbon source for *Shewanella oneidensis* MR-1. This is an important first step toward making a stable co-culture.

Hydrocarbon biosynthesis within the CO_2-fixing cyanobacterium is likely to be more efficient than the co-culture, since the latter requires high levels of lactate export by the cyanobacterium and rapid uptake by the *Shewanella*. This Chapter looks at two

different attempts to construct a cyanobacterium that can produce ketone from CO_2: (1) where the *S. maltophilia oleA* gene is placed under the control of a strong promoter and inserted into the genome of *Synechococcus elongatus*, and (2) where the *S. maltophilia oleA* gene is expressed heterologously in *Anabaena* FC9072.

A.3 Methods

A.3.1 Bacterial strains, culture conditions, and growth

Synecococcus elongatus was donated by the Olszewski Laboratory from the University of Minnesota. *Anabaena* FC9072 was obtained from the Gleason Laboratory at the University of Minnesota. The pSMC188 was obtained from the Callahan Laboratory from the University of Hawaii. *Synecococcus* and *Anabaena* were routinely grown on M9 media at room temperature under light.

For maintenance of certain *Synechococcus* and *Anabaena* constructs, 50 µg/ml of kanamycin and 30 µg/ml of neomycin were added to the medium. For selection of *oleA* gene constructs, kanamycin was routinely added to a final concentration of 50 µg/ml or neomycin was added to a final concentration of 30 µg/ml. All *Escherichia coli* strains were grown aerobically at 37°C in LB. When appropriate, ampacillin (50 µg/ml) and chloramphenicol (10 µg/ml) were added to the media.

A.3.2 Heterologous gene expression

All primers used in this study are listed in Table A.1. For the expression of *S. maltophilia oleA* gene in *Synechococcus elongatus*, the Neutral Site II of the *S. elongatus* genome was utilized along with the Gateway 3-part plasmid construction system (Invitrogen) as described by Ditty *et al.* with a few modifications (40). Briefly,

S. maltophilia oleA was amplified using primers S.m.CompF and S.m.CompR containing the EcoRI and HindIII restriction sites respectively. The resulting product was ligated into the pTrc99 plasmid. Upstream and downstream portions of the Neutral Site II, along with the pTrc-*S. maltophilia oleA*, were amplified using primers attB4, attB1r, attB2r, attB3, attB1, and attB2 (as described by the Gateway Technology Manual). Using the protocol described in the Gateway Technology Manual, the three DNA fragments were introduced into the gateway suicide vector prior to being incubated with the *Synechococcus elongatus*.

For the expression of *S. maltophilia oleA* gene in *Anabaena*, the *oleA* was amplified using the S.m.CompF and the S.m.CompRBam primers containing the EcoRI and BamHI restriction sites. Resulting PCR products were ligated into the Strataclone cloning system followed by ligation into the pSMC188 plasmid. Conjugation with *Anabaena* was performed as previously described (77). Constructs were introduced into *E. coli* UC585. A 1.5 ml overnight (supplemented with ampacillin and chloramphenicol) was pelleted, washed 3 times with phosphate-buffered saline, and mixed with 10µl of the *Anabaena* prior to plating onto a filter lying upon M9 media plates. After two days the filter was transferred to an M9 media plate supplemented with neomycin. After 2 weeks colonies were picked and analyzed using colony PCR.

Table A1 primers used in Chapter A

Primer name	Sequence
S.m.CompF	ATCAGTGAATTCAACCTCGATGCTCTTCAAGAATGTCTC
S.m.CompR	AGTACTAAGCTTGAAGATCATCGCTGTCCGTCGCGAGC
S.m.CompRBam	AGTTAAGGATCCGAAGATCATCGCTGTCCGTCGCG
attB4	GGGGACAACTTTGTATAGAAAAGTTGGGATTCCGATCCGGTTTTGGCCGGTACG
attB1r	GGGGACTGCTTTTTTGTACAAACTTGCAGTAAGAGCGATCGCGCTGGGACGATG
attB2r	GGGGACAGCTTTCTTGTACAAAGTGGGGTGCTGAATCACATTTCCCTTGTCCATT
attB3	GGGGACAACTTTGTATAATAAAGTTGCGAATTGCCGGACTGGGAAGCGATCGCG
attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTGGGTTTGACAGCTTATCATCGACTGCAC
attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTGATCAGGCTGAAAATCTTCTCTCATCCG

A.4 Results and Discussion

A.4.1 Attempt to create a *S. elongatus* strain stably expressing the *S. maltophilia oleA* gene

A *S. elongatus* strain was obtained and the attempt to incorporate the *S. maltophilia oleA* gene under control of the promoter pTrc stably into Neutral Site II was tried. Previous work found pTrc to be a strong lactose-inducible promoter in *S. elongatus* (8). A prior report also found that the Gateway Technology System could stably incorporate a gene of interest into *S. elongatus* Neutral Sites (40). The construction of the *S. elongatus*:pTrc-*oleA* strain was never achieved. Kanamycin-resistant colonies of *S. elongatus* were obtained, but colony PCR analysis of drug-resistant colonies revealed that the *oleA* genes were all truncated (Figure A.1). A total of 20 colonies were tested for *oleA* gene insertion. All PCR products were 600 to 700 base pairs, as compared to the expected wildtype size of 1.1 kilobases. Samples 3 and 5 were sequenced by the BioMedical Genomics Center (University of Minnesota, St. Paul, MN). Sequences are listed in Table A.2.

Table A.2 DNA sequences of truncated gene insertions

DNA Sequence of Sample 3 from Figure A.1.

GGCACACGTTGATGCGCCGCATACGCTGACGACCAAGGAAATCAAC
GAGCGGCTGCAGCCGACGCTCGATCGCCTTGGTATCCGCACCGACGT
GCTCGGCGATATCGCCGGCATCCATGCCCGCCGCCTGTGGGACAACG
GCGTGCTGGCCTCCTCATGGCGCGCTCGGAGCTGGTGCCGGACGCC
CCGCGCTACAAGGGCGGCGTGACCCGCTCGGCCACCGAGTGGAACC
AGCTGTGCCTGGGCAACCTGGACCGCATGGTCACCGACACCCGCCT
GCTGCTGATCGAAGGCATCAAGCTGGCGCAGAAGACCTTCACTGCC
GCCAAGATCGCACTCGGCTGGGCCGTGGAAGAGCTGGACCAGTTCG
TGATCCACCAGGTCAGCCAGCCGCACACCGCTGCGTTCATCAAGAA
CTTCGGCATCGACCCGAAGAAGGTCATGACCATCTTCGGCGAGCAC
GGCAACATCGGCCCGGCCTCGGTGCCGATCGTGCTGAGCAAGCTCA
AGCAGCTGGGCAAGCTGAAGAAGGGCGATCGCATCGCGCTTCTGGG
CATCGGCTCGGGCCTGAACTGCTCGATGGCTGAAGTGGTCTGGTAAA
AAAGTCACCGCTGACCTGC

DNA Sequence of Sample 5 from Figure A.1.

GGCCTGGCACACGTTGATGCGCCGCATACGCTGACGACCAAGGAAA
TCAACGAGCGGCTGCAGCCGACGCTCGATCGCCTTGGTATCCGCACC
GACGTGCTCGGCGATATCGCCGGCATCCATGCCCGCCGCCTGTGGGA
CAACGGCGTGCTGGCCTCCGACGCCGCCACCATGGCCGGCCGCAAG
GCGCTGGAAGACGCAGGCATCAATGCGACGCAGGTCGGCCTGCTGG
TCAACACCTCGGTCAGCCGCGACTACCTGGAGCCGTCCACCGCGTC
GATCGTCTCCGGCAACCTCGGCGTCAGCGAGCGCAGAAGACCTTCA
CTGCCGCCAAGATCGCACTCGGCTGGGCCGTGGAAGAGCTGGACCA
GTTCGTGATCCACCAGGTCAGCCAGCCGCACACCGCTGCGTTCATCA
AGAACTTCGGCATCGACCCGAAGAAGGTCATGACCATCTTCGGCGA
GCACGGCAACATCGGCCCGGCCTCGGTGCCGATCGTGCTGAGCAAG
CTCAAGCAGCTGGGCAAGCTGAAGAAGGGCGATCGCATCGCGCTTC
TGGGCATCGGCTCGGGCCTGAACTGCTCGATGGCTGAAGTGGTCTG
GTAAAAAAGTCACCGCTGACCTGCTGCGCGA

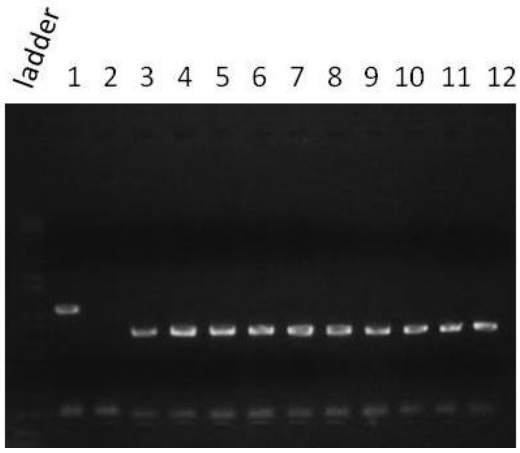


Figure A.1 Colony PCR analysis of *oleA* genes from *Synechococcus elongatus* incubated with Gateway vector-*S. maltophilia oleA*. Lanes (1) Wildtype *oleA* from *S. maltophilia*, (2) wildtype *S. elongatus*, (3-12) *S. elongatus* colonies growing on M9 media supplemented with 50 µg/ml kanamycin.

A gene truncation event had been observed previously, though it was reported to occur in the *E. coli* host strain used for suicide vector creation (personal communications with Invitrogen Customer Services concerning Gateway Technology). A second attempt was made to create the construct where the suicide vector was screened prior to incubation with *S. elongatus*. As depicted in figure A.2, the suicide vectors had the expected *S. maltophilia oleA* gene. Resulting kanamycin-resistant *S. elongatus* all contained truncated *oleA* genes as analyzed through colony PCR analysis.

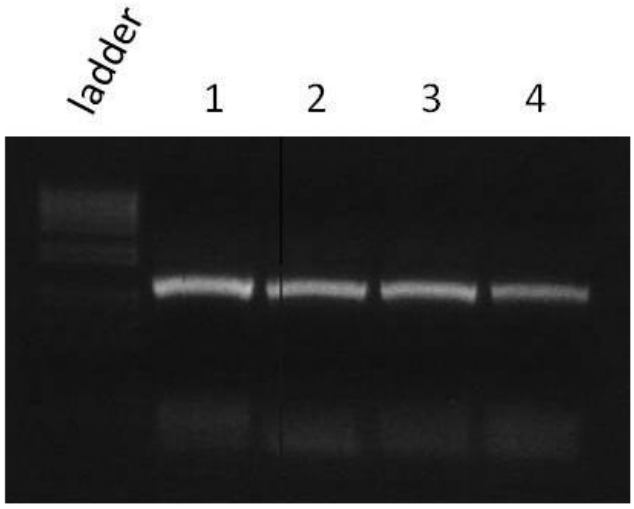


Figure A.2 PCR analysis of *S. elongatus* entry vector containing *S. maltophilia oleA*. All lanes are separate *E. coli* colonies that grew on LB supplemented with 50 µg/ml kanamycin. Expected PCR amplicon = 1.1 KB.

The truncation event may be caused directly by *S. elongatus*. *S. elongatus* is known for its ability to modify foreign DNA (9). There is still the possibility that a cyanobacterium could successfully express the *oleA* gene. A protein expression and ketone production screen of alternative *oleA* genes in the cyanobacteria backgrounds may be beneficial.

A.4.2 Attempt to create an *Anabaena* strain heterologously expressing the *S. maltophilia oleA* gene

An *Anabaena* strain was obtained and the attempt to incorporate the *S. maltophilia oleA* gene heterologously was made. pSMC188 was obtained from the Callahan Laboratory (University of Hawaii). This vector contains the Nir promoter, a

nitrate/nitrite inducible promoter. Previous groups have found that this plasmid can drive expression of genes of interest in *Anabaena* strains (127).

No *Anabaena* colonies grew when plated onto M9 agar plates supplemented with 30µg/ml neomycin, the recommended concentration of drug for selection (E. Reynolds, University of Minnesota, personal communication). Neomycin-resistant colonies did form when neomycin was lowered from 30µg/ml to 10µg/ml. PCR of the resistant colonies revealed that the resistant *Anabaena* did not contain any *oleA* gene (Figure A.3).

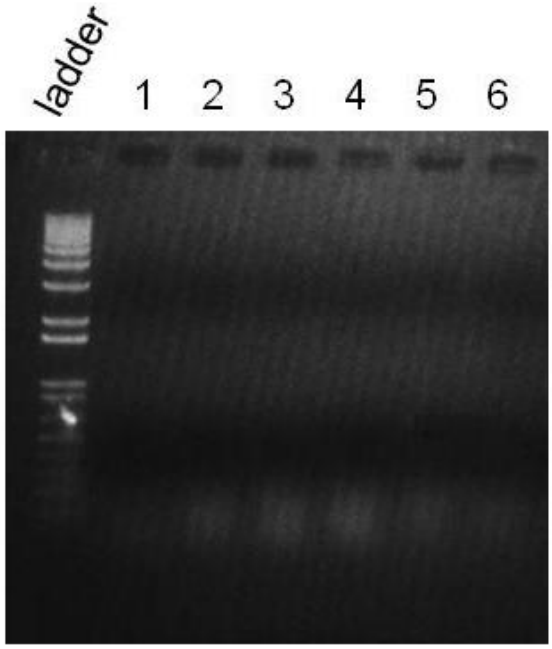


Figure A.3 PCR analysis of *Anabaena* strains resistant to 10µg/ml neomycin. All lanes (1-6) are separate *Anabaena* colony PCR reactions.

Attempts were made to increase the chance of conjugation by varying the time of the biparental mating. *Anabaena* cells were incubated with the carrier *E. coli* for 48,

72 and 96 hours. No neomycin-resistant *Anabaena* grew on M9 media supplemented with 30µg/ml neomycin. The amount of *Anabaena* cells used for conjugation was also increased to try to give a better chance for conjugation. 10µl, 20µl, 30µl, 40µl, and 50µl of concentrated *Anabaena* cells were added to pelleted *E. coli* UC585 containing the pSMC188-*oleA*. Mixtures were allowed to incubate together for 48, 72, and 96 hours prior to plating onto M9 plates supplemented with 30µg/ml neomycin. No neomycin-resistant colonies grew.

A previous observation in the Wackett Laboratory found that the pSMC188 plasmid was not stable in an *Anabaena* strain (E. Reynolds, University of Minnesota, personal communication). Potentially the *Anabaena* is rejecting the plasmid quickly after conjugation. There is the possibility that the plasmid may not be entering the *Anabaena* at all; conjugation may have failed. *E. coli* UC585 is used for biparental mating. Other *E. coli* strains may be used to achieve conjugation. A triparental mating using an plasmid carrier strain (example: *E. coli* DH5α) and a conjugation strain (example: *E. coli* HB101(RP-1)) could be used as an alternative conjugational technique (21).

Appendix B: Ketone Selection

B.1 Overview

In an attempt to find a reagent that could select for ketone-producing *Shewanella oneidensis* strains, various detergents, antibiotics, and dyes were added to cultures grown on liquid or solid media. While some initial results appeared promising, no reagent was identified that could differentiate between producing and non-producing bacteria.

B.2 Introduction

An overall goal of the laboratory is to have *Shewanella* cells produce as much ketone as possible. To accomplish this, one needs to have OleA expressed at high levels, have fatty acid metabolism making the right fatty acyl CoA chains, and have metabolism streamlined to feed most carbon into those pathways. One can genetically manipulate *S. oneidensis* to over-express enzymes (OleA and acetyl-CoA carboxylase are examples) or to knock out others (acyl-CoA dehydrogenase is an example) to accomplish some of this. However, cells are complex and combinatorial methods are very useful. For example mutagenesis via chemicals or transposons can help identify genes necessary for precursor formation (57, 63, 139). It is necessary to be able to pull a small number of useful mutants out of a large number of neutral or deleterious changes when using mutagenesis techniques though. To measure ketone production routinely we use GC-MS. However, GC-MS is very slow and laborious. To obtain one excellent mutant out of millions, one needs rapid diagnostic methods. One such method is screening.

Currently, a high throughput screen for ketone production using Nile Red has been developed. Nile Red is a hydrophobic fluorescent dye that fluoresces at wavelengths greater than 600 nm when excited by light wavelengths of 540 nm (55). Preliminary results in the laboratory are promising. Dr. Neissa Pinzon has shown that Nile Red-treated ketone-producing bacteria appear brighter than control strains when subjected to fluorescent microscopy. A 96-well plate assay using Nile Red is currently being used to screen for over-producing ketone strains of *Shewanella oneidensis* mutants created using transposon mutagenesis.

Screening is limited to thousands or tens of thousands of bacterial mutants. The best mutants may be multiply mutated, and that may be exceedingly rare; chances of finding a multiply mutated bacterium by screening 10,000 mutants are miniscule. Selection may improve success. Natural mutations in bacteria occur frequently. Therefore even in a small culture there are often thousands of spontaneous mutations. When a strong selective pressure is applied on a culture—such as an inhibitory antibiotic—any mutant that has enhanced survival or growth rate advantage under that selective pressure will come to dominate the culture. That mutant strain can then be easily "selected" for and the overall process is called selection. For selection to be effective, an environment where ketone production increases cell viability is needed. Selection is a very robust technique. Whereas screening allows the isolation of a single mutant among thousands of parental bacteria, selection allows the isolation of a mutant among millions or billions of other organisms.

Screening and selection are not mutually exclusive. One might do the following: (1) engineer a strain with the *pfa*- and *fadE*-gene deletion background that

also over-expresses the OleA protein, then (2) subject the strain to mutagenesis and screening, followed by (3) passing it through continuous selective pressure to select for increased ketone production. In this study, various compounds were tested in an attempt to identify one that could be used as a selective reagent for ketone production. These compounds included detergents, antibiotics, oxidizing agents, and hydrophobic dyes. Different hypotheses were tested: (1) that ketones protect the cell membrane from ionic detergent denaturation of membrane proteins, (2) that ketones protect the cell membrane from non-ionic detergent disruption of lipid-lipid interactions, and (3) that ketones protect the cell against hydrophobic antibiotics or oxidizing reagents.

B.3 Methods

B.3.1 Bacteria used

Bacteria used in this study include *Shewanella oneidensis* MR-1 along with *S. oneidensis ΔoleABCD*, *S. oneidensis ΔoleA/ΔpfaA* double knockout, *S. oneidensis ΔoleABCD* heterologously expressing the *S. maltophilia oleA* gene on the pBBR1MCS2 plasmid, and *S. oneidensis ΔoleABCD* heterologously expressing the *S. maltophilia oleA* gene on the pBBAD22K plasmid. All strains were previously constructed (150, Chapter 6). All organisms were grown at 30^{o}C, with kanamycin added to 50 µg/ml when necessary. L-arabinose was added to 1% final concentration when required for induction of the pBBAD22K constructs.

B.3.2 96-well plate

For certain experiments, bacteria were grown in 96-well plates. Bacteria were inoculated to a starting OD 600 of 0.05A and grown at 30^{o}C. The highest inhibitory level of additive was noted for each organism.

B.3.3 Aerosolized plates

For certain experiments, volatile compounds were added to the lids of agar plates. After bacteria were plated onto LB plates with 1% L-arabinose and inverted, crystals of compounds were sprinkled onto the inner lid of the plates. Plates were sealed and placed at 30^{o}C. Differences in growth patterns were noted.

B.3.4 Gradient Plates

For certain experiments, bacteria were grown on plates supplemented with compounds in the media. Plates were setup as followed: LB agar with 1% L-arabinose was mixed with a compound of interest (final concentration ~0.1-2%) and poured onto a plate at an angle. After solidification, the plates were taken off their angle and regular LB agar supplemented with arabinose was poured over the adulterated LB agar. Bacteria were subsequently streaked onto the plates and were placed at 30^{o}C. Differences in growth were noted.

B.4 Results and Discussion

All chemicals and compounds tried and any notable differences in growth patterns are listed in the following tables: Table B.1 Growth on gradient plates, Table B.2 Growth on differential media, Table B.3 Growth in 96-well plates, and Table B.4 Growth on aerosolized plates. In all growth experiments, all cultures grew to the same

extent on the given compound. Compounds tried ranged from ionic detergents, non-ionic detergents, antibiotics, dyes, and differential media.

Table B.1 Growth on Gradient Plates

Detergent/Dye	Concentration	Comments
Sodium Dodecyl Sulfate	1%	No inhibition/no differences
Sodium Dodecyl Sulfate	2%	Limited growth/pBBAD appeared to grow somewhat better
Sodium Dodecyl Sulfate	3%	Limited growth/no differences
Lauroylsarcosine	1%	Limited growth/no differences
Lauroylsarcosine	2%	Limited growth/ pBBAD appeared to grow somewhat poorer
Lauroylsarcosine	3%	Limited growth/ no differences
Alconex	1%	very limited growth/ no differences
Alconex	2%	No growth of any strain
Denzalkonium Chloride	1%	No growth of any strain
Benzalkonium Chloride	2%	No growth of any strain
Deoxycholic acid	2%	No inhibition of growth
Brij 35	2%	No inhibition of growth
Brij 35	3%	No inhibition of growth
Tween 85	2%	No inhibition of growth
Tween 85	3%	No inhibition of growth
Ethylene glycol	2%	No inhibition of growth
Ethylene glycol	3%	No inhibition of growth
Triton 100	1%	No inhibition of growth
Crystal Violet	0.1%	No growth of any strain
Malachite green	1%	No growth of any strain
Malachite green	0.1%	No growth of any strain

Table B.2 Growth on Differential Media

Media	Comments
Eosin Methylene Blue	No differences in growth rate or appearance
Hektoen enteric agar	No differences in growth rate or appearance
Brilliant Green agar	No differences in growth rate or appearance
Simmons citrate agar	No growth of any strain

Table B.3 Growth in 96-well plates

Additive	Concentration ranges	Comments
Streptomycin	1 µM-2 mM	all inhibited between 200 and 250 µM
Tetracycline	1 µM-2 mM	all inhibited between 10 and 20 µM
Sodium Dodecyl Sulfate	10 mM-40 mM	all strains inhibited between 22.5 and 24 mM
Lauroylsarcosine	3 mM-60mM	No growth for any strain
H_2O_2	100nM-10mM	All strains inhibited between 175 and 200µM
2,4-dinitrophenol	10µM-10mM	All strains inhibited between 450 and 475µM
Brij 35	8 mM-80 mM	All strains inhibited between 75 and 80 mM
Benzalkonium chloride	25 µM-1 mM	All strains inhibited between 25 and 100 µM
meta-chloroperoxy benzoic acid	1µM-1mM	All strains inhibited between 50-75µM

Table B.4 Growth on Aerosolized Plates

Additive	Concentrations	Comments
Anthracene	1mg/ml-100mg/ml	No differences in colony phenotype
Azulene	crystals on lid	No differences in colony phenotype
Biphenylene	crystals on lid	No differences in colony phenotype

We hypothesized that the ketones may protect the membrane from ionic detergents. Ionic detergents affect protein-protein interactions (94). They are often used for denaturation of proteins. Sodium dodecyl sulfate (SDS) is an example of an ionic detergent. When initially grown on 2% SDS in the gradient plates, the *S. oneidensis ΔoleABCD* containing the induced pBBAD22K-*S. maltophilia oleA* vector grew better than the control strains. Unfortunately the results were variable. Depending on the experimental replicate, the *S. oneidensis ΔoleABCD* strain containing the pBBAD22K- *S. maltophilia oleA* would grow better, worse, or similarly to the control strains. Irreproducible results were also observed when the bacteria were grown in 96-well plates. Lauroylsarcosine is another ionic detergent tested as a selection reagent. Initial experiments using the gradient plates and 1% lauroylsarcosine found that the *S. oneidensis ΔoleABCD* with the pBBAD22K- *S. maltophilia oleA* vector grew

poorer than the control strains. Subsequent attempts to grow the strains on Lauroylsarcosine ended with mixed results. When grown in 96-well plates, bacterial growth results varied. No differences in how bacteria were grown or how media and compounds were handled between experiments were identified that could explain these mixed results.

We also hypothesized that ketones may protect the *S. oneidensis* membrane against non-ionic detergents. Non-ionic detergents affect lipid-lipid interactions (94). They are often used to help separate membrane fractions in lipid studies. Tween 85 is an example of a non-ionic detergent. While numerous non-ionic detergents were tested as a potential selection reagent (both on a gradient plate and in 96-well plates) no consistent results were observed.

A third hypothesis was the ketone may protect the cell against hydrophobic and/or oxygen radical-forming reagents. Nishida *et al* reported that increases in polyunsaturated fatty acids in *Shewanella* membranes protect against antibiotic infiltration and oxygen radical interactions (112), both of which if not checked would lead to cell death. We hoped that the ketone may provide a similar protection against the cell lysis. While the chemicals and antibiotics tried in these experiments do not support this hypothesis, the ketone may provide some sort of protection to the bacterium. For example, the constructs we have created may not be producing enough ketone to impart protection against these compounds; future strains producing greater amounts of olefin may show greater viability when exposed to these reagents.

Finally we hypothesized that the hydrophobic nature of the ketone would absorb dyes such as anthracene, azulene, and biphenylene. Unfortunately we were not able to

observe this phenomenon with the dyes listed in Table B.1. Potentially the bacteria may not be producing enough ketone for a visual assessment to be applicable; while the polyolefin may be collecting dye, there may be too little ketone in the membrane for an observational difference to be detected. The dye may also not be volatile enough for the compound to reach the growing cells.

While many compounds were screened in an attempt to identify an additive that could be used for ketone selection, none were found to have this desired property. Further screening of compounds may find a substance that can be utilized in a plate or media assay. Future compounds tested could include other detergents, ionophores, antibiotics, and reagents that interact directly with the ketone being produced by the constructs.

Appendix C: Sequences optimized for DNA 2.0 synthesis and primers used for cloning

***Congregibacter litoralis* KT71**

AATATTGGCGCCATGTCTGGTAACGCTAAATTCACTTTAAACGATACTGCTA
TCGTTTCTGTTACTGCTCACCACGCTCCAGAAGTTGTTACTTCTGCTTCTTTA
GATGATCGTATCATGCACACTTACGAACGTTTAGGTACTCAACCAGGTTTAT
TAGAATCTTTAGCTGGTATCTCTGAACGTCGTTGGTGGCCAGAAGGTCACAC
TTTCACTGAAGCTGCTGCTGAAGCTGGTCGTAAAGCTATGGCTGCTGCTAAC
ATCAAACCAGAACAAGTTGGTTTATTAATCGATACTTCTGTTTCTCGTGATC
GTTTAGAACCATCTTCTGCTGTTACTGTTCACCACTTATTAGATTTACCATCT
TCTTGTTTAAACTTCGATATGGCTAACGCTTGTTTAGGTTTCATGAACGCTAT
GCAAGTTGCTGGTATGATGTTAGATTCTCGTCAAATCGATTTCGCTTTAATC
GTTGATGGTGAAGGTTCTCGTCAACCACAAGAAAAAACTTTAGAACGTTTA
GCTTCTGATGAAGCTACTGTTGCTGATTTATTCGCTGATTTCGCTACTTTAAC
TTTAGGTTCTGGTGCTGCTGGTATGGTTTTAGGTCGTCACTCTGAAAACGCT
GGTTCTCACAAAATCATCGGTGGTATCAACCGTGCTAACACTTCTCACCACA
AATTATGTGTTGGTACTTTAGATCAAATGCGTACTGATACTGCTGCTTTATT
AGAAGCTGGTTTAGATGTTTCTGAACGTGCTTGGGCTAACGCTGAAGAATA
CGGTTGGTTAGATATGGATCGTTACGTTATCCACCAAATCTCTTCTGTTCAC
ACTTCTATGTTATGTGAACGTTTAGGTATCGATGTTGATAAAGTTCCATTAA
CTTACCCAAAATTAGGTAACACTGGTCCAGCTGCTGTTCCATTAACTTTAGC

TCAAGAATCTGAATCTTTAAACCCAGGTGATCGTGTTTTATGTTTAGGTATG

GGTTCTGGTATCAACGCTATGGCTTTAGAAATCGCTTGGTAGCTCGAGTATT

AA

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

AATATTGGCGCCATGTTATTCCAAAACGTTTCTATCGCTGGTTTAGCTCACA

TCGATGCTCCACACACTTTAACTTCTAAAGAAATCAACGAACGTTTACAACC

AACTTACGATCGTTTAGGTATCAAAACTGATGTTTTAGGTGATGTTGCTGGT

ATCCACGCTCGTCGTTTATGGGATCAAGATGTTCAAGCTTCTGATGCTGCTA

CTCAAGCTGCTCGTAAAGCTTTAATCGATGCTAACATCGGTATCGAAAAAAT

CGGTTTATTAATCAACACTTCTGTTTCTCGTGATTACTTAGAACCATCTACTG

CTTCTATCGTTTCTGGTAACTTAGGTGTTTCTGATCACTGTATGACTTTCGAT

GTTGCTAACGCTTGTTTAGCTTTCATCAACGGTATGGATATCGCTGCTCGTA

TGTTAGAACGTGGTGAAATCGATTACGCTTTAGTTGTTGATGGTGAAACTGC

TAACTTAGTTTACGAAAAAACTTTAGAACGTATGACTTCTCCAGATGTTACT

GAAGAAGAATTCCGTAACGAATTAGCTGCTTTAACTTTAGGTTGTGGTGCTG

CTGCTATGGTTATGGCTCGTTCTGAATTAGTTCCAGATGCTCCACGTTACAA

AGGTGGTGTTACTCGTTCTGCTACTGAATGGAACAAATTATGTCGTGGTAAC

TTAGATCGTATGGTTACTGATACTCGTTTATTATTAATCGAAGGTATCAAAT

TAGCTCAAAAAACTTTCGTTGCTGCTAAACAAGTTTTAGGTTGGGCTGTTGA

AGAATTAGATCAATTCGTTATCCACCAAGTTTCTCGTCCACACACTGCTGCT

TTCGTTAAATCTTTCGGTATCGATCCAGCTAAAGTTATGACTATCTTCGGTG

AACACGGTAACATCGGTCCAGCTTCTGTTCCAATCGTTTTATCTAAATTAAA

AGAATTAGGTCGTTTAAAAAAAGGTGATCGTATCGCTTTATTAGGTATCGGT

TCTGGTTTAAACTGTTCTATGGCTGAAGTTGTTTGGTAGCTCGAGTATTAA

***Xylella fastidiosa* 9a5c**

AATATTGGCGCCATGTTATTCAACAACGTTTCTATCGCTGGTTTAGCTCACA

TCGATGCTCCATGTACTTTAACTTCTCAAGAAATCAACGCTCGTTTACAACC

AATGTTAGAACGTATCGGTATCAAATCTGATGTTTTCGCTGATATCGTTGGT

ATCAACGCTCGTCGTTTATGGAACACTAACGTTCAAACTTCTGATGTTGCTA

CTATGGCTGCTCGTAAAGCTTTACAAGATGCTGGTGTTGCTGTTGATCGTAT

CGGTTTAGTTGTTAACACTTCTGTTTCTCGTGATTACTTAGAACCATCTACTG

CTTCTATCGTTTCTGGTAACTTAGGTGTTGGTGAACAATGTATCGCTTTCGAT

GTTGCTAACGCTTGTTTAGCTTTCTTAAACGGTATGGATATCGCTGGTCAAA

TGTTAGAACGTGGTGATATCGATTACGCTTTAGTTGTTAACGCTGAAACTGC

TAACCGTGTTTACGAAAAAACTTTAGAACGTATGTCTGCTCCAGGTGTTACT

GAACAAGAATTCCGTGAAGAAATGGCTGCTTTAACTTTAGGTTGTGGTGCTG

TTGCTATGGTTTTAGCTCGTACTGCTTTAGTTCCAGATGCTCCACAATACAA

AGGTGGTGTTACTCGTTCTGCTACTGAATGGAACAAATTATGTTGTGGTAAC

TTAGATCGTATGGTTACTGATACTCGTTTAATGTTAATCGAAGGTATCAAAT

TAGCTAAAAAAACTTTCGTTGTTGCTAAACAAGTTTTAGGTTGGGCTGTTGA

AGAATTAGATCAATTCGTTATCCACCAAGTTTCTCGTCCACACACTGAAGCT

TTCATCAAATCTTTCGGTATCGATCCAGCTAAAGTTATGACTATCTTCCGTG

AATACGGTAACATCGGTCCAGCTTCTGTTCCAATCGTTTTATCTAAATTAAA

AGAATTAGGTCGTTTAAAAAAAGGTGATCGTATCGCTTTATTAGGTATCGGT

TCTGGTTTAAACTGTTCTATGGCTGAAGTTGTTTGGTAGCTCGAGTATTAA

***Plesiocystis pacifica* SIR-1**

AATATTGGCGCCATGCGTTTCGCTAACGTTTCTATCTGTTCTGTTGCTCACGT

TGATGCTCCATACCGTGTTTCTTCTACTGATTTAGAAAACCGTTTAGCTGCTC

CAATGCAACGTTTAGGTTTACCACCAGGTATCTTAGAAACTTTAACTGGTAT

CAAAGCTCGTCGTATGTGGCCAGCTTCTGTTTCTCCATCTGATGCTGCTACTT

TAGCTGCTCGTCGTGCTATCGCTGAATCTGGTGTTGATCCAGAACGTATCGG

TGTTTTAATCTCTACTTCTGTTTGTCGTGATTTCGTTGAACCATCTACTGCTT

GTTTAGTTCACGGTAAATTAGGTTTACCACCAACTTGTTTAAACTTCGATGT

TGGTAACGCTTGTTTAGGTTTCATCAACGGTATGGATATCATCGGTAACATG

ATCGAACGTGGTCAATTAGATTACGGTATCGTTGTTGATGGTGAAGATTCTC

GTTACGTTATCGATAAAACTATCGAACGTTTATCTGCTCCAGATTCTACTCG

TGAAGATTTCTGGTCTAACTTCGCTACTTTAACTTTAGGTGGTACTGCTGCTG

CTATGGTTTTAGCTCGTACTGATTTAGCTCAAGCTTTAGCTGAAAAACGTGC

TGAAGGTGGTTACTCTCACCAATTCTTAGGTTCTGTTATCGTTGCTGCTACTC

AACACTCTGGTTTATGTCGTGGTCAAGTTGATCGTATGGAAACTGATTCTGC

TGAATTATTAACTGCTGGTTTACGTGTTGCTAAAGAAGCTTGGCGTGCTGCT

CAACGTGAATTCGGTTGGACTCCAGGTGCTTTAGATGAATGTGTTATCCACC

AAGTTTCTCGTACTCACACTGATAAATTCTGTGAAACTTTCGAATTAGATCC

AGCTAAATTATTAGCTACTTACCCAGAATTCGGTAACGTTGGTCCAGCTGGT

GTTCCAATGGTTTTATCTAAAGCTGCTTCTTCTGGTCGTTTAGGTCGTGGTGA

TCGTGTTGGTTTAATGGGTATCGGTTCTGGTTTAAACTGTGCTATGGCTGAA

GTTGTTTGGTAGCTCGAGTATTAA

gamma proteobacterium NOR5-3

AATATTGGCGCCATGCACTTCGAATCTGTTGTTATCTTATCTTTAGCTGCTGC

TGATGCTCCAATCTCTTTAACTTCTAAAGAAATCTCTCAACGTTTAAAACCA

ACTATGGATCGTTTAGGTGTTCGTGAAAACTTATTAGAAGAAATCTCTGGTA

TCGCTTCTCGTCGTATCTGGAACCCAGAAACTTCTCCATCTGATGCTGCTAC

TTTAGCTGCTGAAAAAGCTATCCAAGATTCTGGTATCGATCGTTCTCGTATC

GGTGTTATCATCTCTACTTCTGTTTCTCGTGATTTCTTAGAACCATCTGCTGC

TTGTATGGTTCACGGTAACTTAGGTTTAGCTTCTGATTGTTTAAACTTCGATG

TTGCTAACGCTTGTTTAGGTTTCTTAAACGGTATGGATATCGCTGCTCGTAT

GATCGAACGTGAAGAATTAGATTACGCTTTAGTTGTTGCTGGTGAATCTTCT

CGTCCATTAATCGAAGCTACTACTGAACGTTTATTAGATCAAGATGTTGGTG

CTGCTCAATTCCGTGAAGAATTCGCTTCTTTAACTTTAGGTTCTGGTGCTGCT

GCTATGATCATGACTCGTCGTGAATTAGCTCCAGGTGGTCACACTTACCGTG

GTTCTGTTACTCGTTCTGCTACTCAATTCAACCGTTTATGTCAAGGTAACATG

GATCGTATGCGTACTGATACTGGTATGTTATTATCTGCTGGTTTAGAATTAG

CTGCTCAAACTTTCGAAGCTTCTTGTTCTACTTTAGATTGGTCTGTTGATGAA

ATGGATCAATTCATCATCCACCAAGTTTCTAAAGTTCACACTGAATCTTTAG

TTAAAACTTTAGGTTTAAACCCAGATAAAGTTCACGCTATCTACCCACACAT

GGGTAACATCGGTCCAGCTTCTGTTCCAATCGTTTTAGCTAAAGTTGAAGAA

GCTGGTAAATTAAAAAAAGGTGATCGTATCGCTTTATTAGGTATCGGTTCTG

GTTTAAACTGTGCTATGGCTGAAGTTGTTTGGTAGCTCGAGTATTAA

Appendix D: Supplementary data for Chapter 3

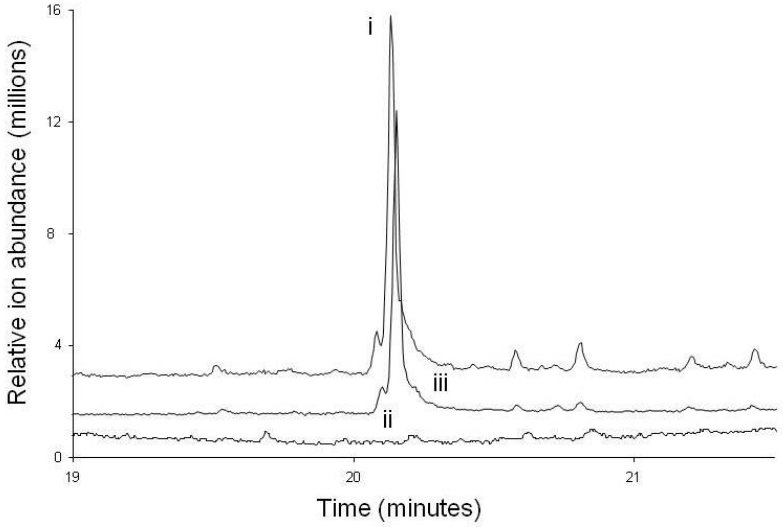


Figure D.1 Gas chromatograph of *S. oneidensis* wildtype (i), the Δ*pfaA* mutant (ii), and the Δ*pfaA* mutant complemented in trans with the *pfaA* gene in the pBBR1MCS2 plasmid (iii).

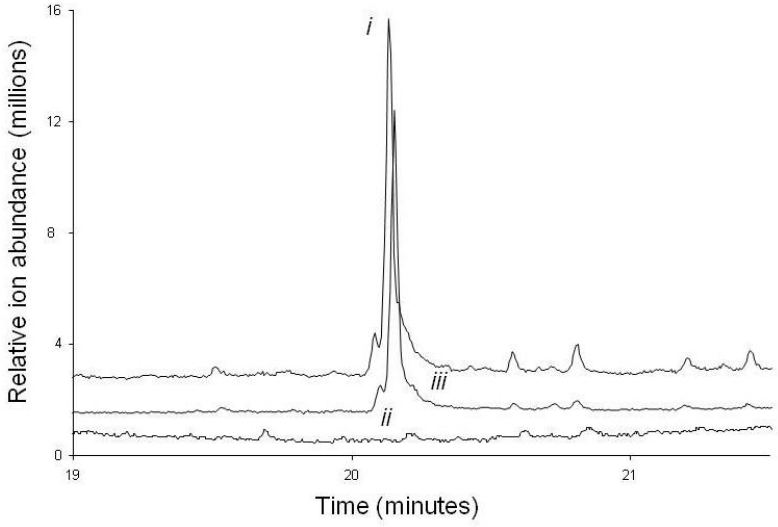


Figure D.2 Gas chromatography of *S. oneidensis* MR-1 strains: (i) wild-type, (ii) mutant lacking *oleABCD*, and (iii) mutant lacking *oleABCD* complemented with pBBR1MCS2 containing *oleABCD*.

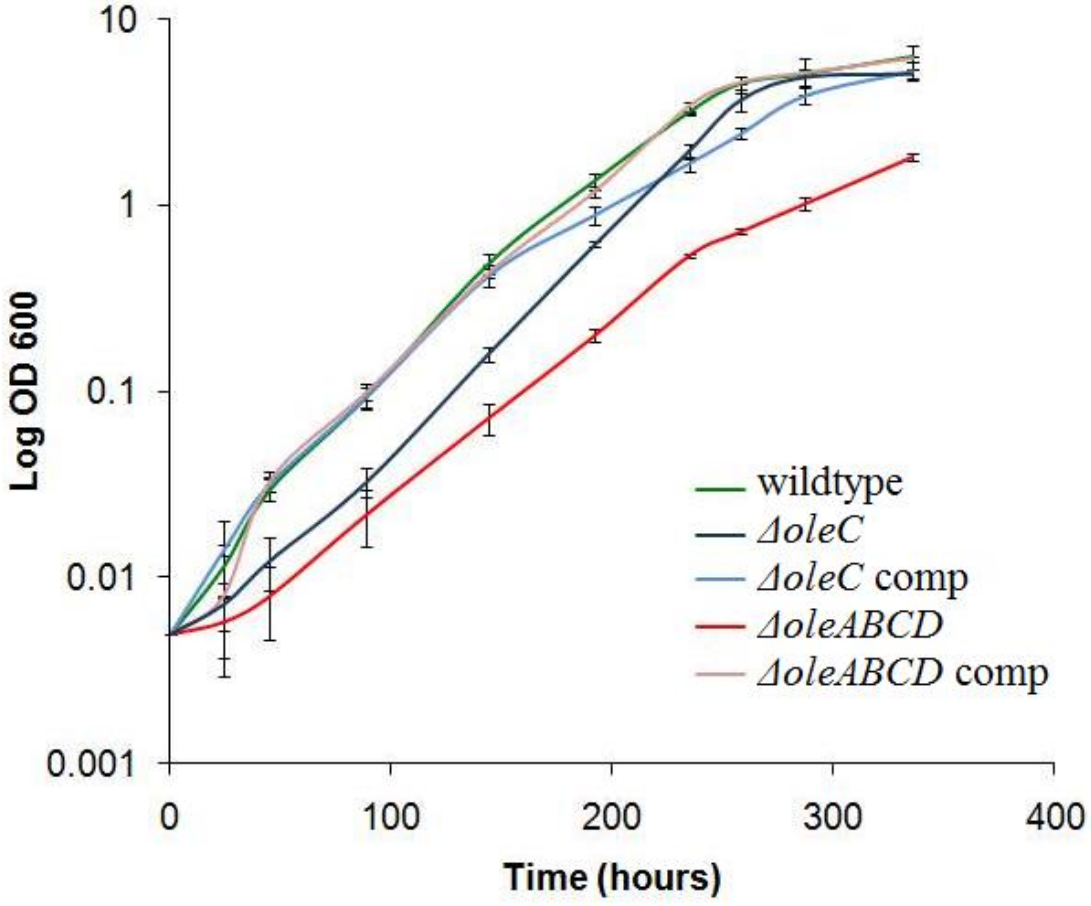


Figure D.3 Wild-type MR-1 (green), the *oleC*-deficient mutant (blue) and complement (light blue), and the *oleABCD*-deficient mutant (red) and complement (light red) were downshifted from 30°C to 4°C, and the cold temperature growth curves are shown. Experimental points are average triplicate samplings from six treatments. Variation is shown as the standard deviation.

Appendix E: Supplementary data for Chapter 4

Table E.1. Organisms with *oleABCD* genes. GI identifiers for each gene are given.

Strain	OleA	OleB	OleC	OleD
Arthrobacter aurescens TC1	119962129	OleBC	119960515	119962242
Arthrobacter chlorophenolicus A6	220911225	OleBC	220911226	220911227
Brachybacterium faecium DSM 4810	62425589	OleBC	237670144	237670143
Brevibacterium linens BL2	237670145	OleBC	62425588	62425587
Chloroflexus aggregans DSM 9485	118047293	118047297	118047295	118047294
Chloroflexus aurantiacus J-10-fl	163849058	163849062	163849060	163849059
Chloroflexus sp. Y-400-fl	187599902	187599906	187599904	187599903
Clavibacter michiganensis subsp. *Sepedonicus*	170782221	OleBC	170782220	170782219
Colwellia psychrerythraea 34H	71279747	71279056	71281286	71280771
Congregibacter litoralis KT71	88700054	OleBC	88705540	88705539
Desulfococcus oleovorans Hxd3	158522019	158522020	158522021	158522022
Desulfotalea psychrophila LSv54	51244593	51246484	51246483	51246482
Desulfuromonas acetoxidans DSM 684	95929232	95929233	95929234	95929235
Gamma proteobacterium NOR5-3	225089533	225089532	225089531	225089530
Geobacter bemidjiensis Bem	197118484	197118483	197118482	197118481
Geobacter lovleyi SZ	189425328	189425329	189425330	189425331
Geobacter sp. FRC-32	110599660	110599661	110599662	110599663
Geobacter sp. M21	191164328	191160706	191160707	191160708
Geobacter uraniireducens Rf4	148264067	148264066	148264065	148264064
Geodermatophilus obscurus DSM 43160	227405470	OleBC	227405471	227405473
Kineococcus radiotolerans SRS30216	152965648	OleBC	152965647	152965646
Kocuria rhizophila DC2201	184200698	OleBC	184200697	184200696
Kytococcus sedentarius DSM 20547	227995171	OleBC	227995172	227995173
Micrococcus luteus NCTC 2665	239917824	OleBC	239917825	239917826
Moritella sp. PE36	149909209	149909208	149909207	149909206
Nakamurella multipartita DSM 44233	229225818	OleBC	229225819	229225820
Opitutaceae bacterium TAV2	225164858	225164859	225155590	**
Opitutus terrae PB90-1	182415091	182415090	182412680	182412679
Pelobacter propionicus DSM 2379	118581504	118581505	118581518	118581519
Photobacterium profundum 3TCK	90413871	90413872	90413873	90413874

Photobacterium profundum SS9	54308655	54308656	54308657	54308658
Planctomyces maris DSM 8797	149174448	149178001	149178707	149178706
Plesiocystis pacifica SIR-1	149918031	149918029	149918030	149918028
Psychromonas ingrahamii 37	119945681	119945682	119945683	119945684
Psychromonas sp. CNPT3	90408674	90408673	90408672	90408671
Shewanella amazonensis SB2B	119774319	119774320	119774321	119774322
Shewanella baltica OS155	126173784	126173785	126173786	126173787
Shewanella baltica OS185	153000075	153000076	153000077	153000078
Shewanella baltica OS195	160874697	160874698	160874699	160874700
Shewanella baltica OS223	217973959	217973958	217973957	217973956
Shewanella benthica KT99	163751382	163751383	163751967	163751968
Shewanella denitrificans OS217	91792727	91792728	91792729	91792730
Shewanella frigidimarina NCIMB 400	114562543	114562544	114562545	114562546
Shewanella halifaxensis HAW-EB4	167624737	167624736	167624735	167624734
Shewanella loihica PV-4	127512642	127512643	127512644	127512645
Shewanella oneidensis MR-1	24373309	24373310	24373311	24373312
Shewanella pealeana ATCC 700345	157962557	157962556	157962555	157962554
Shewanella piezotolerans WP3	212636100	212636099	212636098	212636097
Shewanella putrefaciens 200	124547320	124547321	124547322	124547323
Shewanella putrefaciens CN-32	146292545	146292546	146292547	146292548
Shewanella sediminis HAW-EB3	157374649	157374650	157374651	157374652
Shewanella sp. ANA-3	117921156	117921155	117921154	117921153
Shewanella sp. MR-4	113970883	113970882	113970881	113970880
Shewanella sp. MR-7	114048107	114048106	114048105	114048104
Shewanella sp. W3-18-1	120599457	120599456	120599455	120599454
Shewanella woodyi ATCC 51908	170727499	170727498	170727497	170727496
Stenotrophomonas maltophilia K279a	190572283	190572284	190572286	190572287
Stenotrophomonas maltophilia R551-3	194363945	194363946	194363948	194363949
Stenotrophomonas sp. SKA14	254521309	*	254522078	254520980
Streptomyces ambofaciens DSM40697	117164435	117164436	117164437	117164438
Streptomyces ghanaensis ATCC 14672	239928261	239928262	239928263	239928264
Xanthomonas axonopodis pv. *citri* str. 306	21241007	21241009	77748519	21241012
Xanthomonas campestris pv. *campestris* str. 8004	66766567	66766571	77761077	66766574
Xanthomonas campestris pv. *campestris* str. ATCC	21229690	21229694	77747740	21229697

33913				
Xanthomonas campestris pv. *campestris str*. B100	188989629	188989631	188989633	188989637
Xanthomonas oryzae KACC10331	58583832	58583836	122879327	58583839
Xanthomonas oryzae MAFF 311018	84625635	84625639	84625641	84625642
Xanthomonas oryzae pv. *oryzicola* BLS256	166710197	166710199	166710201	166710202
Xanthomonas oryzae PXO99A	188574836	188574832	188574829	188574828

* There is a large area with no gene identified between *oleA* and *oleC* homologs. When the genome is searched with the *oleB* homolog DNA sequence from *Stenotrophomonas maltophilia K279a*, it hits region with 93% identity. Therefore the nucleotide sequence is there but not predicted.

**Unfinished genome – OleABC are similar to Opitutus terrae PB90-1 (58-71% identity) but the OleD homolog is not adjacent to OleC in the genome, though many OleD homologs are present in the genome

A

ShewMR1	^{94}GAVVYTGVC-	^{113}-TAIYDISN-A**C**LGVLSGI	
	^{271}DKVIC**H**QVG	^{300}TYQLLG**N**MGTVSL	
	^{331}VSFLGIGSGLNCMM		
ArthroTC1	^{90}GLLINTSVT-	^{120}-AMNFDLAN-A**C**LGFVNGL	
	^{264}DRYVT**H**QVS	^{293}TFPHWG**N**VGPASL	
	^{322}VLCMGVGSGLNAGM		
FabH	^{74}GLIVVATTSA	^{103}C-PAFDV-AAA**C**AGFTYAL	
	^{239}DWLVP**H**QAN	^{269}L-DRHG**N**TSAASV	
	^{299}VLLEAFGGGFTWGS		
HMG_CoA	^{72}GMVIVAT**E**SA	^{102}A-RCFEM-KEA**C**YAATPAI	
	^{228}ASLCF**H**VPF	^{270}N-RYVG**N**IYTGSL	
	^{300}IGLFSYGSGSVGEF		
Chalcone	^{125}THLIVCTTSG	^{154}V-KRYMMYQQG**C**FAGGTVL	
	^{299}-FWIA**H**PGG	^{331}L-SEYG**N**MSSACV	
	^{368}GVLFGFGPGLTIET		
Pyrone	^{130}THLIFCTTAG	^{159}V-KRYMLYQQG**X**AAGGTVL	
	^{304}-FWMV**H**PGG	^{336}L-SEYG**N**LISACV	
	^{373}GVLFGFGPGMTVET		
FabB	^{99}VGLIAGSGGG	^{153}HGVNYSI-SSA**C**ATSAHCI	
	^{293}DYLNS**H**GTS	^{328}KAM-TG**H**SLGAAG	
	^{385}VMSNSFGFGGTNAT		
FabF	^{100}IGAAIGSGIG	^{153}RGPSISI-ATA**C**TSGVHNI	
	^{298}GYVNA**H**GTS	^{335}ASM-TG**H**LLGAAG	
	^{393}TLCNSFGFGGTNGS		
Biosynthetic	^{50}NEVILGQVLP	^{80}A-TAWGM-NQL**X**GSGLRAV	
	^{311}DLVEA**N**EAF	^{343}AIA-IG**H**PIGASG	
	^{373}GLATL**C**IGGGMGVA		

B

ShewMR1	^{100}ITLVVH**D**WGG	^{236}ICWGLQ**D**FV	^{267}DCG**H**YILE
ArthroTC1	^{139}VVTVGH**D**WGG	^{271}MLWGPT**D**PI	^{302}GAG**H**LVGE
HAD	^{118}ITLVVQ**D**WGG	^{254}MAIGMK**D**LL	^{286}DAG**H**FVQE
EH	^{101}AYVVGH**D**FAA	^{240}MIWGLG**D**TC	^{272}DCG**H**FLMV
Prolyl	^{548}LTINGG**S**NGG	^{635}LLTADH**D**DR	^{677}KAG**H**GAGK
Lipase	^{147}LHYVGH**S**QGT	^{318}VWNGGK**D**LL	^{350}FYN**H**LDFI
Esterase	^{196}VVVQGE**S**GGG	^{302}VAVNEL**D**PL	^{335}GLV**H**GADV
EH pot	^{99}VFVVAH**D**WGA	^{259}FIVGEF**D**LV	^{297}GAA**H**FVSQ

C

ShewMR1	^{229}TSGSTGTPK	^{373}YGATE	^{475}MGD
ArthroTC1	^{541}TSGSTGPAK	^{684}YGMTE	^{786}TGD
Gram	^{190}TSGTTGNPK	^{323}YGPTE	^{411}TGD
AcCoA	^{320}TSGSTGAPK	^{469}YWQTE	^{557}TGD
Luciferase	^{200}SSGSTGLPK	^{342}YGLTE	^{422}TGD
Benz	^{191}SSGSTGKPK	^{332}IGSTE	^{410}SGD

D

ShewMR1	^{60}GAGGFLG	^{158}LVYTSTP	^{189}YYAHSK	
ArthroTC1	^{7}GASGLLG	^{102}VVYVSSP	^{133}DYARTK	
Udpepim	^{9}GGAGYIG	^{128}VFSSSAT	^{156}PYGKSK	
7a-HOsterDH	^{18}GAGAGIG	^{142}LTITSMA	^{158}SYASSK	
3□HOsterDH	^{8}GCATGIG	^{110}VVISSVA	^{154}AYAGSK	
3□,20□-HOsterDH	^{13}GGARGLG	^{135}VNISSAA	^{151}SYGASK	
FabG	^{14}GSTRGIG	^{140}VNISSVV	^{151}NYSTTK	
D-3-HObut DH	^{11}GSTSGIG	^{138}INIASAH	^{156}AYVAAK	

Figure E.1. Conserved/catalytic residues within OleABCD protein superfamilies – expanded version. Detailed methods for these alignments are described in the method section of the main paper. In summary, structural alignments were done with representative members of each superfamily. A profile-profile sequence alignment using secondary structure was then used to align the structure-based sequence alignments with alignments of Ole proteins. Regions not shown in the sequence alignment are denoted by "…". Significant superfamily conserved residues and catalytic resides are displayed in color and defined below. Accession numbers or pdb identifies of the proteins displayed here are listed below. (A) Thiolase superfamily members used for the alignments include OleA (ShewMR1) from *Shewanella oneidensis* MR1

(gi24373309), OleA (ArthroTC1) from *Arthrobacer aurescens* TC1 (gi119962129), beta-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli* (1EBL), 3-hydroxy-3-methylglutaryl-CoA synthase (HMG_CoA) from *Staphylococcus aureus* (1XPK), chalcone synthase (Chalcone) from *Medicago sativa* (1CGZ), pyrone synthase (Pyrone) from *Gerbera hybrida* (1EE0), beta-ketoacyl-[acyl carrier protein] synthase I (FabB) from *Escherichia coli* (1DD8), beta-ketoacyl-[acyl carrier protein] synthase III (FabF) from *Mycobacterium tuberculosis* (1U6S), and biosynthetic thiolase (Biosynthetic) from *Zoogloea ramigera* (1OU6). Blue and pink residues indicate the glutamate and cysteine that abstracts a proton to produce a carbanion for the non-decarboxylative condensation reaction of HMG CoA synthase and biosynthetic thiolases, respectively; red indicates the absolutely conserved cysteine of the superfamily that covalently attaches to a substrate; and green residues are involved in formation of an oxyanion hole (62, 64, 101, 145). (B) Alpha/beta hydrolase superfamily members included OleB (ShewMR1) from *Shewanella oneidensis* MR1 (gi24373310), OleB domiain (ArthroTC1) from *Arthrobacer aurescens* TC1 (gi119960515), haloalkane dehalogenase (HAD) from *Xanthobacter autotrophicus* GJ10 (1B6G), epoxide hydrolase (EH) from *Agrobacterium radiobacter* AD1 (1EHY), prolyloligopeptidase (Prolyl) from porcine brain (1H2W), gastric lipase (Lipase) from humans (1HLG), brefeldin A esterase (Esterase) from *Bacilius subtilis* (1JKM),

and epoxide hydrolase (EH pot) from potato (2CJP). Red residues indicate the catalytic nucleopile (Ser, Asp, or Cys in the whole superfamily), green indicates the general acid, and blue indicates the conserved histidine that activates water (107, 122, 168). (C) AMP-dependent ligase/synthase superfamily members included OleC (ShewMR1) from *Shewanella oneidensis* MR1 (gi24373311), OleC domain (ArthroTC1) from *Arthrobacer aurescens* TC1 (gi119960515), gramicidin synthetase (Gram) from *Brevibacillus brevis* (1AMU), acetyl-CoA synthetase (AcCoA) from *Saccharomyces cervisiae* (1RY2), luciferase from the Japanese firefly (2D1Q), and benzoate-CoA ligase (Benz) from *Burkholderia xenovorans* LB400 (2V7B). Red indicates absolute conservation in the three consensus regions identified by Conti, et al. ([STG]-[STG]-G-[ST]-[TSE]-[GS]-x-[PALIVM]-K, [YFW]-[GASW]-x-[TSA]-E, [STA]-[GRK]-D). Blue and green indicate Thr/Ser residues thought to be involved in binding the phosphate in ATP and AMP. NOTE: more superfamily members with greater sequence diversity have been identified since the definition of this consensus sequence. Benzoate-CoA ligase, for instance, has an isoleucine in the first position of the second consensus instead of the [YFW] listed but it still is in the superfamily (34, 60, 74, 174). (D) Short-chain dehydrogenase/reductase superfamiy members included OleD (ShewMR1) from *Shewanella oneidensis* MR1 (gi24373312), OleD (ArthroTC1) from *Arthrobacer aurescens* TC1 (gi119962242), UDP-

galactose-4-epimerase (Udpepim) from humans (1EK6), 7-α-hydroxysteroid dehydrogenase (7α-HOsterDH) from *Escherichia coli* (1AHH), 3- α-hydroxysteroid dehydrogenase/carbonyl reductase (3 α -HOsterDH) from *Comamonas testosteroni* (1FK8), 3- α, 20-β hydroxysteroid dehydrogenase (3 α,20β-HOsterDH) from *Streptomyces exfoliatus* (2HSD), 3-oxoacyl ACP reductase (FabG) from *Aquifex aeolicus* (2PNF), and D-3-hydroxybutyrate dehydrogenase (D-3-HObut DH) from *Pseudomonas fragi* (3ZTL). Blue identifies the tyrosine anion that abstracts the proton from the substrate, red is a lysine that stabilizes the tyrosine anion, green is a glycine rich region involved in cofactor $NAD(P)^+$ binding, and pink is the serine that orients the substrate or stabilizes intermediates (47, 48, 79).

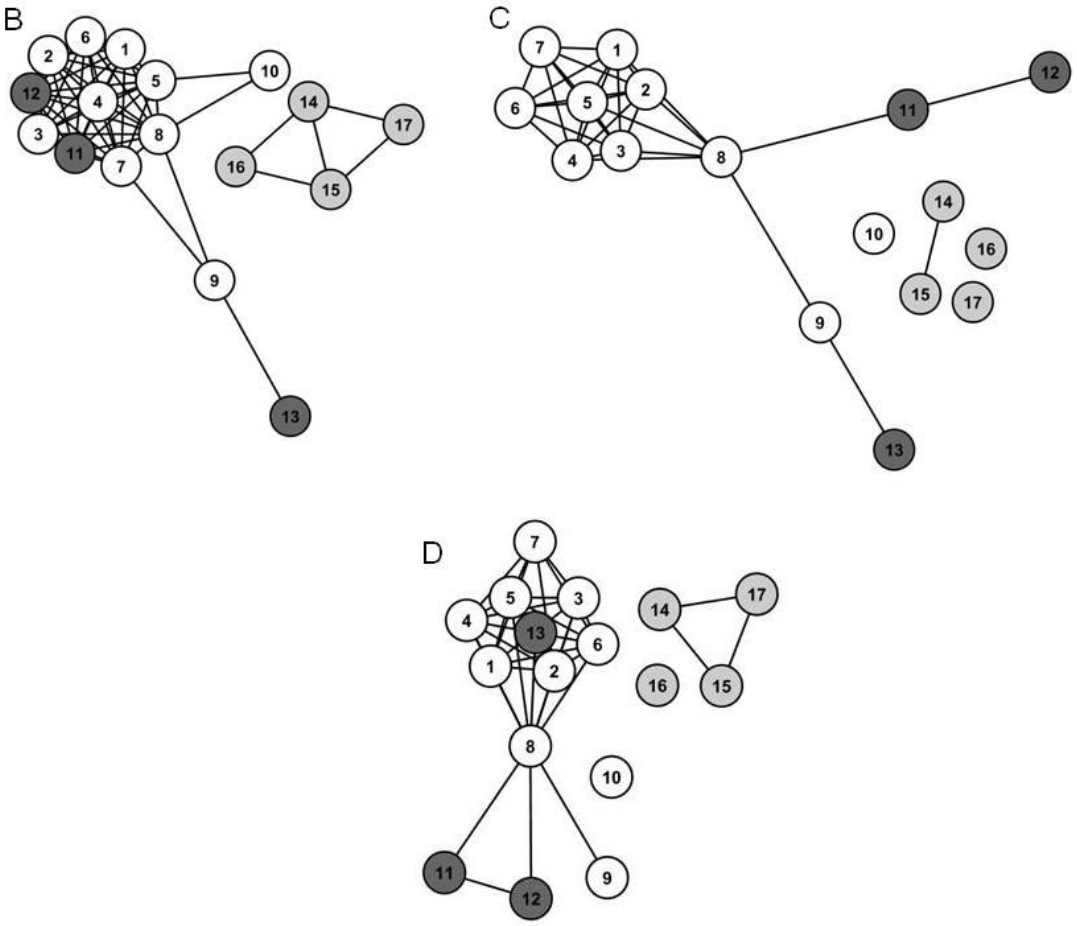


Figure E.2 Network diagrams of OleBCD at the specified cutoffs, showing how the proteins from various organisms brach in a pattern that differs from that of OleA (shown in the paper). Except for OleA, the divergence in protein sequences does not correlate to the products being formed. The organisms that the node numbers represent are the same in the main text. (B) OleB with a 1e-69 cutoff, (C) OleC with a 1e-145 cutoff, and (D) OleD with a 1e-89 cutoff.

Appendix F: Supplementary data for Chapter 5

Table F.1 Strains, vectors and primers used for cloning in Chapter 5

Strains:	Notes	Reference
Chloroflexus aurantiacus sp. J-10-fl	wildtype	120
Shewanella oneidensis Δole	oleABCD knockout	150
Ralstonia eutropha	strain unable to produce PHB	72
Plasmids:		
pBBR1MCS2	5.1KB broad-host range plasmid, *lacZ*, Km^r	89
pChloro	pBBR1MCS2 containing 1.1KB fragment of oleA	Chapter 3
pBBR	vector obtained from Scrienc Laboratory (University of Minnesota)	
pBBRChloro	pBBR containing 1.1 KB fragment of *C. aurantiacus oleA*	This study
pBBRS.m.	pBBR containing 1.1 KB fragment of *S. maltophilia oleA*	This study
pBBRXanth	pBBR containing 1.1 KB fragment of *X. campestris* wild type *oleA*	This study
pBBRCongorSyn	pBBR containing 1.1 KB fragment of Synthesized *C. litoralis oleA*	This study
pBBRXanSyn	pBBR containing 1.1 KB fragment of synthesized *X. campestris oleA*	This study
pBBRXylSyn	pBBR containing 1.1 KB fragment of synthesized *Xylella fastidiosa oleA*	This study
pBBRP.pacificaSyn	pBBR containing 1.1 KB fragment of synthesized *P. pacifica oleA*	This study
pBBRg.p.Syn	pBBR containing 1.1 KB fragment of *gamma proteobacteria* Nor-5 *oleA*	This study

Primers:	
S.m.CompFClaI	ATCTATCGATAACCTCGATGCTCTTCAAGAATGTCTC
S.m.CompRXhoI	CGATCTCGAGGAAGATCATCGCTGTCCGTCGCGAGC
ChlorooleAClaF	CATATTATCGATATGCTATTCAGGCATGTCATGATCG
ChlorooleAXhoR	CAATATCTCGAGTCACCACGTCACACTCATCATTGAAC
XantholeANarIF	ATTAATGGCGCCATGCTCTTCCAGAATGTCTCCATCGC
XantholeAApaIR	AATATTGGGCCCTCACCAAACCACTTCGGCCATCGA
CongoleANarIF	AATATTGGCGCCATGTCTGGTAACGCTAAATTCACT
CongoleAXhoIR	TTAATACTCGAGCTACCAAGCGATTTCTAAAGCCAT
XanSynoleANarIF	AATATTGGCGCCATGTTATTCCAAAACGTTTCTATC
Xan/XylSynoleAXhoIR	TTAATACTCGAGCTACCAAACAACTTCAGCCATAGAAC
XyloleANarIF	AATATTGGCGCCATGTTATTCAACAACGTTTCTATC
PlesioleANarIF	AATATTGGCGCCATGCGTTTCGCTAACGTTTCTATC
PlesioleAXhoIR	TTAATACTCGAGCTACCAAACAACTTCAGCCATAGCAC
gammaoleANarIF	AATATTGGCGCCATGCACTTCGAATCTGTTGTTATC
gammaoleAXhoIR	TTAATACTCGAGCTACCAAACAACTTCAGCCATAGCA

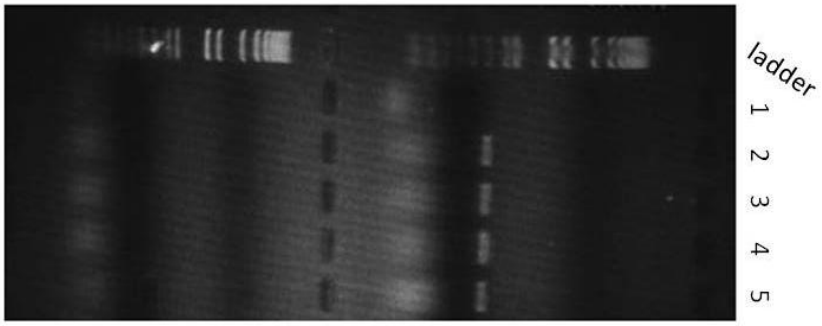


Figure F.1. RT PCR reaction of *R. eutropha* containing the pBBR-vector with different *oleA*s. Lane (1) *Ralstonia* "wildtype", (2) *Ralstonia* with pBBR-*S. maltophilia oleA*, (3) *Ralstonia* with pBBR-*C. aurantiacus oleA*, (4) *Ralstonia* with *C. litoralis oleA*, and (5) *Ralstonia* with *P. pacifica oleA*. Far left is gel of mRNA control while right gel is of the RT- PCR reaction using primers RTPCRF1 and RTPCRR1.