Metabolic Regulation and Genetic Tools for Bacterial Neutral Lipid Production

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

Nagendra Prasad Palani

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Dr. Brett M. Barney

September 2011

Acknowledgements

I express my most sincere thanks to Dr. Brett Barney. He has been an exceptional teacher, advisor and mentor for me. He has been unfailingly supportive during my time here and I cannot thank him enough for all he has taught me. I also thank the Department of Bioproducts & Biosystems Engineering, University of Minnesota, the National Science Foundation and the Department of Energy for their support of research in Dr. Barney's lab.

I thank my committee members Dr. Igor Libourel and Dr. Simo Sarkanen for posing critical questions about my research, helping me think better in the process. Their wisdom and insight often set me into thinking about my research from new perspectives.

I thank fellow graduate and undergraduate students of the Barney lab and members of the Libourel lab for the many favors I have asked of them.

I thank my friends here in the US for their fantastic company. I cannot express enough gratitude to my friends back home, who stepped in as sons and brothers whenever my family needed me.

I thank my parents and my sister for their unwavering love, support and confidence in me.

For her incredible patience and steadfast support, many thanks to my best friend and soon-to-be wife, Ajeetha.

Abstract

Wax esters are neutral lipids that find diverse applications as cosmetics components, lubricants and fuels. Most of the present production of wax esters is derived from non-renewable fossil oil. Given the diversity of uses, essentiality and their burgeoning demand, it is prudent to develop renewable alternatives to current methods of production. The recent discovery of wax biosynthesis in a niche of bacteria has created much interest in exploring these organisms both for their biochemical and biotechnological significance.

In the first section of this thesis, I describe efforts to understand the metabolic regulation involved in bacterial wax biosynthesis. Wax biosynthesis is induced in bacteria like *Marinobacter aquaeolei* VT8 and *Psychrobacter cryohalolentis* K5 when the organisms are starved of nitrogen nutrients. It is of interest to know how bacterial metabolism redirects its carbon flow to wax accumulation in the absence of fixed nitrogen. We reasoned that there would be elevated expression of proteins that formed the metabolic route leading to wax formation. Therefore, it would be possible to develop a hypothesis about the metabolic pathway that leads to wax esters if differences in protein expression could be found out between cells before and during wax biosynthesis. Proteomic experiments were performed to compare the cellular protein composition before and after wax accumulation, resulting in a catalogue of proteins differentially expressed between the two metabolic states. These proteins represented the metabolic changes during wax biosynthesis. To fill in key gaps in the proteomics derived information and aid in further development of the hypothesis, quantitative methods of analyzing gene expression at the transcriptional level were adopted. The current state of progress from these studies is discussed.

In the second section of this thesis, I explain the development of genetic tools that enable recombinant gene expression in select wax ester producing bacteria. A self-transmissible plasmid vector and a conjugal method of DNA transfer were adopted to deliver non-native DNA into the target bacterial strain *Psychrobacter arcticus* 273-4 from a donor *Escherichia coli* strain. In plasmid-bearing colonies of *P. arcticus* 273-4, a

puzzling phenomenon was observed where the circular plasmid delivered into the bacteria had increased in size, potentially capable of undermining recombinant gene expression efforts. However, the cause was identified to be a transposon insertion event originating from the donor *E. coli* strain which harbored the transposon, a parasitic genetic element. A 38 base pair sequence, previously reported to be effective against transposon insertions, was incorporated into the plasmid vector. Testing the efficacy of the solution indicated that presence of the introduced sequence was sufficient to prevent transposon insertion events.

In the final section of this thesis, I describe the construction of plasmid vectors that would enable the biosynthesis of myrcene, a monoterpene hydrocarbon, in recombinant *E. coli*. This system is intended to be a prototype for monoterpene biosynthesis in the cyanobacterium, *Anabaena variabilis*. One of the design goals was to assemble two native and one mutated *E. coli* genes in a single plasmid vector to enable biosynthesis of geranyl pyrophosphate (GPP), the precursor for myrcene. The three genes were assembled into an operon, a gene organization principle observed in bacteria where the expression of contiguous genes is driven by a single promoter. During an intermediate step, an issue perceived to be poor protein expression from one of the genes was resolved. The plasmid vector pPCRISPDSX11 Col 1 contained the final construct of the synthetic operon. Electrophoretic analysis provided visual evidence that all three genes in the vector were being expressed. This would imply that the precursor GPP is being produced in *E. coli*, thereby making significant progress towards the stated aim.

Table of Contents

List of Tables

List of Figures

Chapter I

Metabolic Regulation of Wax Ester Biosynthesis in

Marinobacter aquaeolei VT8 & *Psychrobacter cryohalolentis* K5

1.1 Introduction

1.1.1 Wax Esters: Introduction

Wax esters are carbon rich macromolecules, a class of lipids formed by the condensation of a long-chain fatty acid with a long-chain fatty alcohol (Jetter & Kunst, 2008). They are compact and energy dense neutral lipids, soluble only in non-polar solvents. The length of the fatty precursors, their degree of branching and degree of unsaturation determine the physical properties of wax esters. Generally, wax esters of smaller lengths or with higher degrees of unsaturation or branching are liquids at room temperature (Patel et al., 2001). The neutral chemical nature and the oily character of wax esters make them suited towards varied applications. Wax esters find a range of uses, spanning from candles to cosmetics, lubricants and fuels. Insect derived waxes, like beeswax, are used for manufacturing polishes and even as food additives. Animal derived waxes, typified by the spermaceti wax of the sperm whale, were extensively used in the manufacture of cosmetics and lubricants (Alvarez, 2010). Wax esters of plant origin are often found as a mixture with fatty alcohols and hydrocarbons and some wax esters like jojoba oil have been in high demand from the cosmetics industry for use as skin moisturizers (Jetter & Kunst, 2008). While petroleum derived waxes have been grouped with wax esters, they are typically long-chain hydrocarbons (paraffins) used as solid lubricants. Apart from the most common sources of wax esters, a multitude of organisms ranging from protists to humans have been found to produce wax esters (Brasser et al., 2011; Teerawanichpan & Qiu, 2010).

In chemical notation, wax esters are represented as CX-W:CY-Z where C represents carbon, X represents the alcohol moiety, Y represents the fatty acyl moiety, W & Z represent the degree of unsaturation. Thus C16-0:C16-0 would represent the wax palmityl palmitate, which is completely saturated. A variety of wax esters have been characterized from bacterial sources and also synthesized using the enzymatic machinery of wax ester biosynthesizing bacteria, with variation in size, degree of unsaturation, branching and incorporation of uncommon chemical groups (Rontani et al., 1999; Uthoff et al., 2005).

With growing demand for natural waxes from industries like cosmetics manufacturers, there is a strain on the natural resources that supply these products (Hills, 2003). For example, jojoba oil, a recommended alternative to spermaceti wax ester, is in short supply due to the requirement of a specific arid ecology for the jojoba plant (*Simmondsia chinensis*) to thrive. Other natural sources face similar limitations from being developed into stable sources for wax esters. With the expansion of applications of wax esters to polishes, coatings and recently oleochemicals like fuels, the demand is expected to burgeon over the next few years. The potential for using renewable feedstock for wax ester production makes biological synthesis more attractive than methods of synthetic chemistry. Further, the recent demand for naturally derived additives in cosmetics and food, even at premium pricing, encourages the biosynthesis of wax ester chemicals. Such market considerations warrant the development of a biological system than can supply wax esters sustainably, from renewable sources and which is easy to scale up according to demand.

The recent discovery of monoester waxes in some bacterial species has spurred great biotechnological interest in understanding the biochemical basis of wax synthesis in these micro-organisms (Ishige, 2003). Waxes known as dimycocerosates were known to be produced by mycobacterial species and were hypothesized to be necessary for the infectivity of these organisms. However, the enzymatic pathway for mycocerosate synthesis was not known (Minnikin et al., 1985). Fatty acid based wax esters were initially identified in hydrocarbon-degrading bacteria, isolated from oceanic oil spills and natural hydrocarbon seepage from seafloors (Ishige et al., 2000; Ishige et al., 2002). These hydrocarbonoclastic bacteria were found to accumulate long chain fatty acid based wax esters when grown on hydrocarbons like hexadecane (Ishige et al., 2002). Some of the earliest experiments to probe the biochemical basis of wax production were performed in one such hydrocarbonoclastic species of the *Acinetobacter* genus (Tani et al., 2000; Tani et al., 2002). With the identification of a key enzyme in bacterial wax biosynthesis and subsequent protein homology analysis, a number of bacteria have now been identified including *Marinobacter aquaeolei* VT8 and *Psychrobacter cryohalolentis* K5, that are capable of intracellular production of wax esters.

1.1.2 Wax Ester Synthase

One of the major breakthroughs in bacterial wax ester biosynthesis research was the identification and characterization of the terminal enzyme in the wax biosynthetic pathway, the wax ester synthase/diacyl glycerol acyltransferase (WS/DGAT) (Kalscheuer & Steinbüchel, 2003). This enzyme was found to catalyze the formation of an ester

linkage between a fatty acyl-CoA and a fatty alcohol. It can also link two fatty acid molecules to one of glycerol, leading to the formation of diacylglycerol. The discovery of this enzyme filled a major gap in the knowledge about a hypothetical metabolic pathway for wax synthesis in bacteria. The enzyme and most of its characterized homologs were found to be promiscuous to a broad range of substrates, including unsaturated and branched chain fatty acids and alcohols, and smaller alcohols like ethanol and butanol (Stöveken et al., 2005; Uthoff et al., 2005). The promiscuity of wax ester synthase led to the development of a bioprocess for the production of fatty acid ethyl ester (FAEE) which is similar to the conventional biodiesel FAME (fatty acid methyl ester) derived from the plant triglycerides (Kalscheuer et al., 2006). This enzyme has also been genetically engineered into a fatty acid overproducing strain of *Escherichia coli*, resulting in recombinant production of long-chain linear wax esters (Steen et al., 2010). Recent efforts in our laboratory have been focused on generating rationally designed mutants of the WS/DGAT protein to explore the structural factors that affect the substrate specificity of the enzyme. Information obtained from this study is intended to identify mutant enzymes that exhibit properties desirable for a bioprocess, like high substrate specificity and enzyme stability.

1.1.3 Biochemistry of Wax Ester Biosynthesis

Biochemical studies have outlined two different pathways for bacterial wax biosynthesis. In hydrocarbon degrading bacteria, alkanes are terminally oxidized to form fatty alcohols and then fatty acids. These are then condensed by the wax ester synthase

enzyme. In some bacteria grown on simple carbon substrates like citrate or pyruvate, precursors for wax esters are formed from fatty acids biosynthesized *de novo*. The precursor for this wax synthesis pathway is fatty acyl-CoA which is a fatty acid linked to a Coenzyme A molecule. The first reaction in the pathway is mediated by a fatty acyl CoA reductase enzyme that converts long-chain fatty acyl CoA to fatty aldehydes (Reiser & Somerville, 1997). The aldehydes are then acted upon by a fatty aldehyde reductase to form fatty alcohols (Wahlen et al., 2009). The terminal WS/DGAT enzyme links together a molecule of fatty alcohol and a molecule of fatty acyl-CoA through an ester bond to form wax ester molecule, with the release of free CoA (Kalscheuer & Steinbüchel, 2003).

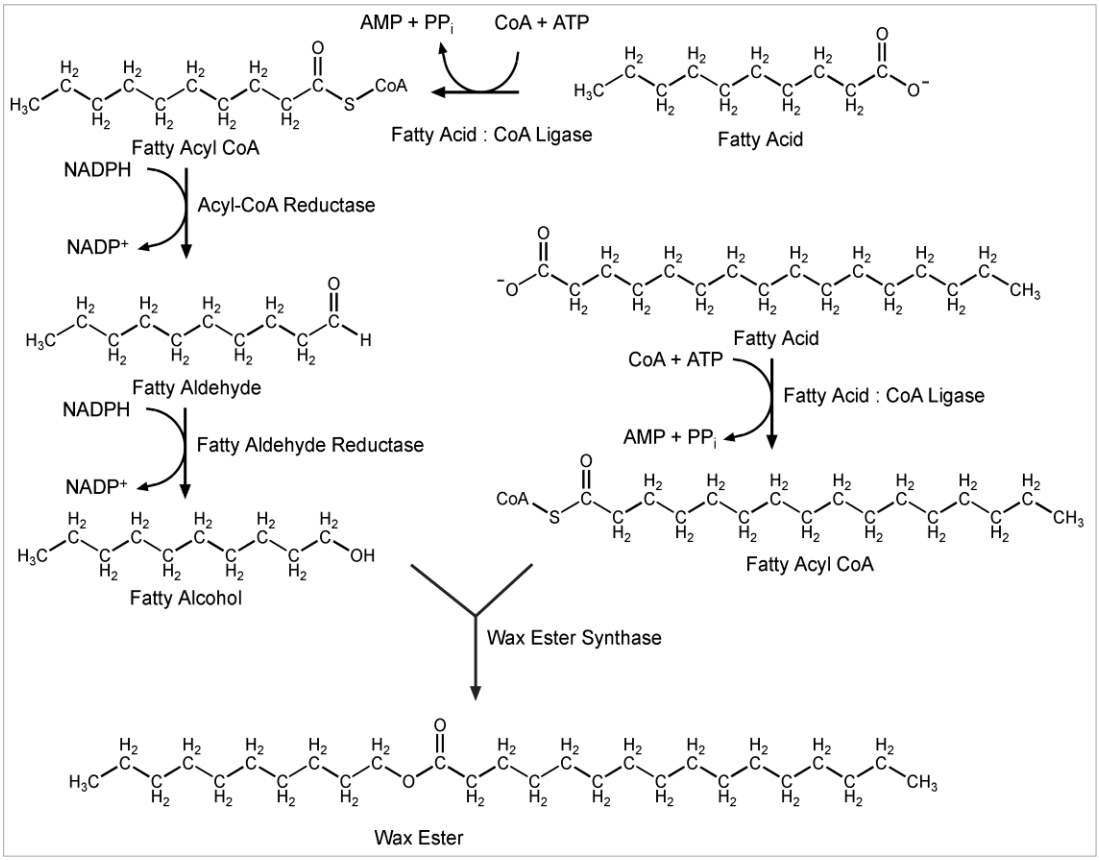


Figure 1.1: Hypothetical pathway for wax ester biosynthesis from fatty acids biosynthesized *de novo*. The terminal enzyme, wax ester synthase, has been used to identify bacteria that can produce wax esters. Image courtesy: Dr. Brett Barney

1.1.4 Conditions for Wax Biosynthesis in Bacteria

While many bacteria have been identified so far that are capable of wax biosynthesis, the select few of them studied under laboratory conditions exhibit wax ester formation and subsequent accumulation only under specific environmental conditions (Wältermann & Steinbüchel, 2005). So far, exhaustion of a soluble nitrogen source in the growth medium in the abundant presence of an organic carbon source or an increase in hydrostatic pressure have been found to be stimuli for induction of wax synthesis(Grossi et al., 2010; Rontani et al., 1999). Since both these conditions represent stressed states for a bacterium, there might other stress inducing factors that could trigger wax ester formation. Most studies regarding the physiology of wax ester producing bacteria have used the former stress response, the depletion of a nitrogen containing nutrient, as the trigger for inducing wax accumulation.

Considering wax formation as a response to nutrient stress, it has been hypothesized that waxes could serve as carbon and energy storage compounds (Kalscheuer & Steinbüchel, 2003). The fact that some bacteria are capable of storing waxes greater than 15% of their dry cell weight(Ishige et al., 2002), and when a nitrogen nutrient is introduced into the medium exhausted of it, the intracellular waxes are completely utilized supports this hypothesis. The high energy density and charge neutrality of wax esters makes them ideal compounds for storage. Furthermore, hydrolysis of a wax ester yields a fatty alcohol and fatty acid which can be fed into the bacterial cell's primary metabolism through the fatty acid degradation pathway. Thus wax esters can be considered functionally analogous to common bacterial storage polymers

like polyhydroxyalkanoates (PHA), glycogen and triacylglycerol (Alvarez & Steinbüchel, 2002; Wilson et al., 2010; Zinn et al., 2001). PHAs are the most common bacterial polymers that serve as storage compounds, and bacteria have been found to commence PHA synthesis under a variety of environmental conditions (Anderson & Dawes, 1990). Similarly, glycogen is a polymer of glucose molecules used as a carbon storage compound in some bacteria, notably *E. coli*. TAGs have been recognized as storage compounds in certain bacterial species, leading to the consideration of such species as a source of substrates for biodiesel (FAME) production (Hernández et al., 2008). One advantage of bacterial TAG based manufacture of FAME is the presence of existing infrastructure that converts TAG from oilseeds into FAME biodiesel. It is of interest that some bacterial species are capable of biosynthesizing both wax esters and TAG as storage molecules (Manilla-Pérez et al., 2010).

To further understand the evolution of wax ester as a storage molecule in a niche group of bacteria, it is helpful to realize that many species that are wax producers have been isolated from marine environments and they are also capable of degrading hydrocarbons (Kalscheuer, 2010). While hydrocarbons provide an abundant source of carbon, biomass growth in the ocean may be limited by availability of a soluble nitrogen source. In such conditions of nitrogen starvation, it would be advantageous for an organism to devise a way of storing the abundant carbon available and use it for rapid growth when a source of nitrogen becomes available again. Thus hydrocarbon metabolism would not only yield energy for the growth of the organism but also provide precursors for storage compounds (Rontani, 2010). However, recent evidence has shown

that some hydrocarbonoclastic bacteria are also capable of biosynthesis of wax esters from simpler carbon sources than hydrocarbons (Alvarez, 2010; Kalscheuer, 2010). This finding is considered to be of biotechnological importance, since the natively wax producing bacteria can be developed into industrial strains that can produce wax esters from simple, non-hydrocarbon substrates. However, to develop such strains, it would be necessary to rationally re-engineer the biochemical networks of these bacteria. To implement such rational design, it is then necessary to understand the systematic changes in cellular physiology during induction of wax esters biosynthesis.

1.1.5 Gene Expression Studies in Wax Ester Synthesizing Bacteria

Biosynthesis of wax esters from simple carbon sources requires *de novo* synthesized fatty acids as precursors. Fatty acids (FA) are produced by the fatty acid anabolic pathway, beginning from acetyl-CoA, a metabolite of the Tricarboxylic Acid Cycle. In the model bacterium *E. coli*, the biosynthesis of fatty acids is tightly regulated under studied physiological conditions through negative feedback mechanisms. Since fatty acids are the primary constituents of the cell membrane, their biosynthesis is tightly linked to exponential phase of cell growth (Fujita et al., 2007). Extensive genetic engineering is required to relax such feedback inhibition and thereby overproduce fatty acids in *E. coli* (Handke et al., 2011). In contrast to such precedent, we hypothesize that wax biosynthesizing bacteria can de-link fatty acid biosynthesis from cell growth. With the exhaustion of nitrogen nutrient, such bacteria would be unable to proceed with cell growth at the maximum possible rate. However, they are able to synthesize significant

amounts of fatty acid based wax esters during this period of nutrient starvation. Such a phenomenon would require that fatty acid synthesis proceed even when cell growth is inhibited. The study of such a natural example of de-regulation of FA biosynthesis could prove useful in the development of fatty acid based renewable fuel development.

In order to unravel the biochemical mechanism(s) that would lead to the de-regulation of FA biosynthesis and to the subsequent synthesis and accumulation of wax esters, it is necessary to obtain a detailed picture of a bacterial cell's transition from a nutrient rich state to a nutrient starved state. A starvation or stress is expected to trigger global changes in the gene expression pattern of a bacterium, as it must modify its physiology to suit the new condition (Hengge-Aronis, 1996). In other studies with prokaryotic systems, stress response has involved changes in the expression of multiple transcription factors, which in turn can regulate the expression of multiple genes. These changes result in enzyme and metabolite concentrations different from before the triggering of the stress response (Jozefczuk et al., 2010). Since the initiation of wax ester synthesis is a part of the stress response towards nitrogen starvation, it then becomes obvious that any study aimed at understanding the nature of *in vivo* wax ester metabolism would benefit from experiments that provide holistic information about a cell. We followed this line of reasoning to adopt methods that were most suited to answering questions like

1. What metabolic changes occur that route intracellular carbon flux towards *de novo* wax biosynthesis?

2. How are such changes effected by gene regulation mechanisms?

Finding answers to these questions would help us understand the mechanism of *in vivo* wax biosynthesis and thus guide us in rational genetic engineering efforts.

1.1.6 Methods for Global Analysis of Gene Expression

To understand global variations in gene expression during wax biosynthesis, we resorted to experimental techniques that were capable of capturing changes in bacterial cellular physiology at the onset of nitrogen starvation and subsequent wax ester production. These techniques were used to probe the proteome and the transcriptome, the protein and mRNA constituents respectively, of bacterial cells capable of wax production.

Analytical methods of exploring the proteome utilize chromatographic and mass-spectrometry instrumentation to identify and if necessary, quantify enzymes and other proteins present in the cell at a particular time (Domon & Aebersold, 2010). A powerful adaptation of this idea is a technique known as isobaric Tagging for Relative and Absolute Quantitation (iTRAQ) (Zieske, 2006). iTRAQ uses isobaric labels, chemical moieties of identical structure that differ only in their isotopic composition, to label protein fragments. The labeled fragments are subsequently identified and quantified using tandem mass-spectrometry. When applied to cell samples exhibiting different physiological conditions, this powerful technique can quantify the ratio of the differentially expressed proteins between the samples, thus providing a global coverage of proteome level differences (Redding et al., 2006).

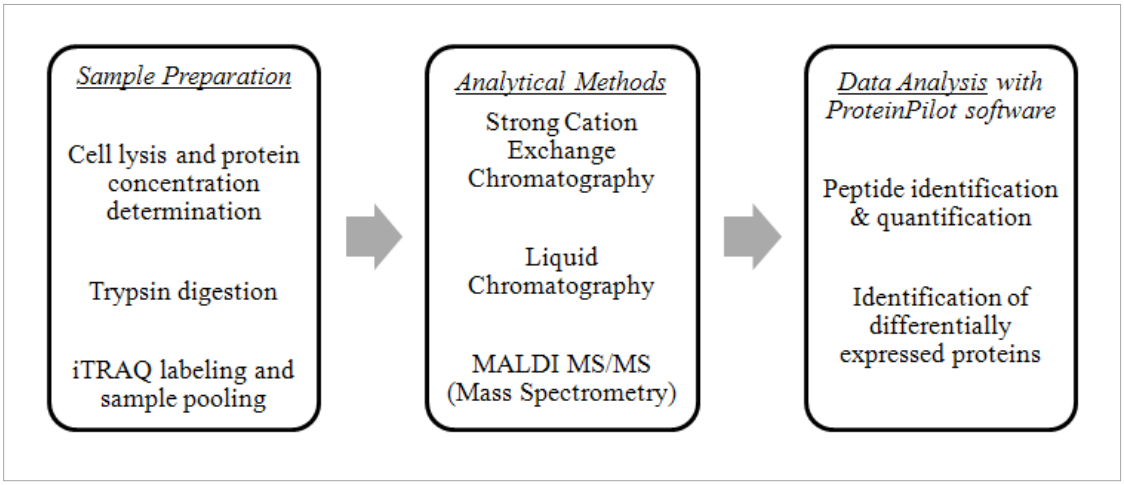


Figure 1.2: Flowchart describing process to identify differentially expressed proteins using iTRAQ labeling.

Transcriptomics is the study of gene expression changes at the level of messenger RNA (mRNA), or the transcriptome. mRNA expression is more transient compared to protein expression and therefore, provides a more precise picture of a cell's physiological state. Mature transcriptomic techniques like microarray hybridization and real time quantitative polymerase chain reaction are routinely used to catalog system wide fluctuations in gene expression.

The Quantitative Real Time – Polymerase Chain Reaction technique was adopted for probing variation in gene expression of select genes during wax ester biosynthesis. The technique uses a modified PCR process to quantify mRNA concentrations present in a sample, thereby possessing the ability to differentiate mRNA levels of a gene across samples. The technique utilizes a reverse transcriptase enzyme to create complementary DNA (cDNA) copies of RNA in a 1:1 ratio. These cDNA are then amplified in a PCR process using primer pairs specific to a single gene. Most often, a fluorescent dye that

specifically binds to double-stranded DNA is used to quantify the amount of DNA present after each cycle of amplification, thereby indicating the abundance of mRNA of that gene present initially in the sample. Comparison of initial mRNA abundances across samples for each gene leads to identification of differential expression of a gene across samples.

The integration of data from multiple 'omics' studies can provide comprehensive systems-level information about a cell. Such knowledge has been used to develop bioprocesses within short timescales. One notable example is the bacterium *Mannheimia succiniciproducens*, a succinic acid secreting anaerobic bacterium discovered in the gut of a Korean cow. The use of Systems Biology approaches led to the design of a bioprocess based on this bacterium for succinic acid production in approximately 7 years (Lee et al., 2002; Hong et al., 2004; Lee et al., 2006; Lee et al., 2008; Oh et al., 2009). Similar strategies in exploring the biosynthetic machinery of wax ester producing bacteria could lead to fruition of a bioproducts platform for renewable production of wax ester based fuel molecules.

1.2 Materials and Methods

1.2.1 Bacterial Strains and Growth Conditions

M. aquaeolei VT8 and *P. cryohalolentis* K5 cell pellets were obtained from Robert Willis at Utah State University, in Logan, Utah. Each species was grown in two flasks that contained 1L of the medium and differed only in the amount of nitrogen source present (3mM versus 14 mM). After 3 days of growth in a shaker table at 30° C, the samples were centrifuged and cell pellets were transferred to 2 mL microcentrifuge tubes, before freezing. Each cell pellet within a 2 mL microcentrifuge tube contained approximately 50 mg of wet cells.

1.2.2 2-Dimensional Gel Electrophoresis

The cell pellets were removed from freezer storage and placed on ice. 100 µL of 2D-gel electrophoresis sample preparation buffer (prepared by Todd Markowski, UMN Center for Mass Spectrometry & Proteomics) was added to each of the samples to resuspend the cells. The cells were lysed by sonication (Misonix XL-2000, Qsonica LLC, Newtown, CT) at power setting 15 in three 10 second bursts with 10 second breaks in between to prevent thermal denaturation of cell components. The microcentrifuge tubes were kept on ice throughout the process. After the sonication step, the tubes were centrifuged at 20,000 g for 4 minutes in a benchtop centrifuge (Centrifuge 5424 - Eppendorf AG, Hauppauge, NY). Immediately after centrifugation, the tubes were placed on ice and supernatant from each tube transferred to a new 0.5 mL microcentrifuge tube,

taking care not to disturb the pellet. Both the pellet and supernatant were then stored at -20° C for later use.

In order to quantify the amount of total protein present in the supernatants, a standard Bradford assay (Thermo Fisher Scientific, Waltham, MA) was used according to the manufacturer's protocol. From the calculated protein concentration, 100 μg of protein for each sample was submitted to the Center for Mass Spectrometry & Proteomics (University of Minnesota) for 2D gel electrophoresis analysis.

1.2.3 iTRAQ Workflow Description

1.2.3.1 Sample Preparation

The cell pellets were removed from freezer storage and placed on ice. 100 μL of iTRAQ dissolution buffer (prepared by Todd Markowski, UMN Center for Mass Spectrometry & Proteomics) was added to each of the samples to resuspend the cells. The cells were lysed and the supernatants collected in a process identical to that followed for the sample preparation of 2D gel electrophoresis. Supernatants containing the extracted proteins were then stored at -20° C for later use.

A standard Bradford assay (Thermo Fisher Scientific, MA) was used as before to quantify protein concentration in the supernatants. From the calculated protein concentration, 50 μg of protein for each sample was submitted to the Center for Mass Spectrometry & Proteomics (University of Minnesota) for iTRAQ analysis.

1.2.3.2 iTRAQ Labeling & Analysis

For *M. aquaeolei VT8* samples, a 2-plex labeling strategy was followed. The cell sample with no wax (control) was labeled with iTRAQ label 114 and the sample with wax was labeled with iTRAQ label 116. For *P. cryohalolentis* K5, a 4-plex labeling strategy was followed and the following isobaric labels were used. Sample with no wax (control) was labeled with iTRAQ labels 114 & 115 and sample with wax was labeled with iTRAQ labels 116 & 117. Actual method of labeling and the subsequent steps involving liquid chromatographic separation, MALDI spotting and MS/MS data collection were done as in Akkina et al., (2009). The mass spectrometry data was analyzed using the software ProteinPilot (AB SCIEX, Foster City, CA) available at the Minnesota Supercomputing Institute (University of Minnesota).

1.2.4 *M. aquaeolei* VT8 *Celstir* Growth

In order to better track wax production in *M. aquaeolei* VT8, a larger scale method of growing cells was required compared to utilizing 2 L shaker flask (1 L volume) cultures. A *Celstir* cell culture system (Wheaton, Millville, NJ) with a capacity to hold up to 4 L of liquid culture was obtained for timed studies of bacterial growth. This culture vessel has an internal magnetic stirrer that can be driven by placing the vessel on a magnetic stir plate. An aquarium pump was used to provide air via an autoclavable 0.2 μm filter (Millipore, Billerica, MA) and delivered through holes drilled in a piece of stainless steel tubing. Liquid medium for the growth of *M. aquaeolei* VT8 was prepared

and transferred to the *Celstir* vessel and the entire apparatus was autoclaved (held at 121° C for 60 minutes).

1.2.5 Growth Medium Composition

For 1 L of medium, the following composition was used: 50 g Sodium chloride, 7 g Sodium citrate, 5 g Magnesium sulfate heptahydrate, 500 mg Potassium phosphate dibasic, 200 mg Calcium chloride dihydrate, 35 mg Ferrous sulfate heptahydrate, 1L of water. The medium composition was derived from Huu et al., (1999). The medium was set to pH 7.3 (Orion 3 Star, Thermo Scientific, MA) prior to autoclaving. For the nitrogen source, either 3.75 mM ammonium sulfate (495 mg/L) or 7.5 mM sodium nitrate (637 mg/L) was used. Growth on 7.5 mM sodium nitrite (517.42 mg/L) was tested and was found to be severely impaired in terms of cell density and wax production even after prolonged culture conditions (~ 1 week). 4 L of medium was prepared as a batch and 3 L were transferred to the *Celstir* vessel. 984 mL of the medium was transferred to a 2 L Erlenmeyer flask. Two test tubes each containing 8 mL of the medium were also prepared. All were autoclaved together and allowed to cool overnight at room temperature.

A single colony was picked from a plate containing *M. aquaeolei* VT8 (grown on solid media with same composition as above and 12 g of agar/L) and inoculated into a prepared test tube (8 mL). The culture was grown for 24 hours at 30° C in a temperature controlled flask shaker (MaxQ 5000, Thermo Fisher Scientific, MA). The 8 mL culture was then transferred to the medium in the 2 L flask and grown for 20 hours, at 30° C and

shaken at 200 rpm in the shaker. After 20 hours, the entire shake flask culture was transferred into the *Celstir* containing 3 L of the medium.

The *Celstir* vessel was placed over a magnetic stir plate (Isotemp, Thermo Fisher Scientific, MA) and the stirring rate set to 60 rpm. The entire setup was placed in a constant temperature room (26.5° C). Sterile air was pumped into the vessel at a constant rate using the aquarium pump and a custom-made tubing system. For sampling cells at different time points of growth, 400 mL of the culture was transferred to a centrifuge bottle using sterile technique. The cells were pelleted in a centrifuge (Avanti J-E, Beckman Coulter, Brea, CA) at 11900g for 8 minutes at 4° C. Once completed, 45 mL of the supernatant was transferred to a 50 mL polypropylene tube (Falcon, Becton-Dickinson, Franklin Lakes, NJ) and stored at -20° C. Nitrate or ammonium test strips (Aquachek, Hach Company, Loveland, CO) were used to qualitatively measure the amount of nitrogen source left in the supernatant. The remaining supernatant was safely disposed. The cell pellet was resuspended in 10 mL of phosphate buffered saline (pH 7.4) and transferred to a 50 mL Falcon tube. The resuspended cells were rapidly frozen in liquid nitrogen and stored at -80° C before being freeze dried (Freezone6, Labconco, Kansas City, MO). Using a sterile transfer pipette, 12 mL of cells were removed from the *Celstir* vessel and transferred to a 15 mL polypropylene tube. The cell sample was filled into three sterile 2 mL microcentrifuge tubes and spun for 1 minute at 20,000 g. The supernatant was removed out and the centrifugation repeated with more cell culture. After pelleting cells from 4 mL of growth medium in each tube, the tubes were flash frozen in liquid nitrogen and stored at -80° C. RNA was extracted from these samples as

required. The sample collection process was repeated at 12 hour sampling intervals spanning 84 hours in initial *Celstir* growth studies to determine the time point of peak intracellular wax concentration. Based on these studies, sampling interval was changed to 4 hours spanning 36 hours of cell growth, with a final reading at 48 hours since start of *Celstir* growth.

1.2.6 Gas Chromatography Analysis of Wax Esters

1.2.6.1 Solvent Extraction of wax esters

To track and quantify the amount of waxes produced in *M. aquaeolei* VT8, an in-house protocol developed by Dr. Brett Barney (unpublished) was used. Solvent extraction of waxes was performed on 100 mg of freeze dried bacterial cells sampled at different time points of growth. The solvent used for extracting neutral lipids was an equivolume mixture of methylene chloride, n-hexane and tetrahydrofuran. Clean glass syringes were used at all stages of extraction (Hamilton, Reno, NV). 5 mL of solvent was added to a measured weight of dry cells in a glass test tube. The mixture was then subjected to sonication for 30 seconds followed by a break of 10 seconds. This process was repeated for 3 more times for a total sonication time of 2 minutes. After sonication, the solution was centrifuged for 2 minutes at approximately 3500 rpm in a bench top test tube centrifuge (Clinical 50, VWR Scientific, Radnor, PA). After centrifugation, the supernatant was carefully transferred to a clean I-Chem bottle (VWR Scientific, PA) for storing the extract, using a syringe fitted with a long needle. Care was taken not to disturb the pellet of cell material. Next, 5 mL of solvent was added and the pellet resuspended.

The sonication, centrifugation steps were repeated and once again, the supernatant was transferred to the I-Chem bottle. The extraction was performed one more time with 5 mL of fresh solvent. The extract was then brought up to 15 mL using fresh solvent. This extract was then prepared for sample analysis on the Gas Chromatography instrument.

1.2.6.2 Wax Esters Analysis

1 mL of the extract was transferred into a GC vial, provided with a Teflon coated septum, using a clean syringe. To this, 10 µL of an octacosane internal standard was added. The internal standard was prepared at a concentration of 5 mg/mL of solvent. To quantify the amount of wax in the test samples, a wax standard was prepared. 5 mg of palmityl palmitate and 5 mg of palmityl stearate were dissolved in 1.5 mL of solvent. From this, 25 µL of the solution was added to 975 µL of fresh solvent in a GC vial. Internal standard was added to this solution. The samples were then analyzed on a gas chromatograph (Shimadzu 2010 Plus, Shimadzu Scientific, Columbia, MD). Analytes were separated on an RTX-Biodiesel TG column (15 m, 0.32 mm ID, Restek, Bellefonte, PA) with a temperature profile of 60° C for 1 minute, and then increased to 360° C for 15 minutes at a ramp rate of 10° C per minute. 1 uL of the samples was injected into a programmable temperature vaporizer (PTV) that had the same temperature profile as the column. The flame ionization detector was set to 370° C and helium gas was used as the carrier gas. The amount of wax present in a sample was calculated by comparing peak area of the wax esters to the peak area formed by the wax standard. Calculation of peak

area was performed with the GC Solutions Postrun analysis software (version 2.32, Shimadzu Scientific, MD)

1.2.7 RNA Extraction Protocol

For all RNA methods, pipettes and working surfaces were wiped clean with an RNAse inhibiting solution (RNAse Zap, Sigma-Aldrich. St. Louis, MO). Barrier tips were used for all pipette transfers. Gloves were changed frequently to avoid RNAse contamination. RNeasy mini kit (Qiagen, Valencia, CA) was used for part of the RNA recovery process and the manufacturer's protocol was modified accordingly. From the -80° C freezer, microcentrifuge tubes containing cell pellets from each time-point were recovered and 1ml of TRIzol reagent (Invitrogen, Carlsbad, CA) was immediately added to each pellet. The cells were thoroughly resuspended in TRIzol and vortexed well for at least 5 minutes. The solution was then left at room temperature for 10 minutes. Then, 200 µL of chloroform was added to each tube and vortexed well for 1 minute. The solution was incubated at room temperature for 1 minute and again vortexed for 1 minute. The microcentrifuge tubes were then centrifuged at 15,000g for 10 minutes at room temperature. After centrifugation, 400 µL of the clear supernatant was gently transferred to a 2 mL microcentrifuge tube. To this, 500 µL of Qiagen RLT buffer and 500 µL of 100% ethanol were added and vortexed. 700 µL of this solution was then transferred to a sterile Qiagen MinElute column and spun at 10,000g for 30 seconds. After discarding the filtrate, the remaining 700 µL of the solution was also filtered through the spin column. The column was transferred to a new collection tube and 500 µL of Qiagen buffer RPE

was added. The column was spun at 10,000g for 30 seconds. After discarding the filtrate, 30 µL of a DNAse I solution (Qiagen, CA) was added directly to the column (This should not be added to the sides of the spin column). The column was incubated at room temperature for 15 minutes. Then, the column was washed with buffer RPE as before. The column was then washed twice with 750 µL of 80% ethanol (10,000g and 30 seconds). Finally, the column was transferred to a sterile 1.5 mL microcentrifuge tube and 50 µL of RNAse-free water was added to elute the RNA. The column was spun at 10,000g for 30 seconds. After centrifugation, the column was discarded and the eluate, containing RNA, was stored at -80° C.

Each RNA sample was quantified using a Nanodrop Spectrophotometer (Thermo Fisher Scientific, MA). 1 µL of each sample was used for the process. Absorbance values were obtained for wavelengths at 230, 260 and 280 nm. Samples with A260/230 and A260/280 ratios greater than 2.0 were considered good for further use. Based on the quantification, approximately 150 ng of RNA was submitted for RNA integrity check using the Agilent Bioanalyzer 2100 system available at the Biomedical Genomics Center (University of Minnesota). The Bioanalyzer system checks the integrity of the RNA sample on a scale of 1 to 10, with 10 being mostly intact RNA.

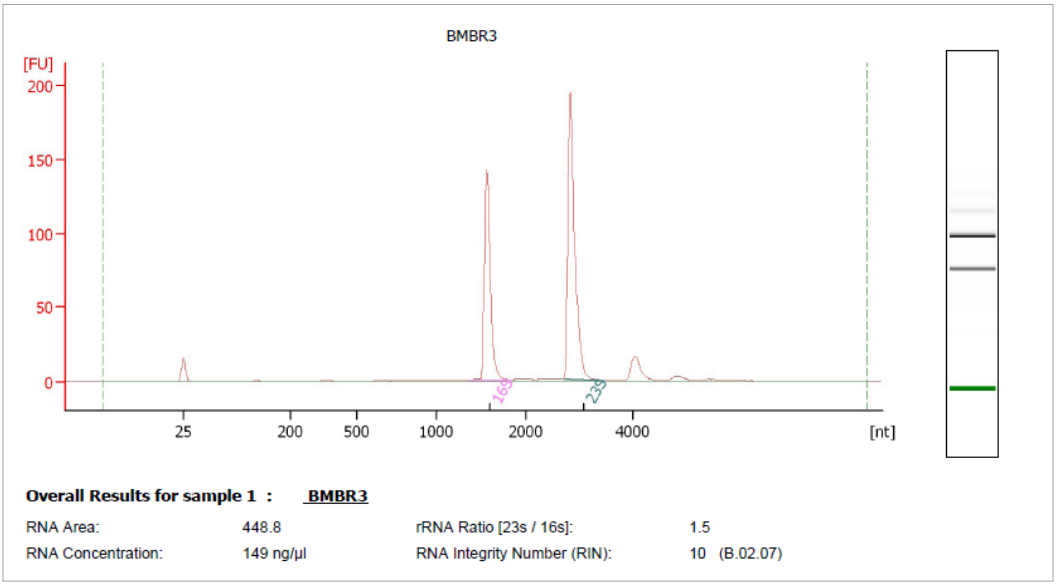


Figure 1.3: Electropherogram of an RNA sample extracted from *M. aquaeolei* VT8. The major peaks correspond to the 16s and 23s ribosomal RNA.

1.2.8 Quantitative Real Time Polymerase Chain Reaction

Primers for qRT – PCR gene targets were designed using NCBI Primer-Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) with custom parameters. PCR product size was set to between 190 and 210 basepairs, and primer melting temperatures were set between 57° C and 63° C with optimal value at 60° C.

qRT-PCR experiments were performed using the SuperScript III Platinum SYBR Green One-Step qRT-PCR Kit (Invitrogen, CA) on a Roche LightCycler 480 instrument (Roche Applied Science, Indianapolis, IN). The one-step kit contains a reverse transcriptase enzyme to reverse transcribe RNA present in the reaction mix into cDNA. This cDNA is then subjected to PCR amplification by the Taq polymerase present in the reaction mix. The fluorescence detection mode in the instrument was set to 'SYBR Green

I’. The PCR was performed in a 96 well plate format. A single well was used for each gene targeted for analysis. Each well contained 5 µL of 2x SYBR Green reaction mix, 0.2 µL of One-Step Enzyme mix, 1.6 µL of RNA at a concentration of 100 ng/µL, 1.6 µL of a primer mix containing 2.5 µM of each of the two primers specific to the gene target and 1.6 µL of RNAse-free water. The 96 well reaction plate was sealed with sealing foil and centrifuged at 500g for 1 minute in a compatible centrifuge (Sorvall Legend RT, Thermo Fisher Scientific, MA). The plate was then loaded on to the LightCycler 480 instrument. A qRT - PCR program was designed as follows. The program began by holding the reaction plate at 50° C for 10 minutes for the reverse transcriptase to create cDNA copies of mRNA. The temperature was then elevated to 95° C and maintained as such for 10 minutes to destroy the reverse transcriptase activity and also activate the Taq DNA polymerase present in the reaction mix. Then, 40 cycles of amplification were performed using a denaturation temperature of 95° C for 30 seconds followed by primer annealing and extension at 60° C for 30 seconds. Fluorescence intensity from the DNA binding dye was acquired at the end of each amplification step. After the amplification process, the temperature was maintained at 60° C for 1 minute before being ramped up to 97° C at a rate of 0.11° C per second. Fluorescence data was acquired throughout the temperature ramp-up process so that melting curve analysis could be performed later. At the end of the program, the plate was cooled to 40° C before being removed. Data analysis was performed using the LightCycler 480 software (Version 1.5, Roche Applied Science, IN) provided with the instrument.

Cp values or Crossing Point values, which correspond to the amplification cycle at which sample fluorescence is detected above the background fluorescence, were calculated using the 2nd Derivative Maximum method available in the LightCycler 480 software. Cp values are therefore inversely proportional to RNA concentration of the template. A reference gene was selected whose expression is expected to be constant under all the conditions considered for this study. Using the Cp value of the reference gene as the control, the fold change in expression of the target gene is calculated.

$$\text{Fold Change in Gene Expression} = \frac{E^{\Delta Cp\ (T_2 - T_1)_{\text{target gene}}}}{E^{\Delta Cp\ (T_2 - T_1)_{\text{reference gene}}}}$$

where E is the efficiency of PCR amplification of the selected gene by its primers, with value ranging from 1 to 2, with 2 corresponding to exponential amplification. T_1 and T_2 are the two different time points or conditions at which samples need to be compared. Target gene is the gene of interest and the reference gene is one whose expression is constant at both conditions T_1 and T_2. The formula presented here was adapted from Schmittgen & Livak (2008).

1.3 Results

1.3.1 2D Gel Electrophoresis

M. aquaeolei VT8 and *P. cryohalolentis* K5 accumulate intracellular wax esters when subjected to nitrogen nutrient limitation. To identify the proteomic changes between the cellular states in nitrogen abundant and nitrogen deficient conditions, two-dimensional gel electrophoresis technique was used. The rationale was to identify differentially expressed proteins to gain an understanding of what enzymes were regulated during the transition to nitrogen limitation and wax ester accumulation. The 2D gel electrophoresis technique achieves separation of proteins based on their size and isoelectric point (pH at which the protein has no net charge). Preliminary gel electrophoresis experiments (at the Center for Mass Spectrometry & Proteomics, University of Minnesota) for both *M. aquaeolei* VT8 and *P. cryohalolentis* K5 produced gels with smears that masked the individual protein spots from being imaged. Thus, a good quality gel separation and subsequent imaging could not be achieved due to unknown reasons. Hence the gel-based technique was not utilized for further studies.

1.3.2 iTRAQ Analysis of *P. cryohalolentis* K5

iTRAQ analysis was performed using *P. cryohalolentis* K5 total cellular protein samples obtained from cells grown in non-limiting and limiting amounts of the nitrogen nutrient. From the experiment and subsequent computational analysis, a set of 51 proteins were found to differentially express with high statistical significance.

Table 1.1: Differentially regulated proteins during wax ester accumulation in *P. cryohalolentis* K5. The test sample was grown in a limiting concentration of nitrogen nutrient compared to the control. A negative value for the Log_2 ratio indicates that the specific protein is present in lower abundance in the test sample compared to the control sample. A positive value indicates higher abundance of the protein in the test sample.

Accession Number	Protein Name (from NCBI annotation)	Log_2 (test:control)
gi|93006563	adenylate kinase	1.38
gi|93006301	cold-shock DNA-binding domain-containing protein	1.04
gi|93007146	30S ribosomal protein S16	0.85
gi|93006774	translation elongation factor P	0.85
gi|93005141	peptidylprolyl isomerase, FKBP-type	0.84
gi|93006258	hydroxyacylglutathione hydrolase	0.81
gi|93004962	HAD family hydrolase	0.81
gi|93005658	GreA/GreB family elongation factor	0.76
gi|93005839	ThiJ/PfpI	0.76
gi|93006022	hypothetical protein Pcryo_1194	0.75
gi|93005477	nucleoside-diphosphate kinase	0.71
gi|93005229	peptidylprolyl isomerase, FKBP-type	0.69
gi|93006890	thioredoxin	0.63
gi|93005103	4Fe-4S ferredoxin, iron-sulfur binding	0.60
gi|93006537	ribosome recycling factor	0.58
gi|93006684	isocitrate lyase	0.55
gi|93006577	outer membrane lipoprotein carrier protein LolA	0.45
gi|93005570	hypothetical protein Pcryo_0740	0.43
gi|93005605	alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	0.31
gi|93005576	bifunctional aconitate hydratase 2/2-methylisocitrate dehydratase	0.27
gi|93006704	hypothetical protein Pcryo_1880	0.26
gi|93005931	amidophosphoribosyltransferase	-0.12
gi|93005326	50S ribosomal protein L5	-0.21
gi|93005357	electron transfer flavoprotein beta-subunit	-0.22
gi|93005185	hypothetical protein Pcryo_0354	-0.23

Table 1.1 Continued: Differentially regulated proteins during wax ester accumulation in *P. cryohalolentis* K5

gi|93005641	17 kDa surface antigen	-0.24
gi|93005862	phosphoenolpyruvate synthase	-0.25
gi|93005354	ketol-acid reductoisomerase	-0.26
gi|93005671	ATPase AAA-2	-0.29
gi|93004849	zinc-binding alcohol dehydrogenase	-0.30
gi|93007051	multifunctional fatty acid oxidation complex subunit alpha	-0.30
gi|93004916	hypothetical protein Pcryo_0085	-0.31
gi|93006509	Lipocalin-like	-0.32
gi|93005992	acetyl-CoA carboxylase, biotin carboxylase	-0.35
gi|93006497	30S ribosomal protein S1	-0.44
gi|93007280	chaperone DnaK	-0.47
gi|93004809	peptidase	-0.48
gi|93006724	superoxide dismutase, copper/zinc binding	-0.51
gi|93005251	beta-lactamase-like	-0.51
gi|93007208	aconitate hydratase	-0.53
gi|93005833	inosine-5'-monophosphate dehydrogenase	-0.54
gi|93006248	hypothetical protein Pcryo_1422	-0.54
gi|93005878	pyruvate dehydrogenase subunit E1	-0.61
gi|93005896	acetyl-CoA acetyltransferase	-0.61
gi|93004815	CsbD-like	-0.70
gi|93006222	phosphoadenosine phosphosulfate reductase	-0.74
gi|93006682	aldehyde dehydrogenase	-0.85
gi|93005322	50S ribosomal protein L29	-0.87
gi|93006911	OmpA/MotB	-0.96
gi|93007164	TraR/DksA family transcriptional regulator	-1.02
gi|93004968	hypothetical protein Pcryo_0137	-1.65

A total of 1968 peptides from 450 proteins were identified at a False Discovery Rate (FDR) of 1%. The FDR is a measure of false positives expected in matching identified peptides to actual protein sequences and is calculated as the number of matches

obtained when matching the peptide sequences identified by the mass spectrometer to a *P. cryohalolentis* K5 protein database that contains the protein sequences in reverse. Differential regulation of proteins was accepted to be statistically significant only if the P-value for test to control ratio was less than 0.05 and the error factor was less than 2. The error factor is an instrumental parameter and gives a measure of confidence for protein quantitation based on the quantitation of each peptide detected for a protein. After applying the necessary conditions for statistical significance, 21 proteins were found to be present in greater abundance in the wax accumulating cells. 30 proteins were found to be of lesser abundance in the same sample.

Among the identified proteins, a number of transcription factors and hypothetical proteins were present. The enzyme *adenylate kinase* which mediates the interconversion of adenosine-diphosphate nucleotide in to adenosine triphosphate (ATP) and adenosine monophosphate (AMP) was found to have the highest positive fold change in expression in the wax producing sample. This enzyme impacts the cellular availability of ATP (energy molecule) and AMP (signaling molecule). A number of ribosome associated proteins were found to be up-regulated, like 30S ribosomal protein S16, translation elongation factor P and GreA/GreB family elongation factor whereas other ribosome associated proteins like 30S ribosomal protein S1 and 50S ribosomal protein L29 were down-regulated in wax containing samples.

1.3.3 iTRAQ Analysis of *M. aquaeolei* VT8

iTRAQ analysis was performed using *M. aquaeolei* VT8 total cellular protein samples obtained from cells grown in non-limiting and limiting amounts of the nitrogen nutrient. From the experiment and subsequent computational analysis performed similar to *P. cryohalolentis* K5, 646 proteins were reliably identified out of which 26 were found to differentially regulated with high statistical significance.

A number of identified proteins cluster together in nucleotide metabolism (*dihydroorotate oxidase, uridylate kinase, adenylosuccinate synthase),* perhaps suggesting a role for conservation of nitrogen present in cellular nucleotides. 3-oxoacyl-(acyl carrier protein) synthase III, an enzyme of the fatty acid biosynthetic pathway, was 1.39 fold more abundant in wax accumulating cells of *M. aquaeolei* VT8. An interesting observation is the lower abundance of the enzyme *aconitate hydratase* in both species under nitrogen limiting condition. The enzyme is 98% similar between the species and suggests a similar role in both species.

Table 1.2: Differentially regulated proteins during wax ester accumulation in *M. aquaeolei* VT8. The test sample was grown in a limiting concentration of nitrogen nutrient compared to the control. A negative value for the Log_2 ratio indicates that the specific protein is present in lower abundance in the test sample compared to the control sample. A positive value indicates higher abundance of the protein in the test sample.

Accession Number	Protein Name (from NCBI annotation)	Log_2 ratio test:control
gi|120556771	extracellular ligand-binding receptor	2.67
gi|120553092	aspartate kinase	1.44
gi|120554298	3-oxoacyl-(acyl carrier protein) synthase III	1.39
gi|120553893	leucyl aminopeptidase	1.06
gi|120553748	phosphoenolpyruvate-protein phosphotransferase PtsP	1.00
gi|120553608	TraR/DksA family transcriptional regulator	0.93
gi|120556561	OmpA/MotB domain-containing protein	0.88
gi|120555128	aminotransferase, class I and II	0.84
gi|120555413	GCN5-related N-acetyltransferase	0.68
gi|120553652	ribosomal protein L3	0.60
gi|120554647	GMP synthase	0.60
gi|120556743	alkyl hydroperoxide reductase/ Thiol specific antioxidant/ Mal allergen	0.59
gi|120554763	glutaminyl-tRNA synthetase	0.47
gi|120554939	phosphoribosylaminoimidazole carboxylase ATPase subunit	0.40
gi|120554394	peptidylprolyl isomerase, FKBP-type	0.37
gi|120553816	ketol-acid reductoisomerase	0.26
gi|120555420	peptidase M16 domain-containing protein	-0.28
gi|120554070	UspA domain-containing protein	-0.29
gi|120553424	porphobilinogen deaminase	-0.67
gi|120553965	dihydroorotate oxidase	-0.67
gi|120555457	uridylate kinase	-0.73
gi|120555418	short chain dehydrogenase	-0.90
gi|120555675	adenylosuccinate synthase	-1.22
gi|120555503	flagellin domain-containing protein	-1.23
gi|120556579	aconitate hydratase 1	-1.40
gi|120556035	hydroxylamine reductase	-2.49

1.3.4 Outcomes from the iTRAQ Experiment

The iTRAQ experiment was successful in identifying 15% of the predicted proteome size in *M. aquaeolei* VT8 and 18% of all predicted proteins in *P. cryohalolentis* K5. The identified proteins were found to be distributed among diverse metabolic pathways. While some proteins could be readily assigned to various pathways, others with broadly defined function like *aldehyde dehydrogenase* could not assigned to any one distinct pathway. The abundance ratios of certain proteins pointed towards possible underlying trends in metabolic regulation. Since enzymes identified with nucleotide metabolism were found to be of lower abundance during wax accumulation, it could be hypothesized that pathways for nucleotide formation might be down-regulated. However, the lack of expression ratios for other enzymes in these pathways does not allow us to validate the hypothesis. A similar case could be made for elevated expression of enzymes of fatty acid biosynthesis and suppression of amino acid biosynthetic pathways, but lack of data on the abundance ratio of some enzymes in these pathways inhibits us from drawing definitive conclusions. An important piece of information that is of much interest is the role of transcription factors that were found to be differentially expressed. Transcription factors act as regulators of multiple genes and thereby act as a global switch in the cell's metabolism. The paucity of data on the particular function of transcription factors in the species studied here can be attributed to the rarity of biochemical research in these species.

The key enzyme in the wax ester biosynthesis pathway, the wax ester synthase/Diacyl glycerol acyltransferase (WS/DGAT) has been proposed to be membrane

bound and hence would not be observed in the iTRAQ experiment. Hence alternate methods were required to quantify the expression of those proteins that were not identified by this experiment. These methods were also needed to complement the information provided by the iTRAQ experiment, so that a rigorous conclusion could be derived regarding metabolic changes during wax biosynthesis.

1.3.5 *M. aquaeolei* VT8 growth in *Celstir* Growth Vessel

M. aquaeolei VT8 was chosen as the prototypical organism for further laboratory studies of gene expression changes in wax biosynthesizing bacteria. This decision was based on the fact that among the wax ester producing bacteria available for our laboratory, *M. aquaeolei* VT8 produced the highest concentration of waxes, is easier to grow in defined media, and has a better growth rate especially when compared to *Psychrobacter* species. Also, a number of genes from the organism have been and are being biochemically characterized in our laboratory and the information arising from such studies could be incorporated in our gene expression studies.

Based on a number of preliminary experiments, nitrate was found to be a better nitrogen source for wax ester biosynthesis compared to ammonium salts. Therefore, nitrate was used as the nitrogen source in the experiment whose data is presented here. Samples were collected at 6, 12, 16, 20, 24, 28, 32, 36 and 48 hours after inoculation of the medium in *Celstir* vessel with *M. aquaeolei* VT8. The *Celstir* vessel enabled growth of *M. aquaeolei* VT8 to a dense culture. Spectrophotometric readings of optical density could not be obtained over the period of the experiment since after approximately 12 hours of growth in the vessel, cells formed floc like aggregates and settled at the vessel's

bottom if left unstirred and undisturbed. Biofilm formation was also observed after this time point, with cells attached to the vessel's stirrer and the walls at the liquid-air interface. Nitrate test strips qualitatively indicated that nitrate present in the medium was exhausted around 12 hours of growth, with only 1- 2 parts per million being detected at 12 hours. No nitrate (zero ppm) was detected at 16 hours.

1.3.6 Gas Chromatography Analysis of Wax Esters

The GC analysis of cell extracts from various time points shows the accumulation of waxes over the period of the experiment. From preliminary experiments, we found that for the given growth condition, the maximum concentration of intracellular waxes was attained between 32 – 48 hours of growth in the *Celstir* vessel. Thus, the growth period for the actual experiment had been restricted to 48 hours. A visual inspection of the chromatograms showed that wax concentration was indeed the highest at around 36 hours of the growth experiment. Surprisingly, wax esters were detected even in the sample from the 6 hour time point, when there was sufficient nitrogen source in the medium. However, the amount of waxes detected at 6 hours was significantly low (0.3% of dry cell weight) compared to values during wax ester accumulating conditions, as reported in prior publications and also observed in experiments in our laboratory. GC analysis of wax ester extracts showed that the predominant intracellular wax esters in *M. aquaeolei* VT8 were composed of C16 and C18 components. Wax esters matching palmityl palmitate (C16-0:C16-0) and palmityl stearate (C16-0:C18-0) but with up to two degrees of unsaturation were predominant in terms of distribution of the wax ester profile. Wax esters containing

C14 components increased in concentration after 16 hours. Maximum intracellular wax concentration was found to be 10.31% of dry cell weight, after 36 hours of survival in the *Celstir* vessel.

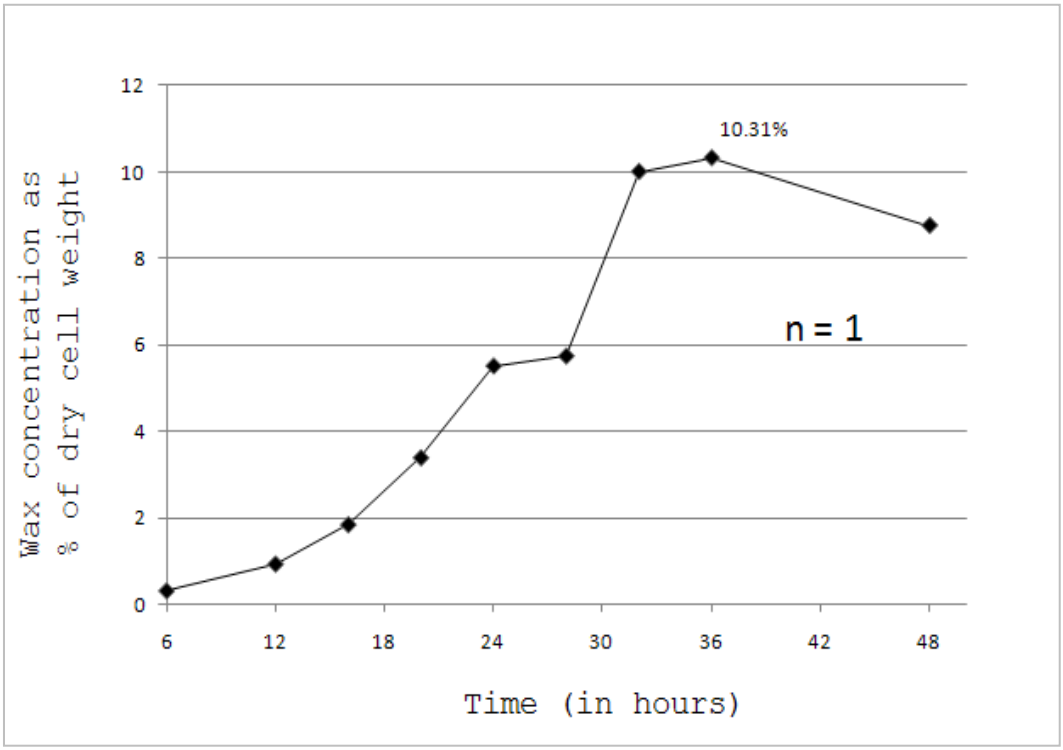


Figure 1.4: Concentration of wax esters as percentage of dry cell weight. Peak concentration was observed at 36 hours, with nitrate being exhausted at the 12 hour mark.

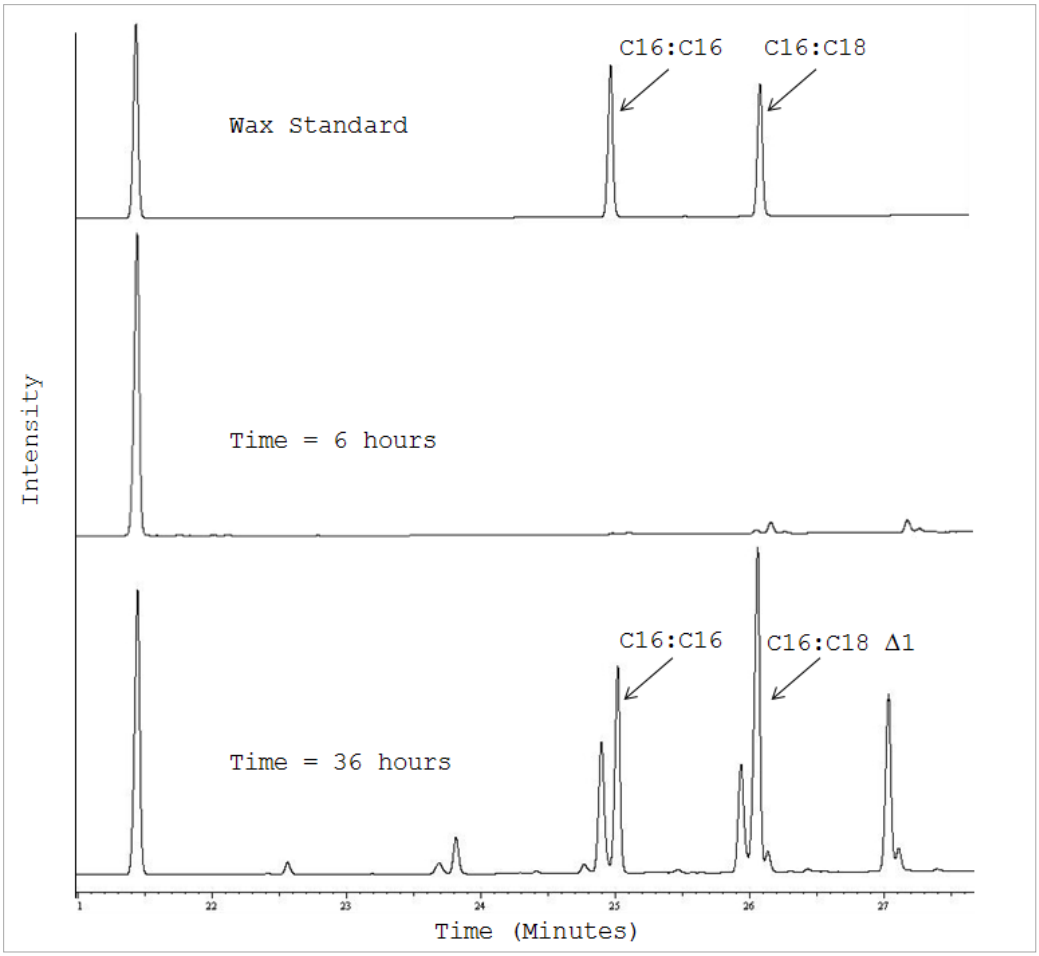


Figure 1.5: Gas chromatogram of samples showing wax esters in *M. aquaeolei* VT8 extract. Only traces of wax are observed at 6 hours compared to waxes at 36 hours. The wax standard was used to identify unsaturated wax esters. The first peak to the left is the internal octacosane standard.

1.3.7 qRT – PCR Analysis of Gene Targets

Based on the results of the proteomic study, a qRT-PCR study was designed. This study was intended to serve two purposes. The first was to confirm if the proteins identified to be differentially expressed by the iTRAQ study were indeed undergoing such a change. The second was to check if during the condition of wax ester biosynthesis, the expression of the wax ester synthase genes is increased above levels observed otherwise. In addition to the four wax ester synthase genes present in *M. aquaeolei* VT8, several other gene targets were chosen for analyses. These targets included genes that were identified by the iTRAQ study to be differentially regulated, like *aconitate hydratase*, *hydroxylamine reductase* and *aspartate kinase*. A qRT-PCR experiment was conducted targeting the expression of the genes including the four wax ester synthases, a fatty aldehyde reductase, *aconitate hydratase*, *hydroxylamine reductase*, *aspartate kinase* and *3-oxoacyl-(acyl carrier protein) synthase III*. *Ribosomal Sigma Factor D* (RpoD) transcript was used as the reference for expression fold change comparison. The abundance of the gene targets was compared between two time points between which a great difference in intracellular wax concentration was observed.

The 12 hours time point was the time at which the first observations of wax formation were noted and at 36 hours, the intracellular wax concentration was found to peak. Therefore, these two time points were chosen for comparison. Compared to cells from 12 hours of growth, cells harvested at 36 hours showed an increase in abundance in the transcripts of all wax ester synthase genes. Of note, the abundance of the *wax ester synthase I* transcript increased more than the other isoforms of the enzyme. Incidentally,

this particular isoform of the wax ester synthase enzyme has been demonstrated to possess the highest enzymatic activity among all the wax ester synthases identified so far in *M. aquaeolei* VT8 (Holtzapple & Schmidt-Dannert, 2007). This leads to the notion that the *wax ester synthase I* enzyme might be most active at oberved physiological conditions. The expression of *fatty aldehyde reductase*, a gene whose protein product is part of the pathway leading to wax biosynthesis, was also increased at 36 hours. The remaining genes were included to test if the iTRAQ results and the qRT-PCR results were in agreement with each other. As observed in figure 1.6, the transcripts for *aconitate hydratase* and *hydroxylamine reductase* were of decreased abundance during wax biosynthesis. *Aspartate kinase* and *3-oxoacyl-(acyl carrier protein) synthase III*, which is part of the fatty acid biosynthetic pathway, were found to be upregulated, all in agreement with the iTRAQ data.

The downregulation of the enzyme *aconitate hydratase* during wax biosynthesis could be a mechanism of re-directing carbon flux towards fatty acid biosynthesis. The lower abundance of *aconitate hydratase* would impair the consumption of citrate by the Tricarboxylic acid cycle and lead to an increased availability of citrate for conversion into acetyl-CoA, which could then be fed into fatty acid anabolism. Thus, if fatty acid biosynthesis from citrate could be considered as the objective of the bacterial cell during wax biosynthesis, downregulation of *aconitate hydratase* can be hypothesized to be minimizing the metabolic flux of citrate in a competing pathway. Also, the upregulation observed with *3-oxoacyl-(acyl carrier protein) synthase III* enzyme, a part of the fatty acid biosynthetic pathway, supports the idea that there is increased metabolic flux along

this pathway.

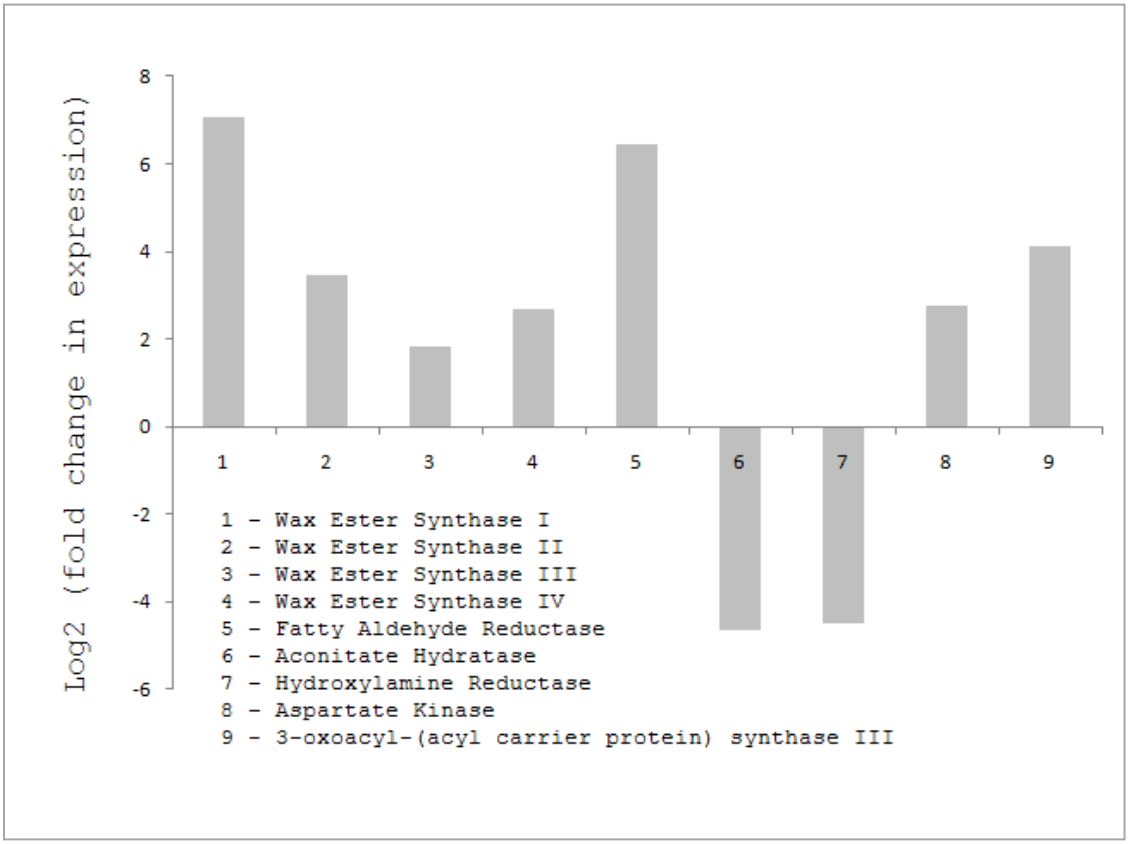


Figure 1.6: Change in expression of select genes during wax biosynthesis in *M. aquaeolei* VT8. The values represent transcript abundance of target genes in wax accumulating cells harvested at 36 hours, compared to cells from 12 hours of growth. The fold change in expression is expressed as Log2 values for ease of representation.

While conducting these experiments, a few technical issues were encountered. Over the course of experiments, many reference genes were considered. The most commonly used reference gene, the 16s rRNA (ribosomal RNA), was found to be a poor reference gene for our studies since preliminary experiments showed that the high concentration of rRNA in our samples quickly saturated the fluorescence detection system, thereby distorting any calculation of fold change in gene expression. Other genes that were considered as references were *Recombinase A*, *Polyhydroxybutyrate synthase*, *Ribosomal Sigma Factor D* (RpoD), *RNA polymerase subunit B* (RpoB) and *Glucose-6-phosphate Isomerase*. Among these genes, RpoD was found to have Cp values that had least deviation for two given samples.

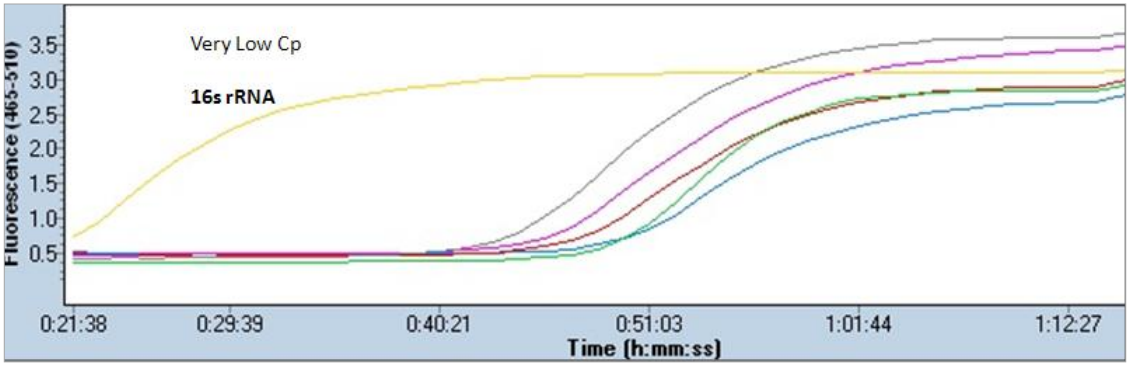


Figure 1.7: Tracking fluorescence intensity over time in a qRT-PCR experiment. Each curve corresponds to a gene target. The great abundance of 16s rRNA in bacterial samples causes rapid saturation of the instrument's ability to detect DNA amplification.

Factors like variation in quality of extracted RNA and RNA degradation seem to affect the qRT-PCR process. Melting curve analysis, which is used to determine the target specificity of primers, found that some primer pairs targeting certain genes were not ideal

for the process. The transcript profiles from those genes were omitted from this analysis. These issues temporarily affect the reproducibility of the obtained results.

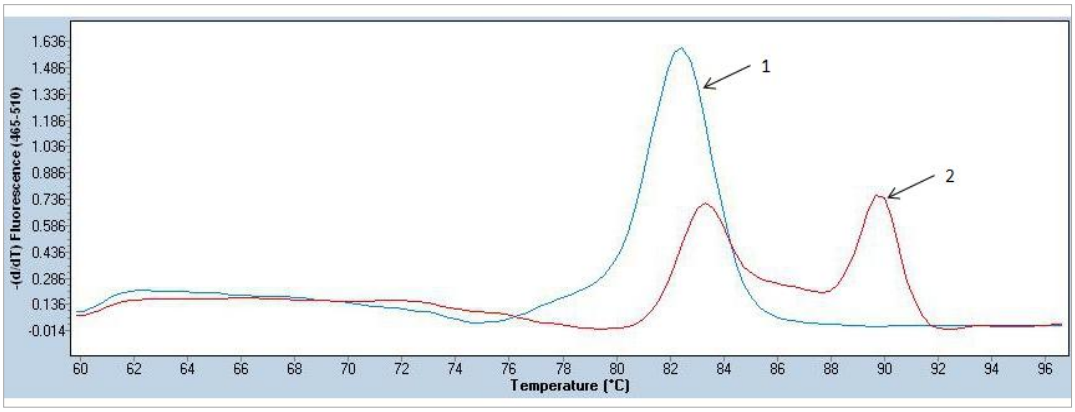


Figure 1.8: Melting curve analysis to determine primer efficacy. Ideal primers are those that show a single peak (1) indicating that the primers are highly specific to the gene target. Multiple peaks as in (2) indicate that the primers may be binding to multiple targets within the gene or in different genes.

1.4 Discussion

We intended to gain an understanding of the cellular mechanism(s) by which biosynthesis of wax esters is regulated in select wax ester biosynthesizing bacteria. In order to induce such bacteria into producing wax esters, we took advantage of the fact that when such bacteria are subjected to starvation of nitrogen nutrients, they start biosynthesizing and accumulating carbon-rich wax esters. This behavior of *de novo* wax biosynthesis has been observed in bacteria like *M. aquaeolei* VT8 and *P. cryohalolentis* K5, two organisms on which we have focused our exploratory studies aimed at understanding regulation of wax biosynthesis. Intracellular wax ester accumulation has been hypothesized to be a mechanism for energy and carbon storage, to be used when utilizable sources of nitrogen are more abundant to the bacterial cell. We have conducted this study to decipher the mechanisms of biochemical regulation that are responsible for triggering wax biosynthesis and accumulation under conditions of nitrogen starvation.

Our initial approach was to compare the proteomes of a bacterial species grown in two conditions that differ only by the availability of nitrogen. One of the conditions was designed to contain a low amount of the sole nutrient that supplies nitrogen to the bacteria. This condition then is limiting to the exponential growth of the cell. When cells senses lack of nitrogen nutrients, they begin to synthesize wax esters from the carbon source available to it. We compared the proteomes of *P. cryohalolentis* K5 grown in limiting and sufficient concentrations of the nitrogen source, using a mass spectrometry based technique called iTRAQ. We also performed a similar study with *M. aquaeolei* VT8. After analytical experimentation and computational analyses, we obtained a list of

proteins that were differentially expressed in each species, between the two conditions compared. We hypothesized on the role of a number of proteins in the list but were hindered from developing a conclusion by the paucity of biochemical data on the hypothetical enzymes and transcription factors of the bacterial species under study.

The success of the proteomic approach urged us to find other methods that could complement the information obtained from the iTRAQ experiments and thereby help us formulate a rigorous explanation of the metabolic regulation during intracellular wax accumulation. But before we sought additional information, we wanted to improve the quality and quantity of our cell sampling methods. We found our flask growth methods to be inadequate for fine-grained tracking of wax biosynthesis and its intracellular accumulation. To alleviate these problems, we customized a commercially available cell culture system to grow batch cultures of *M. aquaeolei* VT8 at volumes up to 4 Liters. Cells were periodically sampled over the course of growth spanning 48 hours and parameters like exhaustion of nitrogen source with time and change in pH were monitored. Cell samples from the different time points were analyzed for wax content and the intracellular wax concentration quantified. We found that only traces of wax esters were present during early stages of growth when the nitrogen nutrient is abundant in the growth medium. However, as this nutrient is exhausted, the intracellular concentration of wax increased rapidly, reaching up to 10% of dry cell weight in 36 hours of growth. This fine-grained tracking of wax ester accumulation gave us the confidence towards studying the gene expression changes associated with the phenomenon.

We identified quantitative real time PCR as a technique that would provide us with results on differential gene expression profiles of certain genes that we reasoned were important for wax biosynthesis. Our genes of interest included genes for the wax ester synthase enzymes, several genes involved in the formation of wax ester precursors, genes in fatty acid biosynthesis and tricarboxylic acid cycle. Based on the supporting data from the proteomic experiments, we hypothesized that expression of some of the selected genes should be suppressed while others should be elevated during wax biosynthesis. The information provided by the iTRAQ experiment, now complemented by transcript abundance profiling, provides subtle hints to how regulation of metabolic flux might be affected during wax ester biosynthesis. Expanding the number of gene targets for analysis would yield further information that could be integrated into a plausible model of metabolic regulation in *M. aquaeolei* VT8 when it exhibits the phenomenon of wax biosynthesis.

1.5 Future Work

Current issues with the qRT-PCR technique pertain to quality of the RNA used for reverse transcription and also, sub-optimal design of some primer pairs. Therefore, priority should be given to developing and standardizing an RNA extraction protocol that can produce consistent, high quality RNA extracts. Certain primer pairs have been deemed inefficient based on the melting curve analysis. These should be redesigned and re-tested to ensure specificity in target gene amplification. The selection of more reference genes will improve the quality and reliability of the gene expression changes observed in target genes.

Chapter II

Development of a Genetic System for Recombinant Gene Expression in *Psychrobacter arcticus* 273-4

2.1 Introduction

Plasmid based vector systems have been quintessential in expression of non-native proteins in bacterial hosts. They have also found use in expression of native proteins driven by various strong promoter systems whenever over-expression of such proteins is desired. The isolation and subsequent modification of a number of plasmids from different species of bacteria has resulted in the availability of a variety of DNA vector systems that provide modularity in heterologous gene expression in both laboratory and industrial bacterial bioprocesses.

Wax ester synthesizing bacteria like *Marinobacter aquaeolei* VT8 possess the cellular machinery required to produce significant amounts of wax esters (Manilla-Pérez et al., 2010). This machinery can be taken advantage of to develop a cellular factory that could potentially lead to industrial scale production of bacterial wax esters. To modify the profile of wax esters produced, some genetic additions and deletions are required in the gene repertoire available to wax producing bacteria. These modifications could be designed to enable the bacteria to produce non-native wax esters, such as fatty acid ethyl esters (biodiesel). Expression of a foreign gene or the over-expression of a native gene under an inducible expression system requires the introduction of *in vitro* engineered DNA into the bacterium. Such a process then requires a vector system that can operate within these bacteria and a method that can deliver the vector DNA into the bacterial cell.

From the literature, conjugal transfer of plasmid was found to be the most effective method of delivering foreign vectors into bacterial species for which direct transformation methods had not yet been developed. The conjugation process involves

transfer of DNA from a host or donor strain of bacteria to a target strain through cell-to-cell physical contact, through a filamentous appendage known as a sex pilus. *Escherichia coli* strain WM3064 was selected as the donor strain from which the plasmids would be conjugally transferred to the various bacterial strains under study. *E. coli* WM3064 is a strain of *Escherichia coli* that is a DAP auxotroph. DAP (diamino-pimelic acid) is a lysine derivative that is incorporated in the bacterial cell wall and hence, required for normal cell growth. Thus, the DAP auxotrophy provides a mechanism for influencing the survival of the WM3064 strain on nutrient plates (Saltikov & Newman, 2003).

We researched the literature on plasmid systems that could operate in a broad spectrum of bacterial species. The pBBR1MCS series of plasmids were identified to be broad spectrum plasmids that were compatible with a wide range of micro-organisms (Kovach et al., 1995; Kovach et al., 1994). The pBBR1MCS series of plasmids contain an antibiotic resistance gene for positive selection on growth plates, a gene coding for a mobilization element (MOB) that is required for transfer of the plasmid from the host to the target strains and an origin of replication where plasmid replication is initiated. Additionally, a blue-white selection scheme for introducing genes into the plasmid was provided via a number of restriction endonuclease sites and the lacZ gene fragment of the β-galactosidase enzyme.

The key advantage of the self-transmissible pBBR1MCS plasmid series is the less time required for transfer of the plasmid into the host organism over similar conjugation techniques like tri-parental mating. Tri-parental mating is a conjugation process involving three strains of bacteria – a donor strain containing a non-self-transmissible plasmid like

a pUC vector, a target host strain and an intervening helper strain that mobilizes the donor's plasmid into the target. Such a system increases the time required for a successful conjugation process and also increases the effort towards isolating pure colonies of the target strain now containing the plasmid.

Our intention was to develop a vector system for expression of foreign genes in *M. aquaeolei* VT8 and other wax producing bacteria from indigenous gene promoters, in order to modify the profile of wax esters produced. Such a vector system will also allow us to express a strain's native genes from non-cognate promoter systems that are functional in the strain, thereby allowing us control over gene expression in these bacterial species.

2.2 Material & Methods

2.2.1 Plasmid Vectors

Plasmids pBBR1MCS1, pBBR1MCS2 and pBBR1MCS3 were obtained from Dr. Jeff Gralnick of the Biotechnology Institute, University of Minnesota. These plasmids were constructed by Kovach et al. (1995)

Table 2.1: The pBBR1MCS series of plasmids used in this study

Plasmid	Antibiotic Resistance
pBBR1MCS1	Chloramphenicol
pBBR1MCS2	Kanamycin
pBBR1MCS3	Tetracycline

2.2.2 Bacterial Strains

E. coli JM109 was used for creating stocks of pBBR1MCS series plasmids and for long term storage as frozen stocks.

For conjugal transfers of the plasmids into target strains, *E. coli* strain WM3064 (W. Metcalf, University of Illinois) obtained from Dr. Gralnick was used. *E. coli* WM3064 was maintained on LB plates which contained 15 mg/ml DAP. Wild type *M. aquaeolei* VT8, *Psychrobacter cryohalolentis* K5 and *Psychrobacter arcticus* 273-4 were obtained from the American Type Culture Collection (Manassas, VA) and maintained on regular LB plates without any DAP.

2.2.3 Bacterial Growth Methods

E. coli WM 3064 was transformed with individual pBBR1MCS plasmids through chemical transformation methods. The transformed cells were grown on LB plates supplemented with DAP and an antibiotic for selection. After overnight growth at 37° C, single colonies were picked and maintained as streaks of pure culture in DAP and antibiotic supplemented LB plates. To prepare cells for conjugation, lawn cultures of *M. aquaeolei* VT8, *P. cryohalolentis* K5 and *P. arcticus* 273-4 were grown on individual LB plates. *E. coli* WM3064 strains with plasmids were grown in an overnight LB liquid culture at 37° C, with DAP & antibiotic. *M. aquaeolei* VT8 cells were scraped from the lawn, mixed with a 100 µL suspension of plasmid carrying *E. coli* WM3064 cells concentrated from 1 mL of culture, and spotted on to LB plates containing only DAP and no antibiotic. The plates were then placed undisturbed at either room temperature or at 30° C. A similar process was done for the *Psychrobacter* strains. After 1 day, the spots of cells were scraped, washed thrice with 1 mL LB liquid medium to remove DAP from the cells which were then plated on LB plates containing a selection antibiotic and no DAP. The plates were incubated at room temperature for *Psychrobacter* species or at 30° C for *M. aquaeolei* VT8 until individual colonies were observed. These colonies represented the target strain now containing the plasmid donated by the host strain. The absence of DAP in the plate medium prevents the growth of the *E. coli* WM3064 strain and hence, colonies of the plasmid-containing target strain were obtained. These colonies were then streaked on antibiotic containing LB plates to isolate pure colonies, without any trace of residual donor strain cells. The colonies were then maintained as pure cultures on LB

plates supplemented with an appropriate antibiotic. Single colonies were picked and grown in LB liquid medium containing an appropriate antibiotic and cells in exponential growth phase were resuspended in 1 mL of LB-Glycerol solution. The cell suspension was then flash frozen in liquid nitrogen and stored at -80° C as a frozen stock.

To verify the presence of plasmids in the target strains, total DNA was extracted from cell pellets using a genomic DNA isolation kit from Zymo Research, Irvine, CA. Polymerase chain reaction with primers specific for the introduced plasmid was used to confirm the presence and size of the plasmid. The PCR was performed using the FailSafe reagents and protocol obtained from Epicentre Biotechnologies, Madison, WI. Oligonucleotide primers were obtained from Integrated DNA Technologies, Coralville IA.

The pBBR1MCS2 plasmid was modified by Dr. Brett Barney to incorporate a 38 base pair sequence hypothesized to confer immunity against transposon insertions. The sequence was chemically synthesized as partly complementary oligonucleotides BBP942 and BBP943. The oligonucleotides were annealed together in a thermal cycler that was gradually cooled down from 95° C to room temperature. The annealed sequence had overhanging ends corresponding to the overhangs generated in pBBR1MCS2 linearized by restriction digestion. The sequences were joined together by ligation.

Table 2.2: List of primers used in the development of plasmid vector for *P. arcticus* 273-4

Primer	Sequence
BBP 828	GCGCAACGCAATTAATGTGAGTTAGC
BBP 829	ACCTCGCTAACGGATTCACCGTT
BBP 840	TGGAACGACCCAAGCCTATGCGAGT
BBP 841	CGCATGATTGAACAAGATGGATTGCAC
BBP 852	GTGCAATCCATCTTGTTCAATCATGCG
BBP 853	ACTCGCATAGGCTTGGGTCGTTCCA
BBP 910	GCGTCCCCGGTATAACGATG
BBP 911	GGGTCTGACGCTCAGTGG
BBP 912	GAACGAAATAGACAGATCGCTGAG
BBP 913	CAGGCAGCCCTGGATCATCTC
BBP 942	5’ - CATGTACTTAACGTGAGTTTTCGTTCCACTGAGCGT CAGACCCCTGAATTCTGCA – 3’
BBP943	5’ – GAATTCAGGGGTCTGACGCTCAGTGGAACGAAAACTCACG TTAAGTA – 3’

2.3 Results

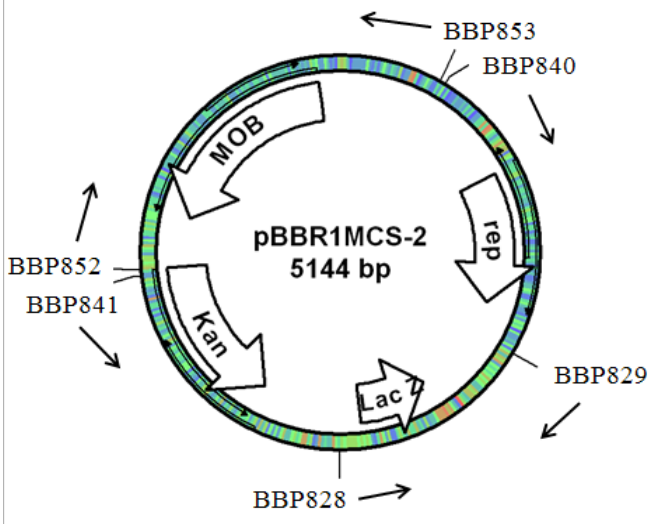
2.3.1 Conjugal Transfer of pBBR1MCS Plasmids

Preliminary growths of the wild-type *M. aquaeolei* VT8, *P. cryohalolentis* K5 and *P. arcticus* 273-4 were performed on LB plates containing varying concentrations of the antibiotics - chloramphenicol, kanamycin and tetracycline. This was done to test the presence of any background resistance of these strains to the mentioned antibiotics. The wildtype strains of *M. aquaeolei* VT8 and *P. cryohalolentis* K5 were found to be resistant to chloramphenicol antibiotic at up to 15 µg/mL concentrations. *P. cryohalolentis* K5 was also found to be resistant to kanamycin and tetracycline at concentrations less than 5 µg/mL of media.

After conjugation, single colonies could not be obtained in some of the strains when selected with specific antibiotics. Strains grown on tetracycline at concentrations of 15 µg/mL or greater formed slimy films rather than single colonies. *P. arcticus* 273-4 formed distinct colonies when pBBR1MCS2 (Kanamycin resistant) was genetically transferred into it and two of these colonies, colony 20 and 21, were isolated for confirming the integrity of the plasmid present *in vivo*.

The two isolates were grown in either LB liquid cultures or LB plates, both containing 10 µg/mL Kanamycin. Total genomic DNA was extracted from each culture and PCR was performed with primer pairs specific to a region of the pBBRMCS2 plasmid. While the expected size of the PCR product for primer pairs BBP828 - BBP829 was 852 base pairs (bp), the size of the PCR product obtained from Colony 21 was approximately around 6000 bp. Colony 20 produced the correct sized PCR product. We

were concerned by the discrepancy in PCR product size for the same plasmid region between the two colonies. We hypothesized that this change in size could be the result of an integration event in which the introduced plasmid had inserted itself into the genome of the organism. To further understand the fate of the introduced plasmid, a number of primer pairs were designed that covered different regions of the circular plasmid. Agarose gel resolution of PCR products obtained from these primers seemed to indicate that both the colonies 20 and 21 harbored circular pBBR1MCS2 plasmids within them but these plasmids had increased from their size of 5144 bp by approximately 5000 - 6000 bp.



Primer Pairs	Expected PCR product size (in basepairs)
BBP828 & BBP829	852
BBP828 & BBP840	2125
BBP829 & BBP841	2041
BBP852 & BBP853	1780
BBP840 & BBP841	3314

Figure 2.1: Map of plasmid pBBR1MCS2 indicating the different elements of the vector. Annealing sites of various primers and their direction of extension are indicated. The table displays the expected sizes of the PCR products for the different primer pairs used in this study.

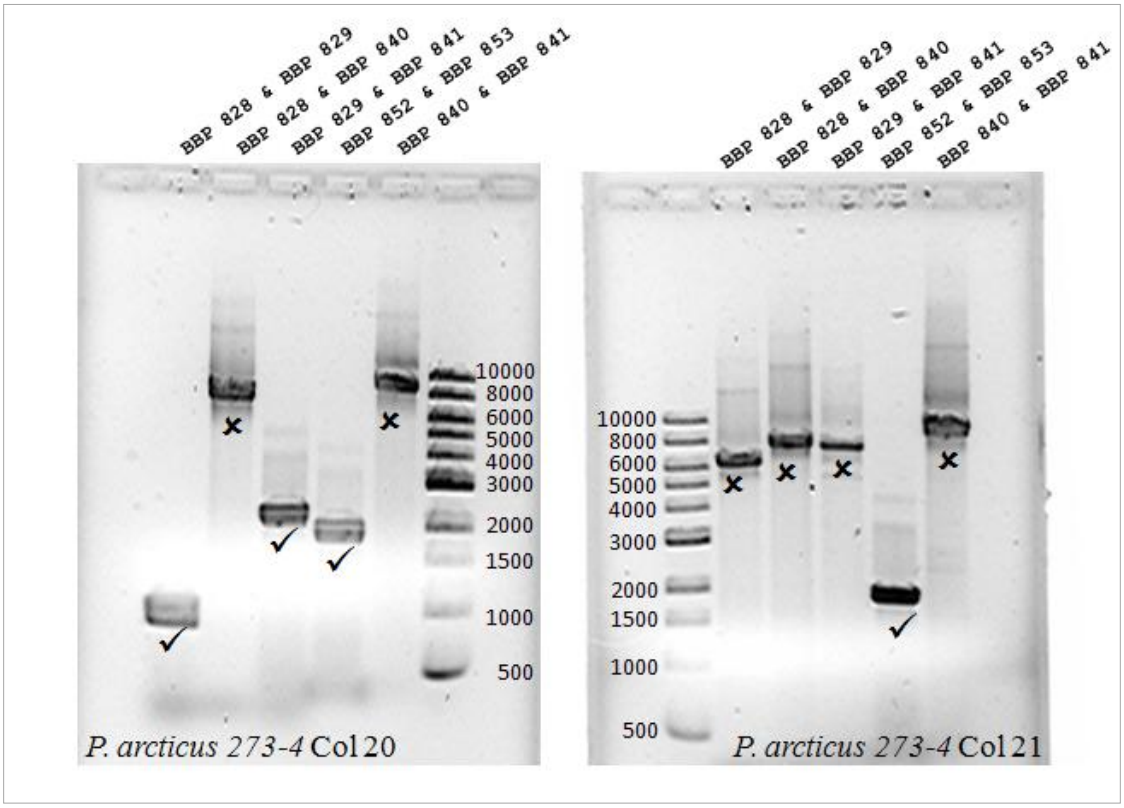


Figure 2.2: PCR based analysis of pBBR1MCS2 plasmid obtained from colonies of recombinant *P. arcticus* 273-4. The symbol (✓) indicates that the observed PCR product is the same size as the expected PCR product for the given primer pair. The symbol (×) indicates mismatch between observed and expected PCR product sizes.

Positive control (Stock plasmid, as received from source) and negative control (wild type *P. arcticus* 273-4 genomic DNA) were used to verify that the PCR products were specific to the plasmid pBBR1MCS2 (data not shown)

2.3.2 Identification of Transposon Insertion in pBBR1MCS2

We then partially sequenced the PCR products whose sizes were greater than the expected size. We found that some stretches of the sequence were not found to match the sequence of the pBBR1MCS2 plasmid. The sequences were then matched against a database of known nucleotide sequences using the BLAST algorithm made available by NCBI. The best results matched that of transposon sequences found in a wide variety of

bacterial species. Primers from BBP910 to BBP913 were used in primer walking techniques to obtain more sequence information on the inserts within pBBR1MCS2, from which we confirmed that the nearly 6000 bp insert was a transposon genomic element that had integrated itself into pBBR1MCS2.

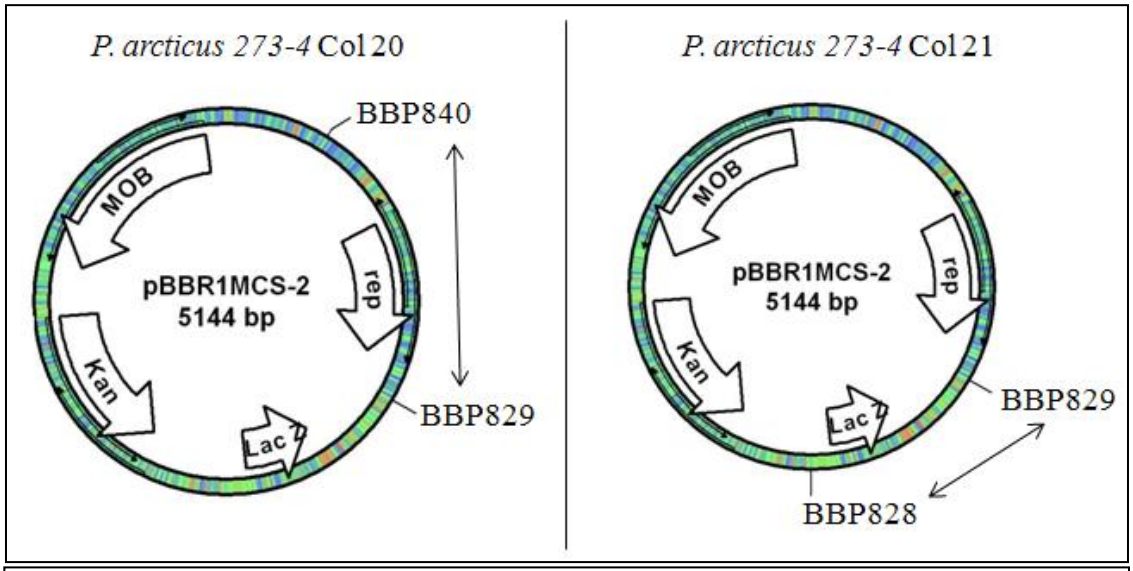


Figure 2.3: Location of transposon inserts as deduced from PCR analysis of plasmids extracted from colonies of *P. arcticus* 273-4 containing the foreign plasmid pBBR1MCS2. The transposon was present between the binding sites of the indicated primers.

Transposons are mobile DNA elements that are able to move from one bacterial organism to another through the various gene transfer mechanisms present in bacteria. They randomly integrate in the genome of the new host and seek to enter new hosts when the current host bacteria are involved in gene transfer, thus ensuring their survival. In order to identify the source of the transposon integrated into the pBBR1MCS2 plasmid in *P. arcticus* 273-4, PCR was performed with primers specific to the transposon sequence on the genomic DNA extract of *E. coli* WM3064 and wild type *P. arcticus* 273-4, both

without any recombinant plasmid in them. A PCR product was obtained only for the *E. coli* WM3064 strain and not for *P. arcticus* 273-4. Thus we found that the transposon DNA element was present in the genome of the *E. coli* WM3064 strain and had integrated itself into the pBBR1MCS2 plasmid when transformed with the vector. The modified plasmid had then been transferred to *P. arcticus* 273-4 through the mechanism of conjugal transfer.

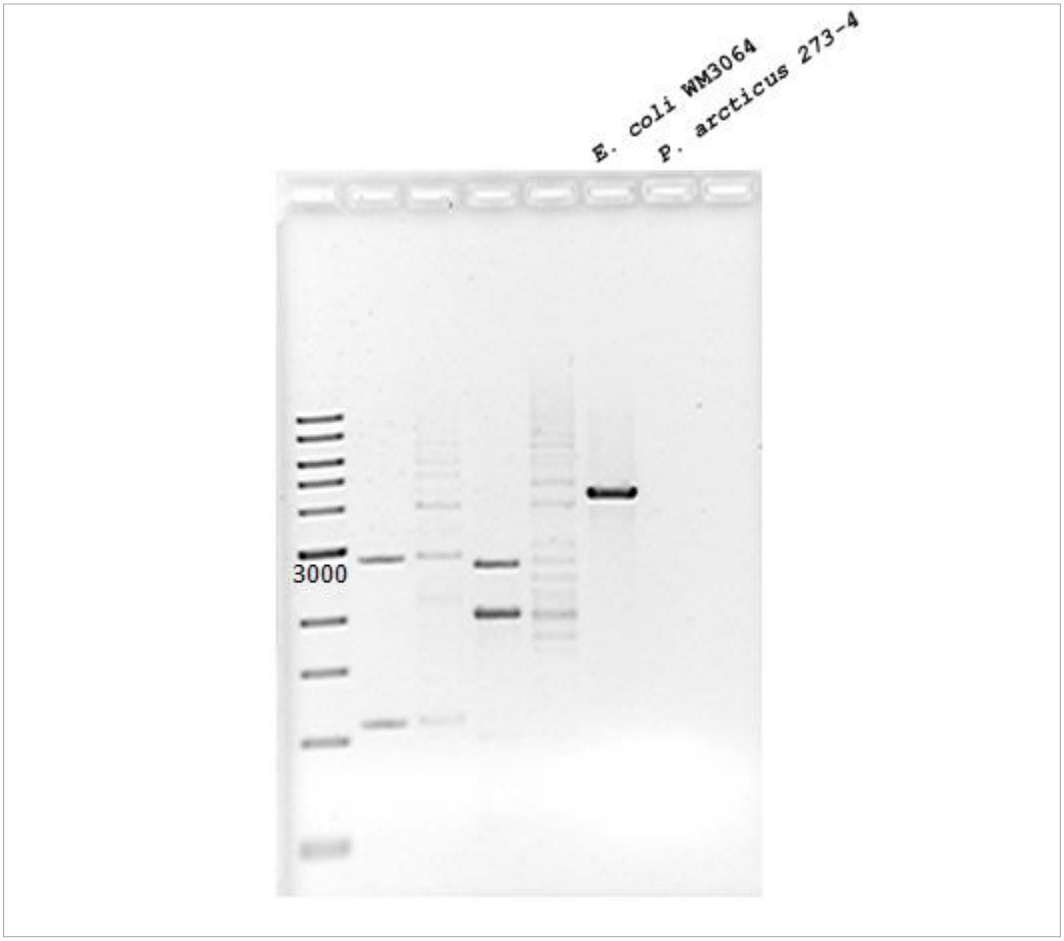


Figure 2.4: Determination of the source bacterial strain containing the transposon. Genomic DNA was extracted from wildtype cells of *E. coli* WM3064 & *P. arcticus* 273-4 and subjected to PCR with the transposon sequence specific primer pair BBP910 – BBP911. A PCR product was obtained only with the genomic DNA of *E. coli* WM3064, which proved that this was the strain that carried the transposon sequence in its genome.

2.3.3 Construction of a Plasmid with Immunity against Transposon Insertion

A literature search for methods to mitigate transposon insertion in DNA sequences revealed the report of Lee and others (Lee et al., 1983), that discussed the role of a 38 base pair DNA sequence which if present within a DNA sequence, could confer immunity to that stretch of sequence from transposon insertion. This 38 bp sequence was a recognition site for one of the transposon proteins that regulates transposon DNA insertion into other DNA strands. Presence of just the recognition site was enough to prevent insertion of the entire transposon element. Furthermore, sequencing and analysis of the transposon-inserted plasmids from colonies 20 and 21 of *P. arcticus* 273-4 revealed the presence of the exact 38 bp transposon recognition sequence within the transposon region.

```
5′- CTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCC - 3′
3′- GAATTGCACTCAAAAGCAAGGTGACTCGCAGTCTGGGG – 5′
```

Figure 2.5: The 38 bp sequence that was reported by Lee et al., (1983) to be a part of the Tn3 transposon. This sequence was also found in the plasmids recovered from *P. arcticus* 273-4 colonies 20 & 21. Addition of this sequence to the pBBR1MCS2 plasmid prevented insertion of the transposon sequence into the vector from *E. coli* WM3064

The pBBR1MCS2 plasmid did not contain this sequence. An analysis of pUC and pET derived vectors available in the laboratory identified the presence of the 38 bp sequence in each of them. Introducing those vectors into *E. coli* WM3064 and testing re-isolated plasmids indicated that the vectors were of the expected size, with no change in

their size. To test if transposon insertion immunity can be conferred to plasmid pBBR1MCS2, the 38 bp recognition sequence was inserted into pBBR1MCS2 to obtain the vector pBB109 Col 1.

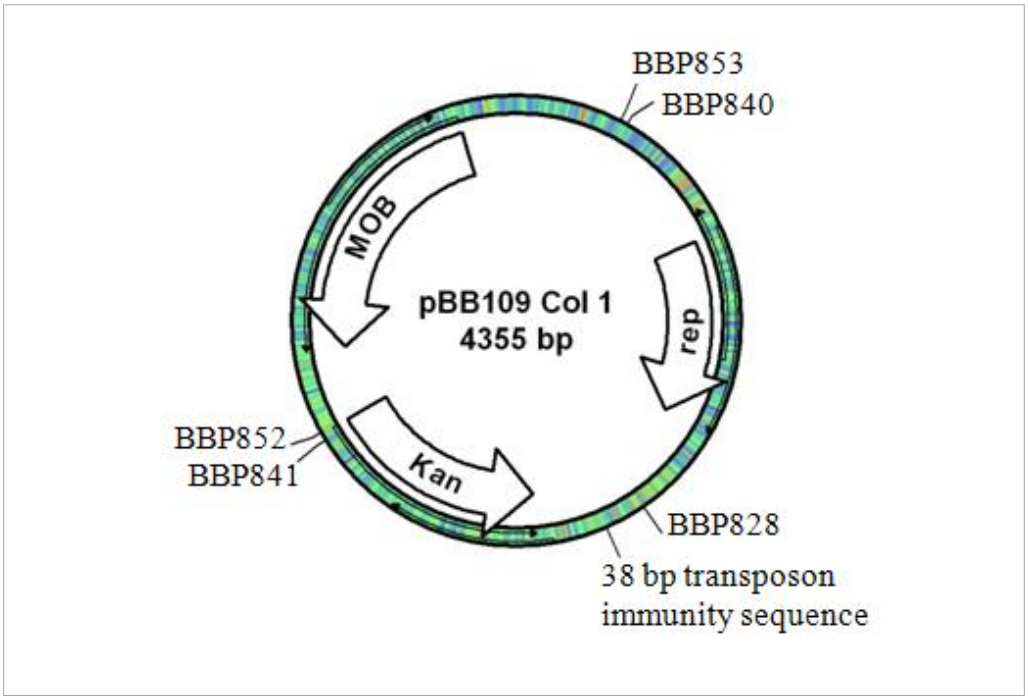
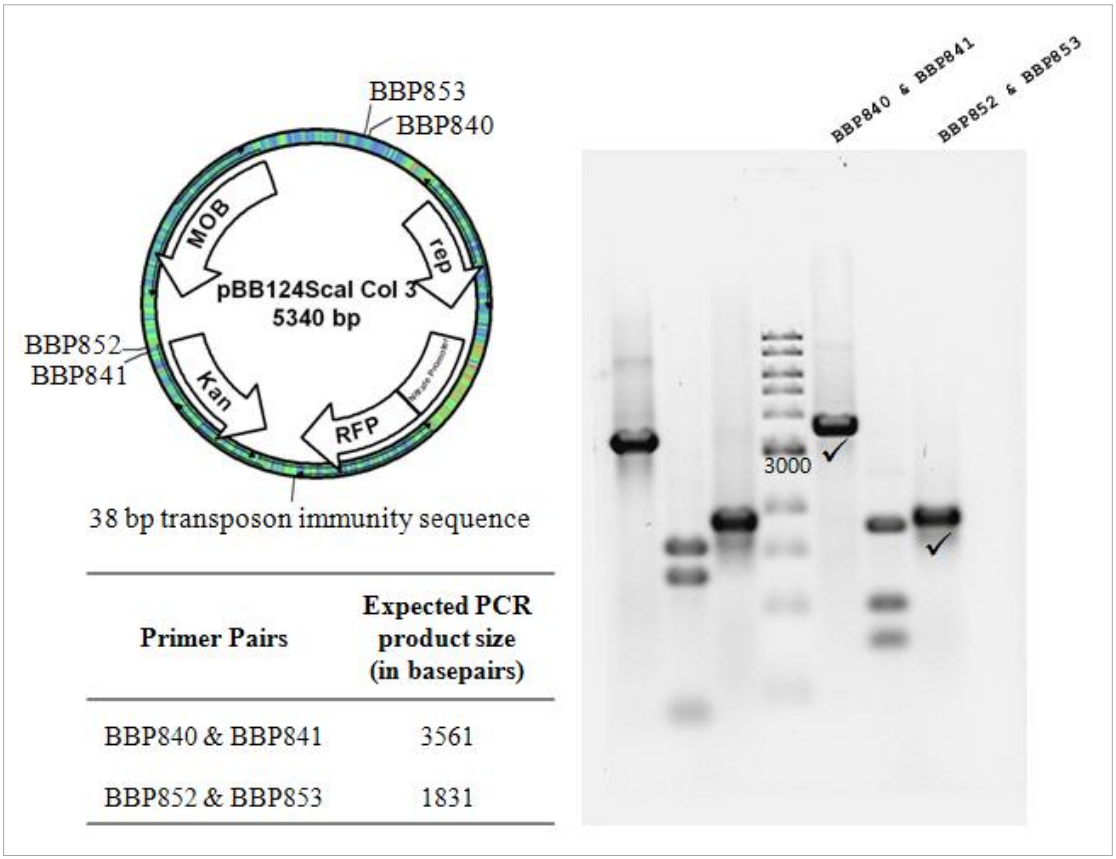


Figure 2.6: Vector pBB109 Col 1, a derivative of pBBR1MCS2 that incorporates a 38 base pair sequence that has been previously demonstrated to prevent transposon insertions in plasmids.

This plasmid was further modified to incorporate the promoter region of a putative nitrate reductase gene present in *P. arcticus* 273-4. The open reading frame of a red fluorescent protein (RFP) was introduced behind the promoter, to test for heterologous RFP expression in the presence of nitrate. The modified plasmid, designated pBB124ScaI Col 3, was then transferred into *P. arcticus* 273-4 through conjugation, using *E. coli* WM3064 as the host strain. PCR analysis of the total DNA from antibiotic-

screened colonies of the target strain indicated the presence of a circular plasmid whose observed size was equivalent to the expected size (5340 bp) of pBB124ScaI Col 3. No transposon insertions were observed.



Primer Pairs	Expected PCR product size (in basepairs)
BBP840 & BBP841	3561
BBP852 & BBP853	1831

Figure 2.7: Vector pBB124ScaI Col 3, a derivative of pBB109 Col 1 was introduced into *P. arcticus* 273-4 through conjugation. PCR amplification was performed using plasmid-specific primers on total DNA extracted from a transformed colony. Agarose gel analysis indicated that the observed PCR products were of the same size as expected for pBB124ScaI Col 3 as a circular plasmid.

2.4 Discussion

The ability to introduce foreign genes and to re-introduce native genes driven by strong, inducible promoters is a key step towards developing recombinant bioprocess systems in wax ester biosynthesizing bacteria. A broad host range plasmid vector was chosen as a genetic carrier for wax ester synthesizing bacteria like *M. aquaeolei* VT8, *P. cryohalolentis* K5 and *P. arcticus* 273-4. The natural bacterial gene transfer system of conjugation was adapted to transfer the plasmid vector from a donor *E. coli* strain into the target bacterial species of choice. While the conjugation process was successful in introducing the plasmid into all of the bacterial species under study, an unexpected phenomenon was observed – the increase in size of the introduced plasmid by approximately 6000 basepairs. This insertion of additional nucleotide sequence might disrupt the integrity of the plasmid vector when it is used as a carrier for a recombinant gene. Therefore, efforts were undertaken to identify the inserted sequence, the organism from which it was inserted and finally, a solution to prevent such an insertion in future genetic transfers.

Using total DNA extracted from colonies of *P. arcticus* 273-4 maintaining the vector pBBR1MCS2, we identified the inserted sequence as a Tn3 transposon, a mobile genetic element that integrates randomly into stretches of DNA. We also deduced the site of transposon insertion in the introduced plasmid through PCR analysis. We then found out that the transposon originated from the *E. coli* strain WM3064 used as a donor strain in the conjugation process. Based on an earlier report, we identified a 38 base pair nucleotide sequence recognized by certain transposon-specific proteins, that when present

in a plasmid, was demonstrated to prevent transposon insertions. We introduced this sequence into the pBBR1MCS2 plasmid using synthetic oligonucleotides. After minimal modifications to the plasmid to allow expression of a reporter gene in *P. arcticus* 273-4, we introduced the derivative plasmid (pBB124ScaI Col 3) into *P. arcticus* 273-4 through conjugation with a donor *E. coli* WM3064 strain. *P. arcticus* 273-4 colonies that were selected on a kanamycin antibiotic screen were subjected to PCR analysis to check if incorporating the transposon immunity sequence prevented insertion of the transposon into the introduced plasmid. The analysis indicated that the plasmid present *in vivo* in the recombinant *P. arcticus* 273-4 colony was present in circular form, without the insertion of any additional nucleotide sequences. Thus, incorporating the 38 base pair transposon immunity sequence seems to have prevented the insertion of the Tn3 transposon present in the genome of donor *E. coli* WM3064 into the plasmid transferred to *P. arcticus* 273-4 through conjugation.

2.5 Future Work

The development of a vector system that can express recombinant genes within wax ester synthesizing bacteria opens up a number of applications including metabolic engineering for modifying the profile of wax esters produced. The pBB124ScaI Col 3 plasmid contains a red fluorescent protein that can act as a reporter of gene expression. RFP should potentially be expressed when the plasmid-containing *P. arcticus* 273-4 cells are grown in a nitrate rich medium, thus indicating the effectiveness of the vector system for heterologous gene expression. Other promoter systems native to *P. arcticus* 273-4 also need to be explored to identify and utilize such systems for driving heterologous or native gene expression.

From the plasmid pBB109, derivatives specific to the different bacterial species studied in the laboratory need to be developed. Each one should potentially possess an organism-specific promoter driving the expression of a reporter gene. Such plasmids should be introduced into *M. aquaeolei* VT8 and *P. cryohalolentis* K5. Colonies of bacteria harboring the plasmid should be analyzed to verify integrity of the introduced vectors and also to verify if reporter gene expression is being driven by the selected promoter system. If the pBB109 system is modified into a vector that can integrate into the target strain's genome through homologous recombination, it would serve to create useful gene knockouts in wax ester biosynthesizing bacteria. Also, this system would be useful to integrate multiple foreign genes into the genome of the desired bacterial species, thus stably incorporating multi-enzyme heterologous pathways for bioproduct biosynthesis.

Chapter III

Multi-Gene Constructs to Facilitate Myrcene Biosynthesis in *Escherichia coli*

3.1 Introduction

3.1.1 Isoprenoids: Introduction

Isoprenoids constitute one of the largest classes of natural compounds, with estimates greater than 50,000 (McCaskill & Croteau, 1997). While the extent of the role and distribution of isoprenoids among living organisms is expanding with research advances, isoprenoids of plant origin have gained much attention during the recent years for a number of their uses. Traditionally, plant isoprenoids, biosynthesized as secondary metabolites, have found use in the manufacture of aromatics. They have been essential aroma compounds in essences, fragrances and cosmetic preparations. The discovery of several isoprenoids from traditional herbal medicines led to a new interest in isoprenoids as potential drugs (Newman & Cragg, 2007). Research into identifying such chemicals has led to the identification of isoprenoids like Artemisinin that has anti-malarial activity, Taxadiene (Taxol) that has anti-neoplastic activity and Casbene that possesses potent antifungal action (McGarvey, 1995). Since many of these isoprenoids are biosynthesized in minute quantities in plants, a great deal of effort has been dedicated towards engineering recombinant microbial systems with isoprenoid biosynthetic capabilities (Ajikumar et al., 2010; Leonard et al., 2010; Martin et al., 2003).

Isoprenoids, also referred to as terpenoids, are biosynthesized from the common precursor molecules Isopentenyl diphosphate (also referred to as Isopentenyl pyrophosphate or IPP) and Dimethylallyl diphosphate (also referred to as Dimethylallyl pyrophosphate or DMAPP). These two precursors are five carbon molecules and are isomers. Isoprenoid formation occurs by the successive addition of a precursor molecule,

usually IPP, to a molecule of DMAPP. Thus isoprenoids are classified based on the number of C5 subunits that they contain. The simplest isoprenoid is isoprene, formed by the dephosphorylation of IPP. It is a five carbon molecule that is a constituent of natural rubber. Monoterpenes are C10 compounds that are formed from combining a single molecule each of IPP and DMAPP. Many monoterpenes are then cyclized and oxidized to produce a variety of common aroma chemicals like limonene (aroma of citrus) and camphene (aroma of camphor). Some cyclized monoterpenes like pinene have also been explored as missile fuel due to their favorable thermal properties at high altitudes (Harvey et al., 2010). However, linear monoterpenes also exist, like myrcene and ocimene. Non-oxidized linear monoterpenes like myrcene and its isomers are classified as olefinic hydrocarbons. Sesquiterpenes are C15 compounds, with farnesene being the prototypical sesquiterpene. Diterpenes are C20 chemicals that are mostly relevant as biologically important antimicrobial compounds. Of the higher terpenes, triterpenes (C30) have been studied extensively since they encompass the steroid group of chemicals.

With increasing interest and investments in developing renewable sources of energy to replace fossil oil, we became interested in developing a biosystem that was capable of producing a broad range of fuel molecules and had a competitive nutrient consumption cost. Isoprenoids were a clear choice in terms of biosynthetic flexibility since they all share the same precursor molecules (IPP and DMAPP), and the size of the fuel molecules can be varied according to needs. We realized that for our initial studies on exploring terpenoid biosynthesis in a novel host, we required an easily recoverable

terpenoid fuel molecule to make product recovery trivial. We found that monoterpene hydrocarbons, which are volatile in nature, will have the least issues in terms of product recovery. As a prototypical monoterpene hydrocarbon, myrcene fit our requirements and was decided as the target molecule for biosynthesis.

3.1.2 Myrcene Biosynthesis

Myrcene is found in number of plants including hops, lemon grass, cannabis etc. It is used widely in the perfume industry to produce fragrant alcohols. However, its fuel potential has not been explored until now. Myrcene is formed from the precursor geranyl pyrophosphate (GPP), which is formed from the addition of a molecule of IPP to DMAPP. The myrcene synthase enzyme, a member of the terpene synthase class, removes the pyrophosphate group to produce myrcene. The monoterpene hydrocarbon is volatile and easily permeates through the cell membrane into the atmosphere. This property would be useful to assay and track the hydrocarbon synthesis over a period of time.

Research from other groups has demonstrated the heterologous expression of multi-gene enzymatic pathways leading to monoterpene biosynthesis in laboratory strains of *Escherichia coli* (Carter et al., 2003). The expression level of these enzyme pathways has also been optimized to report high levels of monoterpene production (Reiling et al., 2004). Our laboratory wanted to replicate the effort of the Keasling group (Reiling et al., 2004) independently, to develop in-house capabilities in multi-gene pathway expression. We reasoned that reconstructing this pathway would help us in learning the practical

difficulties faced with complex metabolic engineering projects. Since the prior results were made publicly available through peer-reviewed research, we felt that our independent efforts could be validated with results from the published research. We selected *E. coli* as the test platform for construction and validation of the myrcene biosynthesis pathway.

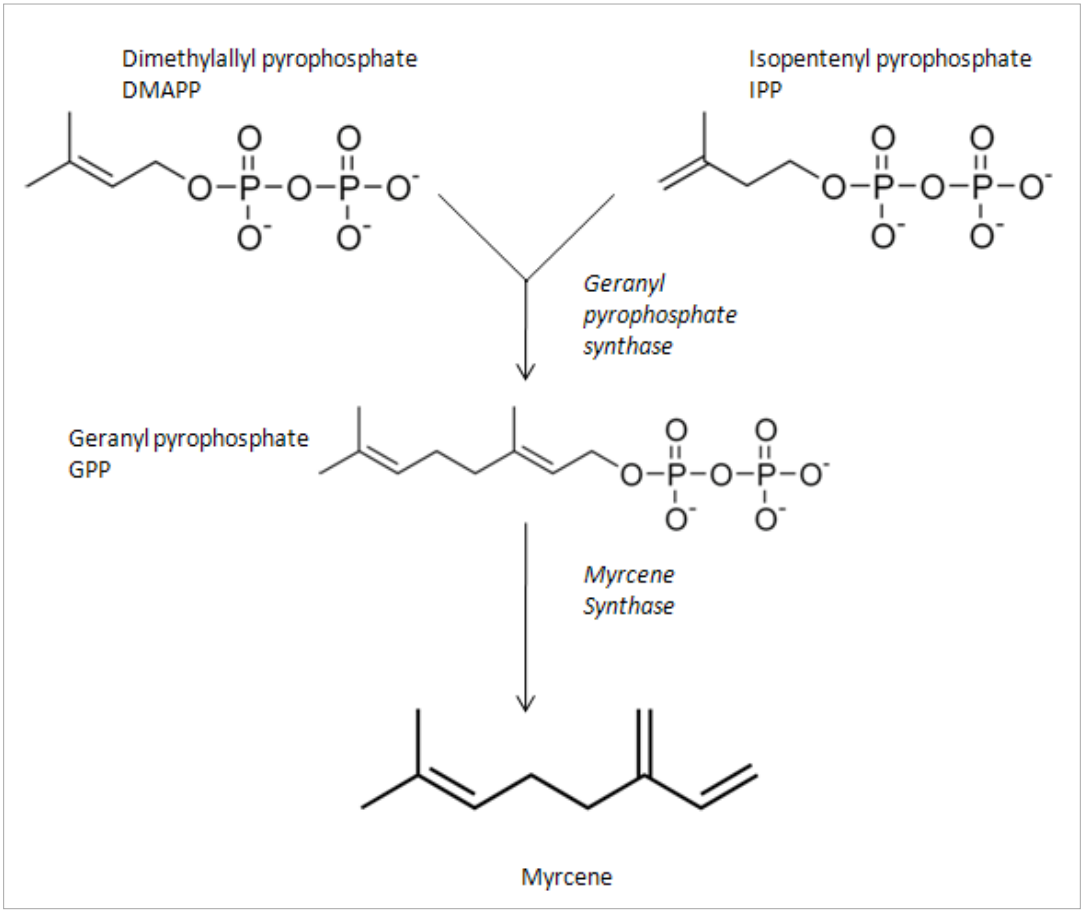


Figure 3.1: Enzymatic pathway for biosynthesis of monoterpene hydrocarbon Myrcene from isoprenoid precursors DMAPP and IPP.
Image adapted from Wikimedia.

3.1.3 *Anabaena variabilis* ATCC29413 – Cyanobacterial Host

As for the production host, we looked at a variety of novel hosts that fit our criteria of having low feedstock costs. We found cyanobacteria to be ideal hosts since they are photosynthetic and fix CO_2, thus eliminating the cost of carbon feed. A number of them have also been successfully genetically engineered (Atsumi et al., 2009; Ducat et al., 2011; Liu et al., 2011) and are being actively explored as recombinant hosts for the commercial biosynthesis of fatty acid based fuels (Berry, 2010). Among cyanobacteria, we were drawn to *Anabaena variabilis* ATCC 29413 (hereafter referred to as *Anabaena variabilis* or *A. variabilis*), a cyanobacterium that is capable of nitrogen fixation apart from photosynthetically fixing atmospheric CO_2.

A. variabilis is a filamentous cyanobacterium that is capable of fixing both gaseous CO_2 and N_2 for its growth. Therefore, its nutrient requirements exclude fixed nitrogen and carbon sources. The organism is known for producing heterocysts, an anoxygenic compartment that can fix gaseous N_2 into ammonia that is then used for growth (Adams & Duggan, 1999). *A. variabilis* has had its genome sequenced, which is important for genetically manipulating it, and also has a well-developed genetic transformation system. Engineered *A. variabilis* has also been demonstrated to be safe in both laboratory and field cultures (Chaurasia & Apte, 2011). Additionally, the organism has a native pathway for synthesis of terpenoids including sesquiterpenes and diterpenes. This pathway can be directed to produce GPP, the precursor of myrcene. These advantages persuaded us to select *A. variabilis* as the host for the myrcene biosynthetic pathway.

3.1.4 Isoprenoid Biosynthetic Pathway in *Escherichia coli*

We chose *E. coli* as the prototype platform for our initial studies before we moved on *A. variabilis*. *E. coli* possesses the MEP/DOXP pathway through which it produces IPP and DMAPP (Rohmer, 1999), which are utilized for the biosynthesis of Ubiquinone, a key component of the cell's electron transport chain. Previous studies have shown that three enzymes encoded by the genes *dxs* (1-deoxyxylulose-5-phosphate synthase), *ispA* (farnesyl pyrophosphate synthase) and *isi* (isopentenyl diphosphate isomerase) were the bottleneck steps that could limit biosynthesis of the prenyl precursor FPP (farnesyl pyrophosphate) synthesized in *E. coli* (Kim & Keasling, 2001; Wang et al., 1999). The genes coding for these enzymes were the targets for elevated *in vivo* expression. Further, *E. coli* does not produce detectable quantities of geranyl pyrophosphate (GPP), the substrate precursor for myrcene. However, mutating the *ispA* enzyme (which synthesizes farnesyl pyrophosphate) at a single amino acid residue has been shown to restrict the synthesis of FPP and instead produce GPP (Ohnuma et al., 1996). To accomplish these targets, we followed the ideas outlined by (Reiling et al., 2004) and planned to develop a vector that expressed the three genes from an inducible promoter system.

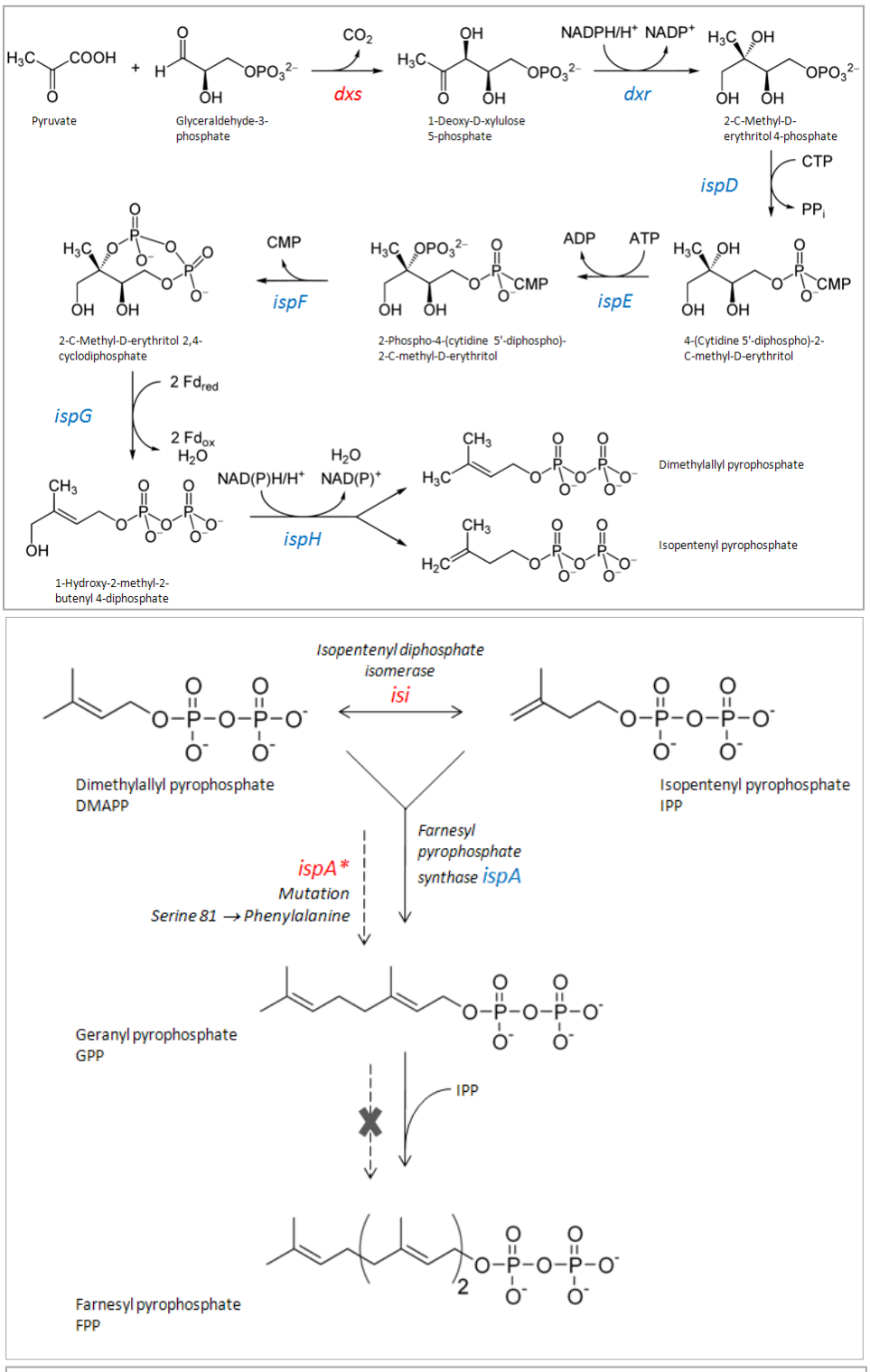


Figure 3.2: MEP/DOXP pathway for prenyl pyrophosphate biosynthesis in *Escherichia coli* K-12 MG1655. Bottleneck enzymes whose genes were over-expressed in a plasmid are indicated in red. Image adapted from Wikimedia.

3.1.5 Gene Design of Myrcene Synthase

The enzyme Myrcene synthase, which converts GPP into myrcene, is not found in *E. coli* or other common laboratory strains of bacteria. While many terpene synthase genes have been characterized from plants through creation of cDNA libraries, the expression of plant terpene synthase cDNA in *E. coli* strains has mostly been at less than desirable levels (Martin et al., 2001). A primary reason attributed to such poor expression levels is the difference in codon usage among different organisms (Gustafsson et al., 2004). Each species of organism has a preferred choice of triplet codon for amino acids encoded by multiple codons. This codon choice varies between organisms and corresponds to factors like the abundance of complementary tRNA of each codon. Therefore, a gene sequence transcribed into an mRNA consisting of rare codons will be translated poorly or at sub-optimal levels (Kane, 1995). Heterologous expression of such genes, characterized by rare codons, requires that such rare codons be replaced by codons better preferred by the recombinant host (Welch et al., 2009).

From research literature, we found a gene for a myrcene synthase enzyme that had high product specificity, unlike the majority of monoterpene synthases that produced an array of monoterpenes including myrcene (Fischbach et al., 2001). This gene was present in the Holm Oak tree and its protein product had been biochemically characterized. Since we intended to express the myrcene synthase gene in a cyanobacterium, *A. variabilis*, we decided to synthesize the gene so that we could optimize the expression of myrcene synthase gene for *A. variabilis*.

3.2 Materials & Methods

3.2.1 Bacterial Strains Used and Growth Methods

E. coli JM109 was used for all gene cloning experiments. *E. coli* JM109, *E. coli* strain BL21 (DE3) or *E. coli* Rosetta were used for protein expression experiments. The strains were grown at 37° C in Lysogeny Broth medium in either plates or test tubes. *E. coli* JM109 and BL21 (DE3) cells transformed with plasmids were maintained in LB medium with either 50 µg/mL Kanamycin or 100 µg/mL Ampicillin, according to the resistance gene present in the introduced plasmid. *E. coli* Rosetta (EMD Chemicals, Gibbstown, NJ) strain was maintained in LB medium containing 35 µg/mL Chloramphenicol antibiotic. Transformed cells of *E. coli* Rosetta were grown on LB plates containing both chloramphenicol and the antibiotic whose resistance gene is present in the introduced plasmid. Test tube cultures were grown in a VWR shaker incubator at 37° C and 225 RPM (VWR, PA). Introduction of plasmids into strains was performed by calcium chloride mediated chemical transformation technique. Long term frozen stocks of strains were maintained at - 80° C in cryovials. Briefly, cells at the exponential growth phase (OD_{600} = 0.6 ~ 0.8) were resuspended in a LB-glycerol solution (containing 20% glycerol by volume) and flash frozen in liquid nitrogen (- 196° C) before being stored at - 80° C.

3.2.2 Gene Cloning

Restriction endonuclease enzymes were obtained from New England Biolabs, Ipswich, MA. Restriction digestions were performed for at least 1 hour at 37° C

according to suggestions from the manufacturer. DNA ligation reactions were performed using T4 DNA Ligase (New England Biolabs, MA) for at least 30 minutes at room temperature. Wherever required, DNA cleanup was performed to remove buffers, free nucleotides or enzymes using DNA Clean & Concentrator kit from Zymo Research, CA. Plasmid extractions were performed using GenElute miniprep kit (Sigma-Aldrich, St.Louis, MO). Site directed mutagenesis of plasmids was carried out using Pfu Turbo Polymerase (Agilent, Santa Clara, CA). All gene cloning experiments were carried out using standard molecular biology protocols. Antibiotic screening and X-gal chromophore based Blue/White screening were used for increasing the odds of selecting desired transformants.

Agarose gel electrophoresis was used to verify the size of plasmids, PCR products and also to troubleshoot ligation reactions. Plasmids were verified for size by initially digesting them with restriction enzymes and then comparing the digested DNA's migration in an agarose gel to that of a standard DNA ladder. Tris-borate-EDTA was used as both the gel-preparation buffer and electrophoresis buffer. Gels contained between 0.5% - 1% of agarose by weight, depending on the expected size of the DNA samples to be run and the amount of separation required between strands for unambiguous resolution. Gels were run at 65V for 2 hours and 15 minutes. Gels were stained in dilute ethidium bromide solution and visualized using a gel documentation system (GelDoc-It Imaging system, UVP LLC, CA).

The software pDRAW32 (Acaclone Software, http://www.acaclone.com/) was used for creating and maintaining digital constructs of all cloning operations and resulting

plasmids. The software was also used to simulate restriction digestion patterns in agarose gel electrophoresis, as a design step before performing actual restriction analysis. The 'Tm Calculator' module of the software was used to design primers for the various polymerase chain reactions described herein.

3.2.3 Recombinant Protein Expression

Test tube cultures of recombinant *E. coli* BL21 (DE3) containing pET derivative vectors were used for small scale protein expression studies. A single colony of the recombinant strain was isolated from a clean LB plate and grown overnight at 37° C in a test tube containing 8 mL of LB medium, along with appropriate antibiotic. From the overnight culture, 100 µL of culture was transferred into fresh medium and grown at 37° C. After 5 hours of growth, at OD_{600} ~ 0.8, 10 µL of 25 mg/mL IPTG was added as inducer. The cells were then placed back in the incubator. After 3 hours of IPTG induction, cells were harvested in 2 mL microcentrifuge tubes by centrifuging (Centrifuge 5424 - Eppendorf AG) for 1 min at greater than 20,000 g and room temperature. The cell pellets were stored at - 20° C for later use.

Stored cell pellets were resuspended in 300 µL of 1% SDS (sodium dodecyl sulfate) solution. The resuspended cells were lysed by sonication (Misonix SL-2000, Qsonica LLC, CT) at power level 15 for 30 seconds while placed on ice. The lysed cell material was then centrifuged at greater than 20,000 g for 4 minutes at room temperature. Meanwhile, an equivolume mixture of 2-mercaptoethanol and Laemmli buffer was prepared. 50 µL of the supernatant from the centrifuged cell lysate was mixed well with

50 µL of the buffer mix in a microcentrifuge tube and placed in boiling water for 10 minutes.

SDS-Polyacrylamide gels (Simpson, 2006) were used with a Bio-Rad Mini-PROTEAN Tetra System for resolving the protein samples on the gels. 30 µL of the prepared sample was loaded into a single lane and a protein ladder was added as a size marker to a separate lane. Gels were run at 165V for 50 minutes in a Tris-Glycine-SDS buffer. Staining was performed using an acidic Coomassie blue solution and destaining was done with a methanol-acetic acid solution. Subsequently, gels were imaged with a white background using a digital camera (Canon PowerShot SD1200 IS).

Table 3.1: List of plasmid vectors developed towards myrcene biosynthesis in *E. coli*

Plasmid	**Source Plasmid**	**Antibiotic resistance**	**Genes contained**
pPCRISPDSX6 Col 1	pUC19	Ampicillin	*ispA*^{wt}, *dxs*
pPCRISPDSX7 Col 3	pUC19	Ampicillin	*ispA* ^{S81F}, *dxs*
pETISPDSX7 Col 1	pETBlueKan Col 3	Kanamycin	*ispA* ^{S81F}, *dxs*
pPCRISPDSX8 Col 1	pUC19	Ampicillin	*ispA* ^{S81F}, *dxs*
pPCRISPDSX9 Col 1	pETBlueKan Col 3	Kanamycin	*ispA* ^{S81F}, *dxs*
pPCRISI1 Col 1	pUC19	Ampicillin	*isi*
pPCRISI2 Col 2	pBB052	Kanamycin	*isi*
pPCRISI3 Col 1	pUC19	Ampicillin	*isi*
pPCRISI4 Col 1	pETBlueAmp Col 3	Ampicillin	*isi*
pPCRISPDSX10 Col 2	pUC19	Ampicillin	*ispA* ^{S81F}, *dxs, isi*
pPCRISPDSX11 Col 1	pETBlueKan Col 3	Kanamycin	*ispA* ^{S81F}, *dxs, isi*
pMK-MyrSyn4	pUC derivative	Kanamycin	*Myrcene Synthase*
pETMyrSyn Col 1	pETBlueAmp Col 3	Ampicillin	*Myrcene Synthase*

Table 3.2: List of primers used in construction of plasmid vectors towards myrcene biosynthesis in *E. coli*. Restriction enzyme recognition sites are underlined. Starred primers were used for sequencing the plasmids.

Primer	**Sequence**
BBP717	GAGTGTATCCACGCTTACT*TC*TTAATTCATGATGATTTAC
BBP718	GTAAATCATCATGAATTAAGAAGTAAGCGTGGATACACTC
BBP874	GACA*GGATCC*TAAGGAGGTGATCTGATGAGTTTTGATATTGCCAAATACC
BBP875	GTCTA*AGATCTGGTACC*TTATTTATTACGCTGGATGATGTAGTC
BBP925	AT*CCATGGAATTC*ATAAGGAGGTTGATCATGCAAACGGAACACGTCATT
BBP926	ATA*GGTACCTGCAGGATCC*TTATTTAAGCTGGGTAAATGCAGATAATCG
*BBP129	GGCTCGTATGTTGTGTGGAATTGTGAGCGG
*BBP159	GGTCACAGCTTGTCTGTAAGCGGATGC
*BBP703	GACCGTCAGCACCAACAATG
*BBP804	GTACTCGACAAGTATGCAGAGAGCATC`
*BBP275	GGAATTGTGAGCGGATAACAATTCCCCTCT

3.3 Results

3.3.1 Construction of the *ispA* S81F – *dxs* operon

The contiguous genes *ispA* and *dxs* were previously cloned from *E. coli* JM109 into a pUC based vector by Dr. Brett Barney. The vector containing these genes was designated as pPCRISPDSX6 Col 1. The *ispA* prenyltransferase gene leads to the formation of the fifteen carbon farnesyl pyrophosphate which is a precursor for sesquiterpenes. However, monoterpenes require the ten carbon geranyl pyrophosphate (GPP) as a precursor. In order to shift the product specificity of the *ispA* prenyltransferase towards GPP, site-specific mutagenesis was performed on pPCRISPDSX6 Col 1 to mutate the Serine (S) 81 of the *ispA* protein into a Phenylalanine (F). A polymerase chain reaction was performed with primers BBP717 & BBP718, using Pfu Turbo polymerase to obtain plasmid-length circular DNA that had the desired mutation. To remove the pPCRISPDSX6 Col 1 template plasmid, the reaction mixture was digested overnight using the enzyme DpnI at 37° C. DpnI specifically recognizes the DNA sequence 5'-GATC -3' and cleaves the site only if the DNA is methylated. The template plasmid, purified from *E. coli* JM109, is methylated *in vivo*. However, the DNA strands formed during the PCR process are formed from unmethylated nucleotides. Hence, digestion with DpnI selectively cleaves the template plasmid without affecting the newly formed DNA which contains the site-specific mutation. This digestion step is necessary since both the template and the mutated plasmid differ only by one codon and hence, screening for the mutated plasmid can become an extremely time consuming process. The digested mixture containing the mutated plasmid was transformed into *E. coli* JM109, selecting on

ampicillin antibiotic. Four random colonies were picked, grown overnight and plasmids extracted from them. Sequencing with primer BBP159 and subsequent analysis identified the plasmid pPCRISPDSX7 Col 3 as the plasmid with the desired S81F mutation in the *ispA* gene. Further sequencing of the downstream regions was performed to confirm the integrity of the *ispA* S81F *dxs* gene operon.

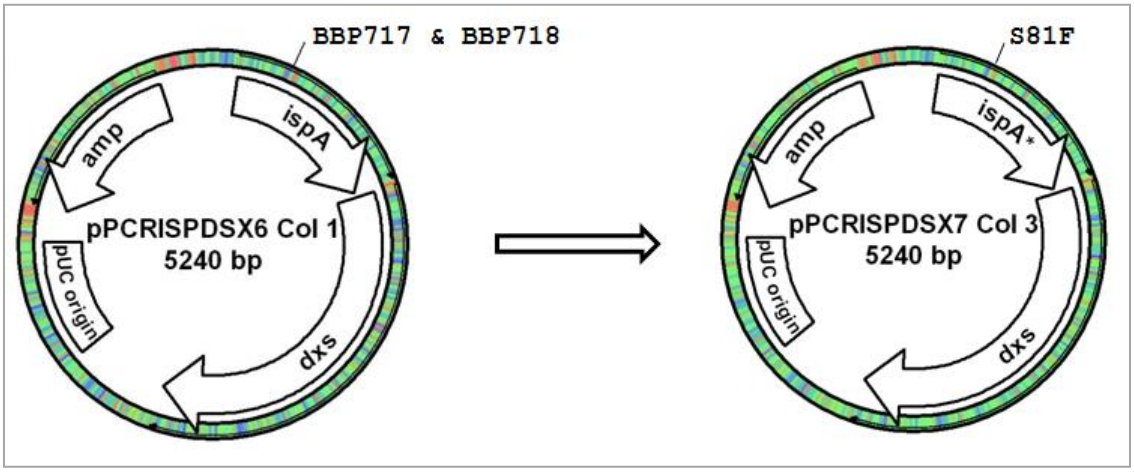


Figure 3.3: Site directed mutagenesis – PCR of pPCRISPDSX6 Col 1. Primers BBP717 & BBP718 incorporated a single amino acid change (S81F) in the *ispA* gene.

The *ispA* S81F gene and the adjacent *dxs* gene were transferred from the pUC based pPCRISPDSX7 Col 3 plasmid into a pET expression system for protein expression. The pET expression system, characterized by the T7 promoter system, has been demonstrated to be superior to the *lac* promoter based expression of the pUC system for high volume protein expression (Studier & Moffatt, 1986). The two gene operon consisting of *ispA* S81F - *dxs* was moved from pPCRISPDSX7 Col 3 into pETBlueKan Col 3 between the restriction sites NdeI - EcoRI. The new plasmid, designated pETISPDSX7 Col 1, contained the *ispA* S81F gene immediately downstream of the T7 promoter with a strong ribosome binding site in-between. The *dxs* gene was downstream

of the *ispA* S81F gene. Protein expression experiments with pETISPDSX7 Col 1 indicated the over-expression of only the *ispA* S81F gene product. The expression of the *dxs* gene could not be differentiated from the wild type control strain which did not contain any plasmid.

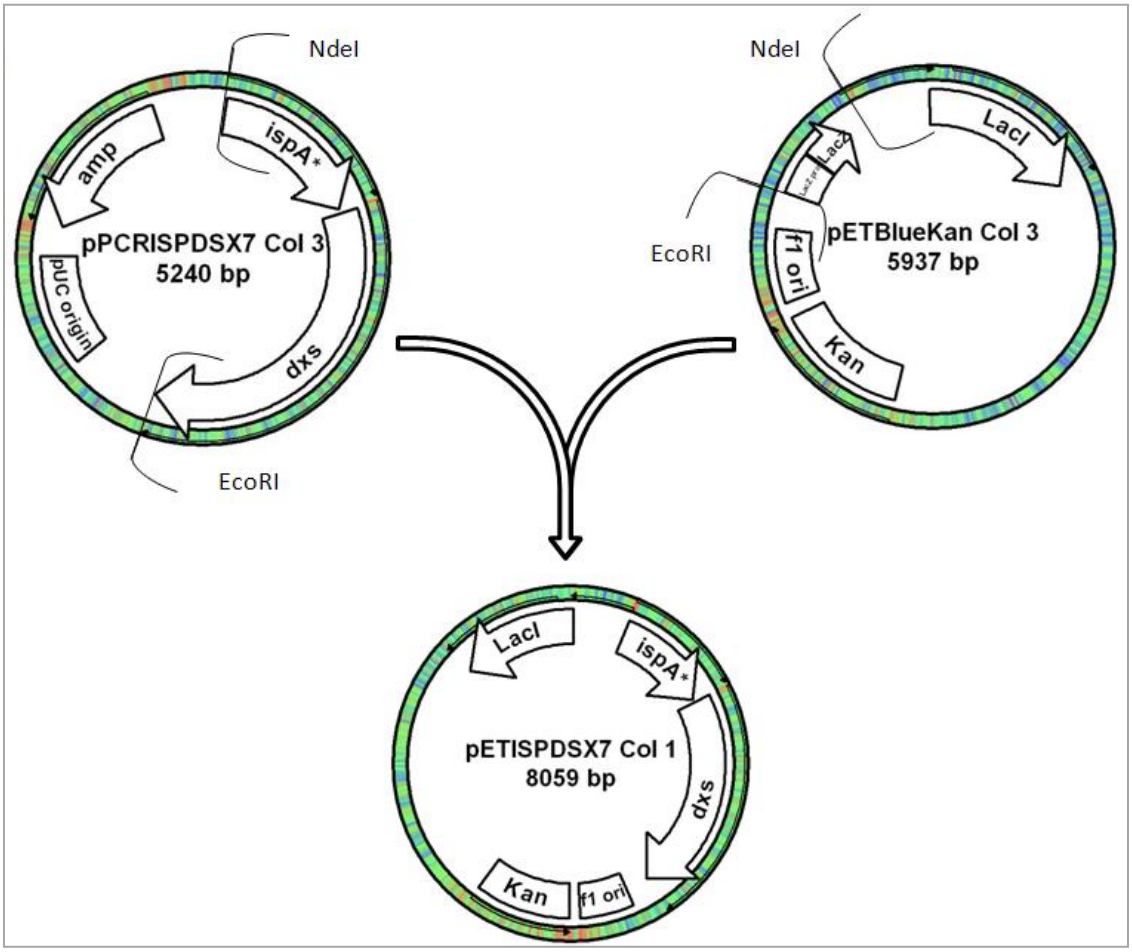


Figure 3.4: Transfer of the *ispA** - *dxs* operon into pETBlueKan Col 3

3.3.2 Incorporation of a Strong Ribosome Binding Site for *dxs*

The previous report on which this work was based on (Reiling et al., 2004) had constructed the *ispA* S81F - *dxs* operon system using genes from different organisms. The authors of the report therefore had to incorporate a ribosome binding site between the two genes to obtain over-expression of both genes. In this work, we had cloned both the genes as a single contiguous sequence from *E. coli* and had expected the natural RBS present in between *ispA* and *dxs* to be of sufficient strength to drive high level expression of *dxs*, given a strong promoter like the T7 system. However, this seemed to not be the case. With available precedent that the sequence of the ribosome binding site is one of the determinants of variation in amount of protein expression, we hypothesized that replacing the natural RBS of *dxs* with a synthetic strong RBS would increase the observed protein expression of *dxs*.

To test this hypothesis, an inverse PCR of pPCRISPDSX7 Col 3 was performed with primers BBP874 and BBP875. The inverse PCR technique uses a circular template and primers that amplify away from each other, unlike a conventional PCR. The resulting product is a linear double strand that lacks the template sequence between the two primer binding sites. This linear strand can then be subjected to restriction digestion and ligation to circularize the DNA. The primers used for the inverse PCR of pPCRISPDSX7 Col 3 were designed in such a way that not only did they eliminate the natural RBS sequence between *ispA* S81F - *dxs*, but also incorporated a new synthetic RBS (5' - AAGGAGGT - 3') between the genes. The new RBS was incorporated between the position -14 to -6 upstream of the start codon (+1) of *dxs*. The linear PCR product was circularized between

the complementary restriction sites of BamHI and BglII. The hexanucleotide recognition sites of these two restriction enzymes differ only in their first and last nucleotide bases and produce a common overhang. Ligation between these two restriction sites forms a hybrid sequence that lacks the recognition sequence of both enzymes, thus eliminating both restriction sites in the circularized DNA. Before the circularization of the linear DNA, the end product of the polymerase chain reaction was digested with DpnI to remove the methylated template plasmid since both the template and desired PCR product would be carrying the same antibiotic resistance gene used for screening. Transformation, subsequent screening and sequencing of a number of random colonies resulted in the identification of the plasmid pPCRISPDSX8 Col 1, that had the new RBS replacing the native RBS between the *ispA* S81F and *dxs* genes.

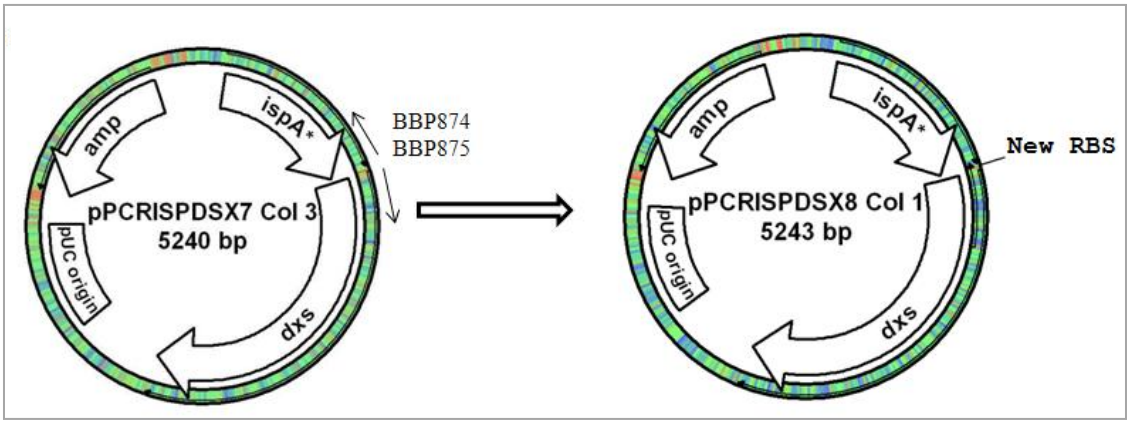


Figure 3.5: Inverse PCR of pPCRISPDSX7 Col 3. Primers BBP874 & BBP875 incorporated a synthetic ribosome binding site upstream of the *dxs* gene.

To test whether the newly incorporated RBS indeed improved the expression of *dxs* gene, the *ispA* S81F - *dxs* operon was moved from pPCRISPDSX8 Col 1 plasmid to pETBlueKan Col 3 plasmid between the restriction sites NdeI - EcoRI, with selection for the new plasmid being done on kanamycin antibiotic. Random colonies were screened for a plasmid of the correct restriction pattern as when digested by select restriction enzymes. Subsequently, a plasmid of the correct size was identified and designated as pPCRISPDSX9 Col 1.

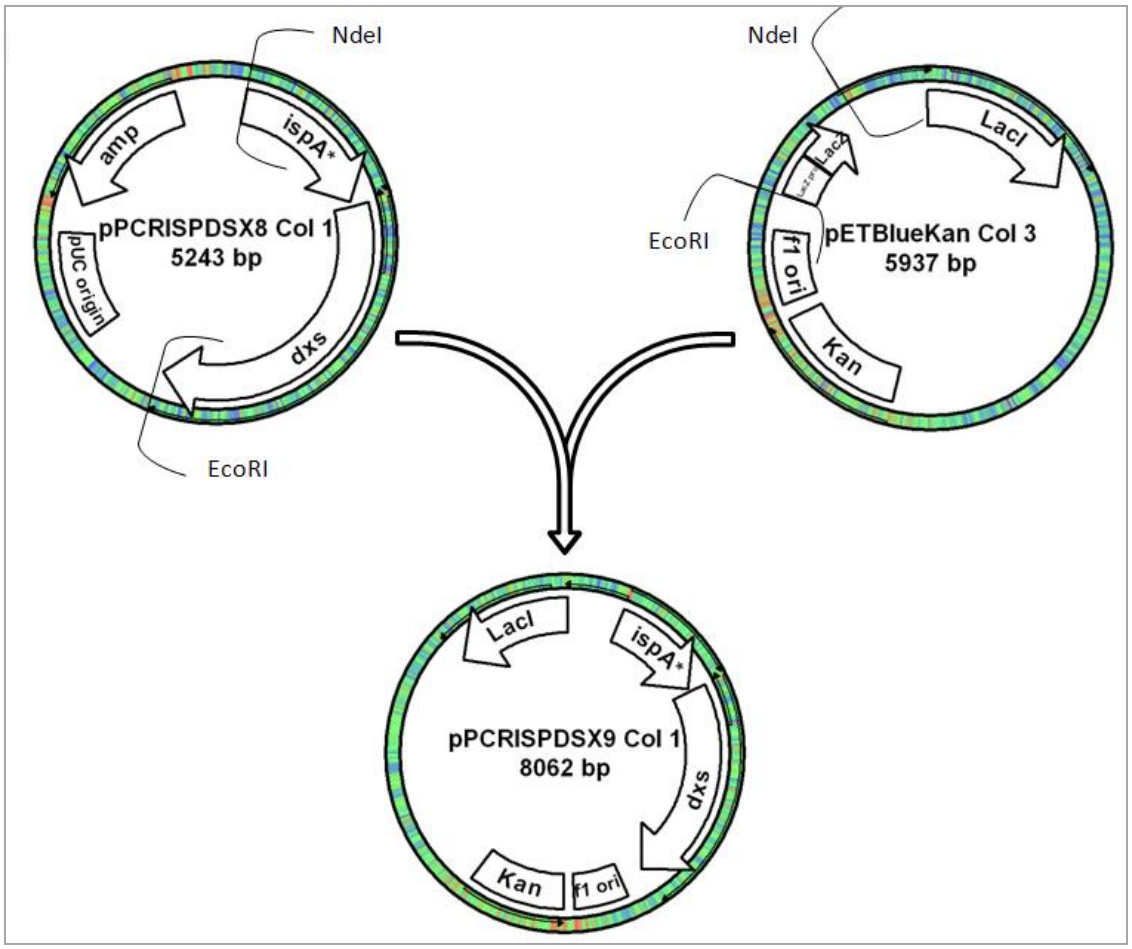


Figure 3.6: Transfer of the *ispA** (S81F) - Strong RBS - *dxs* operon into pETBlueKan Col 3

The pPCRISPDSX9 Col 1 plasmid, which contained the *ispA* S81F - modified RBS - *dxs* system behind a T7 Promoter system, was transformed into *E. coli* BL21 (DE3) strain, with selection being done on kanamycin. A test tube scale protein expression was performed to ascertain whether the new RBS increased protein expression from the *dxs* gene. Performing a SDS-PAGE experiment revealed the presence of a stronger protein band at the location where the *dxs* protein is expected to migrate to in the gel. Thus, this proved that the perceived protein expression of the *dxs* gene was improved by replacing the native RBS with a synthetic strong RBS.

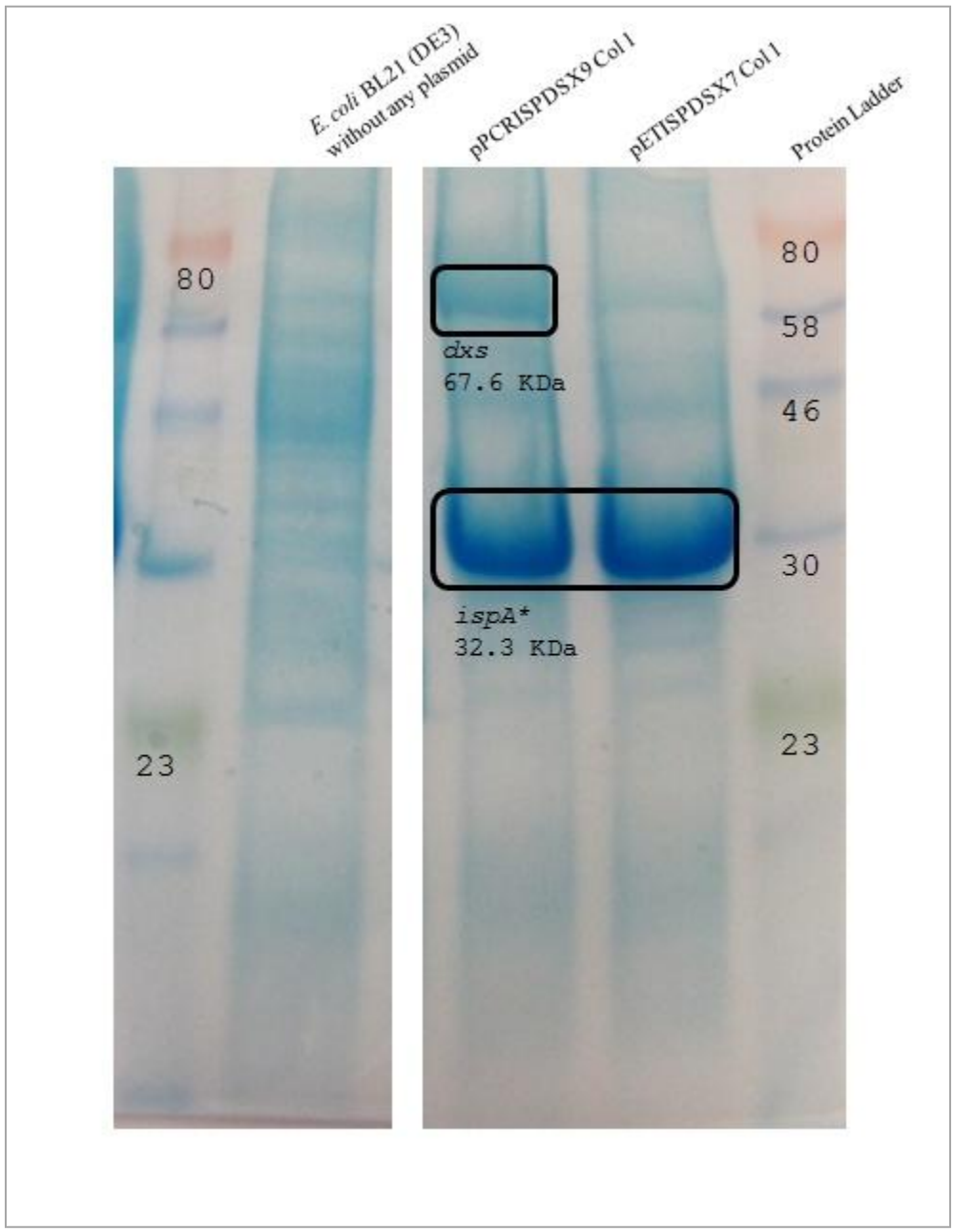


Figure 3.7: PAGE analysis of protein expression from the vector pPCRISPDSX9 Col 1. The new band corresponds to the size of the *dxs* protein and is conspicuously absent in *E. coli* strains with pETISPDSX7 Col 1 plasmid and without any plasmid. The ladder size is in Kilo-Daltons.

3.3.3 Expansion of the *ispA* S81F - *dxs* operon to include *isi*

The isopentenyl diphosphate isomerase (*isi*) is an enzyme that mediates the isomerization of isopentenyl pyrophosphate (IPP) into dimethylallyl pyrophosphate (DMAPP), one of the reactants for *ispA* prenyltransferase. Increased expression of *isi* is expected to increase the availability of DMAPP to *ispA*, thus reducing any bottleneck at this step. In wild type *E. coli*, the *isi* gene exists in a genomic location away from the *ispA-dxs* operon. We wanted to form an artificial operon that had *ispA* - *dxs* – *isi* contiguously, due to the advantages of using a single plasmid and promoter system to drive expression of all three genes. A single plasmid containing all the desired genes reduces the metabolic burden on a recombinant cell when compared to using a separate plasmid for each gene. It is stable, since each separate plasmid must contain an origin of replication that is compatible with the other plasmids. Furthermore, a single plasmid eliminates the use of multiple antibiotics that must be added to select and maintain multiple plasmids types in a cell. Expressing genes as an operon driven by a single promoter has the advantage of regulating multiple genes with a single regulatory system. Also the presence of multiple copies of a promoter system in a plasmid, each responsible for the expression of a downstream gene, can lead to plasmid loss through recombination events due to the high sequence similarity between the promoter copies. It was thus our intention to create an artificial operon containing the *ispA* S81F - *dxs* - *isi* genes that would be regulated from a single promoter system present in the same plasmid.

We decided to introduce *isi* behind the *dxs* gene of pPCRISPDSX8 Col 1. To obtain expression from *isi*, it would be necessary to introduce a RBS upstream of *isi*,

behind *dxs*. The *isi* gene had been previously cloned from *E. coli* JM109 into a pUC derived vector pPCRISI1 Col 1 by Dr. Brett Barney. From this vector, the *isi* gene was amplified by PCR using the primers BBP925 & BBP926 and cloned into a new vector pBB052 between the restriction sites NcoI - KpnI. The forward primer used for the amplification included a RBS sequence (5' - AAGGAGGT - 3') that would be incorporated into the final PCR product as part of the *isi* sequence. Screening for the new vector yielded pPCRISI2 Col 2, which on partial sequencing indicated the presence of the *isi* gene with a RBS upstream of it. Transfer of the RBS + *isi* sequence to pPCRISPDSX8 Col 1 was attempted between the restriction sites EcoRI - PstI. However, screening more than 8 random colonies did not yield the desired plasmid with correct incorporation of *isi*. The presence of a plasmid fragment similar in size to the *isi* DNA fragment, both created by the restriction digestion of pPCRISI2 Col 2 with EcoRI - PstI, undermined gel electrophoresis efforts to resolve a plasmid with the correct insert.

To improve the odds of obtaining a correct sequence in the transfer of the *isi* fragment into pPCRISPDSX8 Col 1, the RBS + *isi* fragment was cloned in the vector pUC19, between the restriction sites NdeI - KpnI to create pPCRISI3 Col 1. From this vector, the desired gene fragment was moved into pPCRISPDSX8 Col 1 between the restriction sites EcoRI - PstI. The plasmid ligation mixture was then transformed in to *E. coli* JM109 cells. Screening random colonies yielded the plasmid pPCRISPDSX10 Col 2 that was verified to contain the RBS + *isi* DNA downstream of the *dxs* gene. Thus the construction of an artificial operon containing *ispA* S81F - *dxs* - *isi* was now complete, with synthetic RBS placed upstream of *dxs* and *isi* genes for driving strong expression of

these genes.

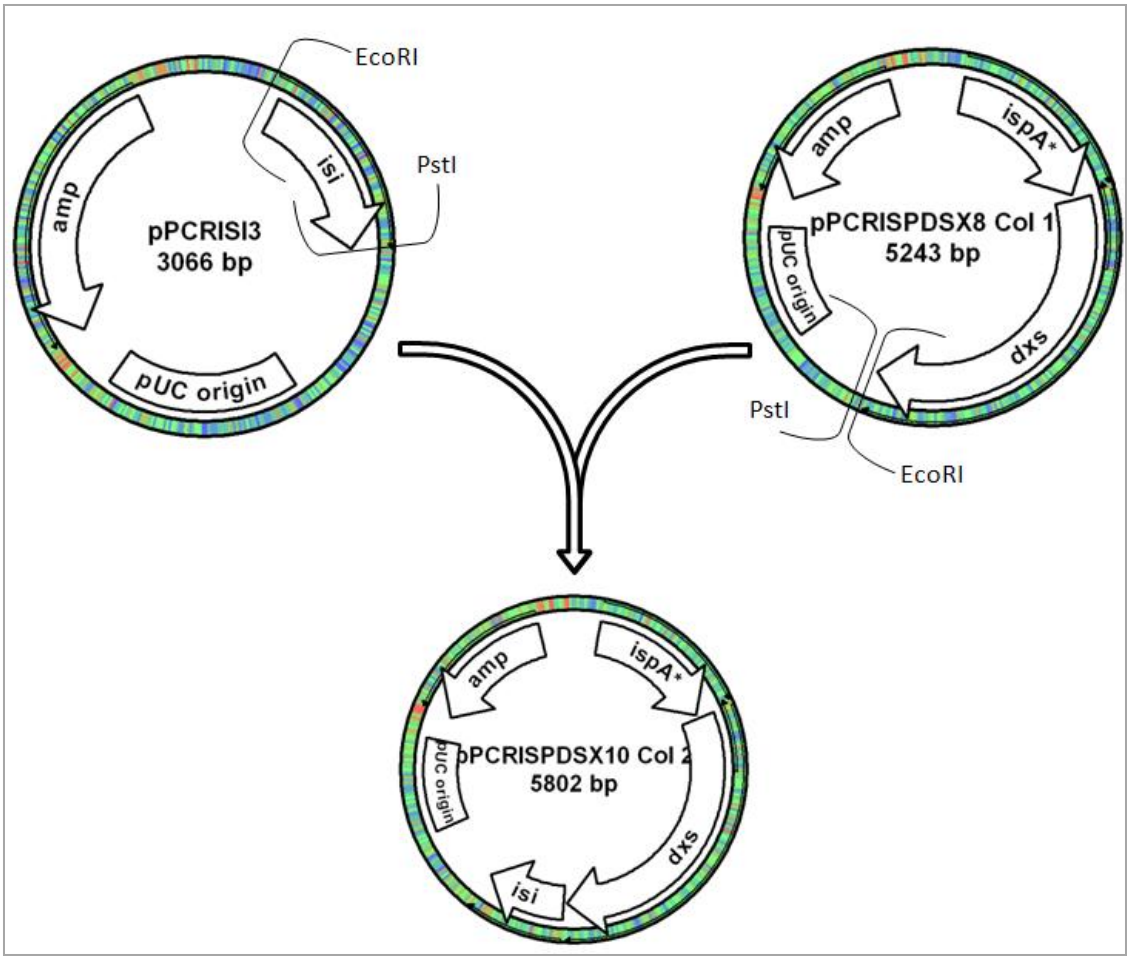


Figure 3.8: Addition of the *isi* gene to the *ispA** (S81F) – *dxs* operon. This step completed the construction of the *ispA* S81F – *dxs* – *isi* synthetic operon. The dxs and isi genes are preceded by non-native strong ribosome binding sites.

The artificial operon needed to be tested for protein expression from all three constituent genes. A pET expression vector, pPCRISPDSX11 Col 1 was constructed by moving the synthetic operon from pPCRISPDSX10 Col 2 to pETBlueKan Col 3 using the restriction sites of NdeI - BamHI. The plasmid was then transformed into *E. coli*

BL21 (DE3). A test tube scale protein expression experiment indicated that all three gene products were expressed in the cell. Meanwhile, the *isi* gene was moved into a pET vector, pETBlueAmp Col 3 from pPCRISI2 Col 2 between the restriction sites NcoI – BamHI to form the vector pPCRISI4 Col 1. This new vector was also subjected to protein expression within the E. coli BL21 (DE3) strain, to serve as a marker for the gel electrophoretic migration of the *isi* protein product.

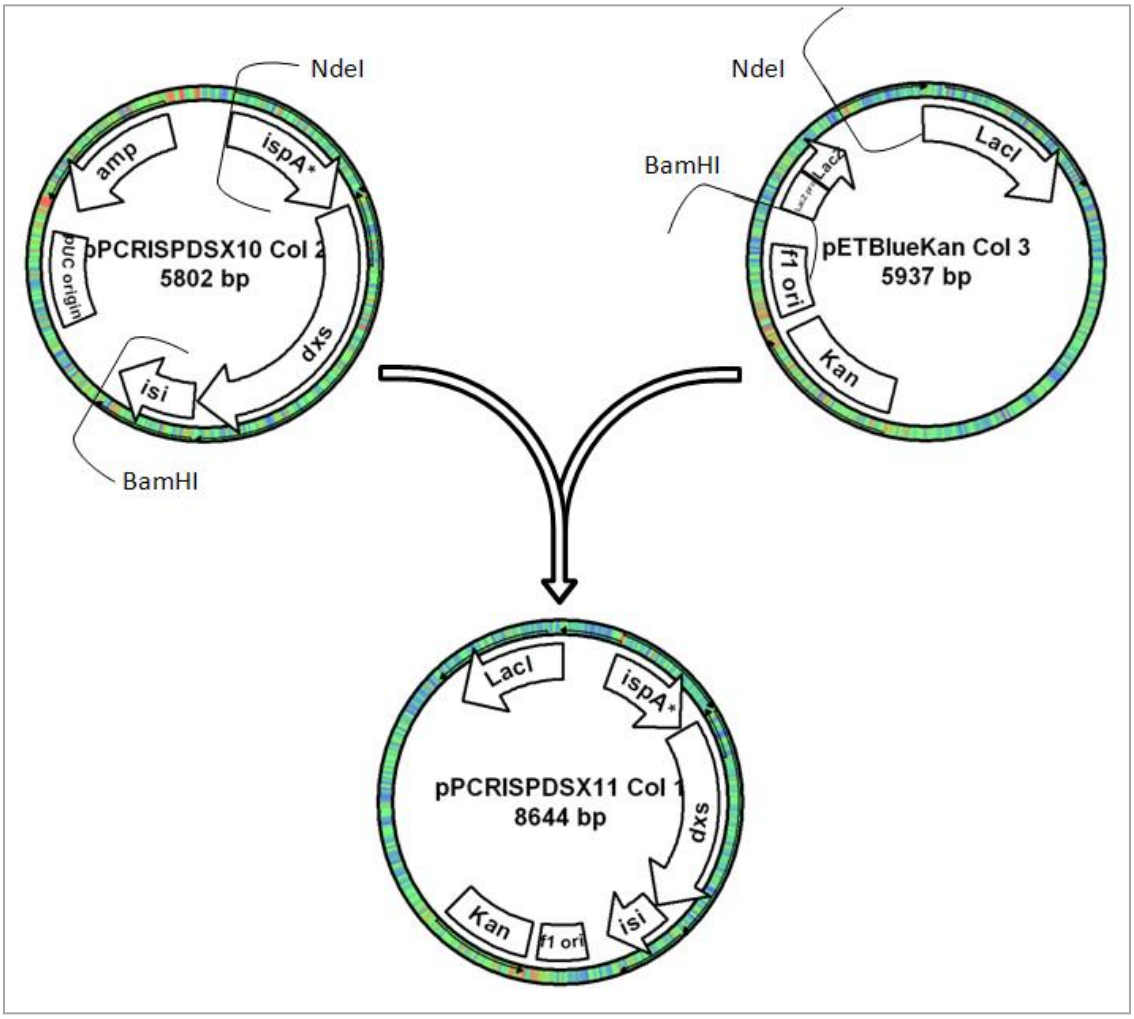


Figure 3.9: Transfer of the *ispA** (S81F) - *dxs* - *isi* operon into pETBlueKan Col 3

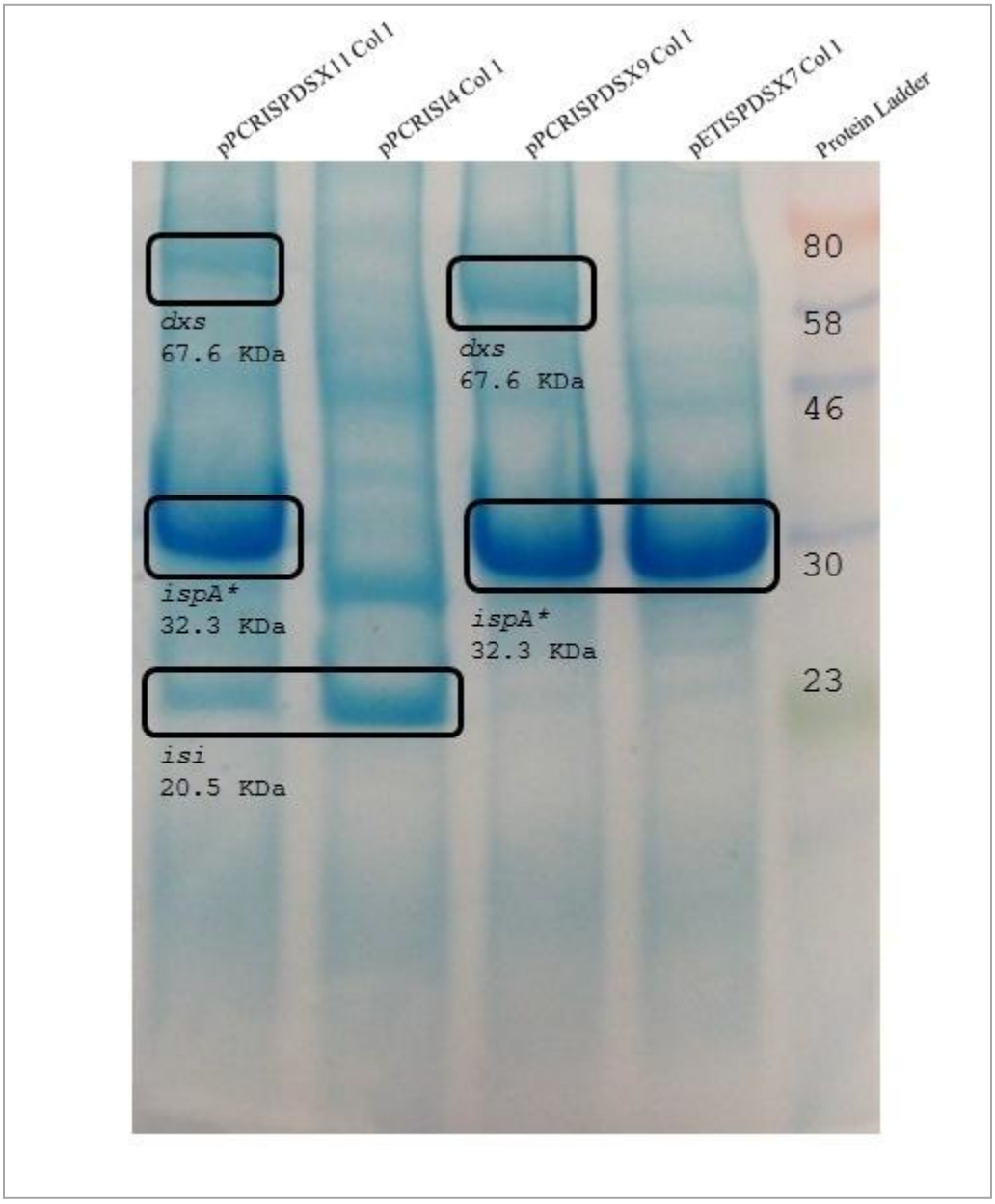


Figure 3.10: PAGE analysis of protein expression from the vector pPCRISPDSX11 Col 1. All three genes of the operon are expressed. The *isi* gene was expressed from a pET vector for comparison. The protein band corresponding to *isi* is absent in the vectors containing only *ispA* – dxs*.

3.3.4 Design Parameters & Synthesis of Myrcene Synthase

The software, Gene Designer 2.0 (DNA 2.0, CA), was used to generate a gene sequence by the reverse translation of the myrcene synthase protein sequence. The reverse translation was performed in such a way that the gene sequence obtained would contain the mean codon distribution of genes in *A. variabilis*. The mean codon distribution frequency for *A. variabilis* was obtained from the Codon Usage Database (www.kazusa.or.jp/codon/index.htm). The gene sequence optimized for *A. variabilis* was presented by the software and the free energy of the corresponding RNA sequence was determined using the tool NUPACK (www.nupack.org). We also wanted to check the expression of the synthesized gene in *E. coli* to test its catalytic activity before introducing into *A. variabilis*. Therefore, we edited certain codons in the artificial gene sequence that were rare in *E. coli* with codons that were more tolerated in both *E. coli* and *A. variabilis*. After such changes, we again determined the free energy of the resulting RNA sequence to make sure that our manual changes to the gene sequence did not negatively affect RNA secondary structure or its free energy. The synthetic gene sequence was then flanked with recognition sites for restriction enzymes, to facilitate cloning of the synthesized gene into different vectors available in the laboratory. Once the artificial sequence was found to satisfy our design criteria, we placed an order for gene synthesis with GeneArt (Life Technologies, CA). The synthesized sequence was received as an insert in a pUC derivative vector (pMK-MyrSyn4) possessing kanamycin resistance.

3.3.5 Expression of Myrcene Synthase in *E. coli*

The synthetic gene sequence had been designed with a *lac* promoter sequence at the 5' end and also included a ribosome binding site upstream of the gene's start codon. Therefore, the pMK-MyrSyn4 vector was expected to express the myrcene synthase enzyme when transformed into an *E. coli* strain. However, protein expression and gel analysis could not help identify a distinct protein product corresponding to the expression of the introduced gene. We reasoned that this could due to poor gene expression from the *lac* promoter. Therefore, the myrcene synthase gene open reading frame was transferred from the pMK-MyrSyn4 vector into pETBlueKan Col 3 between the restriction enzyme sites NcoI – HindIII to form the T7 promoter driven vector, pETMyrSyn. Expression of this vector in the host strain *E. coli* BL21 (DE3) failed to reveal any protein band discernible from the wild type control. We reasoned that since the synthetic gene sequence had been predominantly optimized for expression in *A. variabilis*, the protein translation machinery of *E. coli* might have difficulties processing all the rare codons present in the sequence. We decided to test the expression of the myrcene synthase gene sequence in *E. coli* Rosetta, a strain which provides tRNA complementation for rare codons, thereby improving protein expression from gene sequences containing an unusually high percentage of rare codons. pETMyrSyn was transformed into *E. coli* Rosetta and protein expression was tested. Subsequent polyacrylamide gel electrophoresis failed to distinguish a protein product corresponding to the expression of myrcene synthase gene.

The reason for this perceived lack of protein expression from the synthetic gene

sequence could be that any expression of the protein causes stress on the *E. coli* strain carrying the gene. Therefore, the organism might restrict the rate at which the protein is being produced. Low intracellular concentration of an heterologous protein is not easily discerned in the small format protein gels used in this study. Hence it cannot be ruled that there is complete lack of protein expression from the synthetic gene in *E. coli*. Another cause for not observing protein expression could be that the expressed protein is not folded correctly into a stable structure. *E. coli* has limited ability to guide non-native proteins to fold into a stable structure and misfolded proteins are either degraded *in vivo* or form aggregates known as inclusion bodies. A helpful tactic to resolve formation of misfolded proteins would be to increase the intracellular concentration of native protein chaperone molecules that guide correct folding of proteins. This can be achieved by either plasmid based or genome based strategies for increased chaperone protein expression.

Myrcene Synthase Protein Sequence :

RRSANYQPSIWNHDYIESLRIEYVGETCTRQINVLKEQVRMMLHKVVNPLEQLELIEILQRLGLSYHFEEEIKRILD
GVYNNDHGGDTWKAENLYATALKFRLLRQHGYSVSQEVFNSFKDERGSFKACLCEDTKGMLSLYEASFFLIEGENIL
EEARDFSTKHLEEYVKQNKEKNLATLVNHSLEFPLHWRMPRLEARWFINIYRHNQDVNPILLEFAELDFNIVQAAHQ
ADLKQVSTWWKSTGLVENLSFARDRPVENFFWTVGLIFQPQFGYCRRMFTKVFALITTIDDVYDVYGTLDELELFTD
VVERWDINAMDQLPDYMKICFLTLHNSVNEMALDTMKEQRFHIIKYLKKAWVDLCRYYLVEAKWYSNKYRPSLQEYI
ENAWISIGAPTILVHAYFFVTNPITKEALDCLEEYPNIIRWSSIIARLADDLGTSTDELKRGDVPKAIQCYMNETGA
SEEGAREYIKYLISATWKKMNKDRAASSPFSHIFIEIALNLARMAQCLYQHGDGHGLGNRETKDRILSLLIQPIPLN
KD

Myrcene Synthase Gene Sequence in Holm Oak (Genbank AJ304839) :

CGGAGATCAGCAAATTACCAACCTTCCATTTGGAACCATGATTACATTGAGTCCCTAAGGATTGAATATGTGGGGGA
AACGTGCACTAGACAAATTAATGTGTTGAAGGAACAAGTGAGGATGATGCTTCACAAAGTGGTGAACCCTCTCGAGC
AACTTGAGTTGATAGAAATCTTGCAAAGACTAGGATTATCTTATCACTTTGAAGAAGAAATTAAAAGAATATTAGAT
GGTGTGTACAATAATGATCATGGTGGTGATACGTGGAAGGCGGAGAATTTGTATGCCACAGCTCTTAAATTTAGACT
CCTAAGACAACATGGATATAGTGTTTCTCAAGAAGTTTTCAATAGTTTCAAGGATGAGAGAGGGAGTTTCAAGGCAT
GCCTTTGTGAGGATACCAAGGGCATGCTATCCTTATATGAAGCCTCATTCTTTTTGATAGAAGGTGAAAATATCTTG
GAAGAAGCAAGAGATTTCTCAACCAAACATCTAGAAGAATACGTTAAGCAGAATAAAGAAAAAAATCTTGCTACTTT
AGTAAACCATTCCTTGGAGTTCCCACTGCATTGGAGGATGCCTAGGTTGGAAGCAAGATGGTTCATTAACATATATA
GACATAATCAAGACGTGAACCCTATCTTGCTTGAGTTTGCAGAATTGGATTTCAACATTGTACAAGCAGCCCACCAA
GCAGATCTAAAACAAGTGTCAACGTGGTGGAAGAGCACAGGCCTTGTAGAGAATTTGAGTTTTGCGAGGGATAGACC
GGTGGAGAATTTCTTTTGGACAGTGGGATTAATATTTCAGCCTCAATTTGGATATTGTAGGAGAATGTTTACCAAGG
TCTTTGCACTGATAACAACAATTGATGATGTTTATGATGTATATGGAACTTTGGATGAACTTGAGCTCTTCACAGAT
GTTGTTGAAAGATGGGATATCAATGCAATGGATCAACTCCCAGATTATATGAAGATATGTTTCCTTACTCTACATAA
TTCAGTTAATGAAATGGCTTTGGATACTATGAAGGAACAAAGATTCCACATCATTAAATACCTCAAAAAAGCGTGGG
TAGATTTATGTAGATATTATTTGGTGGAGGCAAAGTGGTACTCCAACAAATATAGACCGAGTCTACAAGAATACATT
GAGAATGCATGGATTTCAATAGGAGCACCAACTATATTAGTGCATGCTTATTTTTTTGTCACAAACCCAATTACAAA
GGAAGCCTTGGATTGCCTAGAAGAGTACCCCAACATAATCCGTTGGTCATCAATAATTGCACGACTTGCAGATGATC
TTGGAACATCTACGGATGAGTTAAAAAGAGGTGATGTCCCAAAAGCGATCCAGTGTTACATGAATGAAACCGGTGCA
AGCGAAGAAGGTGCCCGTGAGTATATAAAGTATTTGATTAGTGCAACATGGAAGAAAATGAATAAAGACCGAGCTGC
GAGTTCTCCCTTCTCTCATATATTTATTGAAATTGCATTGAACCTTGCAAGGATGGCCCAATGCTTGTACCAGCATG
GAGATGGACATGGTCTTGGGAACCGTGAAACTAAGGACCGAATATTGTCATTACTTATTCAGCCCATTCCCCTAAAC
AAAGATTGA

Synthetic Gene Sequence for Myrcene Synthase, optimized for *A. variabilis*:

CGCCGTTCTGCTAATTACCAGCCGAGCATTTGGAATCATGATTATATCGAATCGTTACGGATTGAGTATGTTGGTGA
AACTTGCACCCGGCAAATAAACGTTTTAAAAGAACAAGTAAGAATGATGTTGCATAAAGTTGTAAATCCATTAGAAC
AACTTGAACTTATTGAAATTTTACAACGTTTAGGATTAAGTTACCACTTCGAAGAAGAGATTAAGCGTATCTTAGAC
GGGGTGTATAACAACGATCATGGTGGCGATACATGGAAGGCTGAAAATTTATACGCCACAGCGTTGAAATTCCGTTT
ATTACGTCAACATGGATACTCTGTGAGTCAGGAAGTGTTTAATTCATTTAAAGACGAACGTGGCTCATTTAAGGCCT
GTTTGTGTGAAGATACGAAGGGCATGTTAAGTTTATACGAAGCTAGTTTTTTTTTAATCGAAGGTGAGAATATTCTG
GAAGAAGCCCGGGACTTTTCTACTAAACACTTAGAAGAGTACGTGAAACAGAATAAAGAAAAAAATTTAGCAACTTT
AGTTAATCACAGTCTCGAATTTCCATTGCATTGGCGCATGCCACGCTTGGAGGCCCGATGGTTCATCAATATCTATC
GTCATAACCAGGATGTCAATCCTATTCTCTTGGAGTTTGCCGAACTCGACTTTAACATTGTGCAGGCTGCCCATCAG
GCTGACTTAAAACAGGTTAGCACCTGGTGGAAAAGTACTGGTTTGGTAGAAAATCTGTCGTTTGCAAGGGATCGCCC
AGTCGAAAACTTCTTTTGGACTGTCGGGCTGATTTTCCAACCCCAATTTGGGTATTGTCGTCGCATGTTTACTAAAG
TTTTCGCGTTAATCACTACCATTGATGATGTGTACGATGTGTATGGCACACTGGACGAACTGGAATTGTTTACCGAT
GTGGTCGAACGTTGGGACATTAATGCTATGGACCAGCTGCCAGATTACATGAAAATCTGTTTTCTCACCTTACATAA
TAGCGTCAATGAAATGGCATTGGATACGATGAAAGAACAAAGATTTCACATCATCAAATACCTTAAAAAAGCGTGGG
TTGACTTGTGTCGCTACTACTTAGTCGAGGCAAAATGGTACTCGAATAAATATCGGCCGTCCTTACAAGAATATATT
GAGAATGCGTGGATCAGTATTGGCGCTCCTACAATCTTGGTCCACGCATATTTTTTTGTAACAAATCCGATTACTAA
GGAAGCATTGGATTGCTTAGAGGAATATCCTAATATTATCCGTTGGAGTAGTATTATTGCTCGCTTAGCCGATGATT
TAGGAACTAGTACAGACGAACTGAAACGGGGTGATGTGCCTAAAGCTATTCAATGTTATATGAATGAAACGGGCGCC
AGTGAAGAAGGCGCTCGCGAATATATTAAATACTTAATCTCGGCGACATGGAAAAAAATGAATAAGGACAGAGCTGC
TAGTTCTCCTTTTAGCCACATTTTTATCGAGATAGCGCTGAACTTGGCTCGTATGGCTCAGTGTTTGTATCAGCATG
GTGACGGTCATGGTCTAGGTAACCGTGAAACCAAAGATCGCATCTTGTCACTGTTAATCCAGCCAATTCCATTAAAT
AAAGACTAA

Figure 3.11: The synthetic Myrcene Synthase gene sequence that codes for the same protein as the gene sequence found in Holm Oak.

3.4 Discussion

Monoterpenes are organic molecules of increasing commercial importance, in part due to their wide use ranging from the manufacture of cosmetics, fragrances, essential oils, medicines to novel fuels. While plants produce a variety of such terpene compounds, the volatility of these chemicals makes it inconvenient to develop plants as viable sources of large scale production. Single celled organisms like bacteria and yeasts that can be grown in bioreactors are ideal hosts for production of such volatile chemicals. Genetic engineering of *E. coli* and *Saccharomyces cerevisiae* strains has led to the demonstration of terpene synthesis and its viability in an industrial setting. Still, these organisms depend on externally supplied sugar molecules for their carbon source. Therefore cyanobacteria like *Anabaena variabilis*, which are able to fix atmospheric CO_2 for their carbon requirements, can be considered as ideal recombinant hosts for production of carbon-rich molecules like terpenes. One of the primary drawbacks that prevent widespread adoption of cyanobacteria as recombinant hosts is the prolonged timeframe required for genetic modification compared to rapid turnovers with *E. coli* and *S. cerevisiae*. In our intent to establish monoterpene synthesis in *A. variabilis*, we wanted to establish a prototype in *E. coli*. This work described our efforts towards that goal.

Genes from *E. coli* that could together lead to formation of a sesquiterpene precursor, farnesyl diphosphate were cloned into plasmid vectors. The *ispA* prenyltransferase gene was mutated at a specific site so that its product specificity was changed from farnesyl pyrophosphate to geranyl pyrophosphate. The *ispA* S81F, *dxs* and *isi* genes were assembled into an artificial operon so that all genes were expressed in a

recombinant host above wild type levels, from a single inducible promoter system. The development of an artificial operon involved introducing synthetic ribosome binding sites upstream of the *dxs* and *isi* genes, replacing the natural RBS of these genes. The synthetic RBS provided elevated expression of the corresponding genes. The entire operon was placed in a single low copy plasmid, which provides us with the convenience of expressing multiple genes from a single vector, thereby reducing the metabolic burden associated with maintaining multiple plasmids. Also, an artificial gene was designed and synthesized to circumvent the gene expression problems associated with expressing natural DNA sequences in recombinant organisms. Optimized for expression in *A. variabilis*, no protein expression has so far been perceived from this synthetic gene when introduced into *E. coli*. Probable causes and remedies are currently being explored to alleviate the protein expression problem.

Thus this section of work demonstrates the utility of modular approaches to genetic engineering. This work also contributed towards understanding the nuances in assembling multi-gene constructs. This is a key learning experience for our laboratory that would be important for future work in heterologous expression of multi-enzyme pathways towards bioproduct synthesis.

3.5 Future Work

An expression vector containing the myrcene synthase gene should be co-transformed into an *E. coli* strain containing the vector pPCRISPDSX11 Col 1. Expression of the genes in the respective plasmids should be induced and the headspace of the culture vessel should be analyzed using gas chromatography - mass spectrometry methods for the presence of any myrcene produced. Also, the myrcene synthase gene should be cloned into a vector that contains an *A. variabilis* specific promoter system. This vector should then be introduced into *A. variabilis*. Wildtype *A. variabilis* has a number of uncharacterized prenyltransferase genes that can putatively synthesize GPP. Headspace analysis of a batch culture of recombinant *A. variabilis* which contains the myrcene synthase gene will indicate whether the organism is able to produce myrcene, even if the initial yield is low. If successful, this demonstration would then be a stepping stone to further genetic modifications of *A. variabilis* in order to redirect metabolic flux towards increased myrcene biosynthesis in the cyanobacterium *Anabaena variabilis*.

4.1 References

Adams, D. G., & Duggan, P. S. (1999). Tansley Review No. 107. Heterocyst and akinete differentiation in cyanobacteria. *New Phytologist*, *144*(1), 3–33

Ajikumar, P. K., Xiao, W.-H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., Pfeifer, B., & Stephanopoulos, G. (2010). Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in *Escherichia coli*. *Science*, *330*(6000), 70–74

Akkina, S. K., Zhang, Y., Nelsestuen, G. L., Oetting, W. S., & Ibrahim, H. N. (2009). Temporal Stability of the Urinary Proteome after Kidney Transplant: More Sensitive than Protein Composition? *Journal of Proteome Research*, *8*(1), 94–103

Alvarez, H. M. (2010). Biotechnological Production and Significance of Triacylglycerols and Wax Esters. In K. N. Timmis (Ed.), *Handbook of Hydrocarbon and Lipid Microbiology* (pp. 2995–3002). Springer Berlin Heidelberg. Retrieved from http://dx.doi.org/10.1007/978-3-540-77587-4_222

Alvarez, H. M., & Steinbüchel, A. (2002). Triacylglycerols in prokaryotic microorganisms. *Applied microbiology and biotechnology*, *60*(4), 367–376

Anderson, A. J., & Dawes, E. A. (1990). Occurrence, metabolism, metabolic role, and industrial uses of bacterial polyhydroxyalkanoates. *Microbiological reviews*, *54*(4), 450–472.

Atsumi, S., Higashide, W., & Liao, J. C. (2009). Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. *Nature Biotechnology*, *27*(12), 1177–1180

Berger, R. G., Babel, W., Blanch, H. W., Cooney, C. L., Enfors, S.-O., Eriksson, K.-E. L., Fiechter, A., Klibanov, A. M., Mattiasson, B., Primrose, S. B., Rehm, H. J., Rogers, P. L., Sahm, H., Schügerl, K., Tsao, G. T., Venkat, K., Villadsen, J., Stockar, U., & Wandrey, C. (Eds.). (1997). *Advances in Biochemical Engineering/Biotechnology*: Springer Berlin Heidelberg.

Berry, D. A. (2010). Engineering organisms for industrial fuel production. *Bioengineered Bugs*, *1*(5), 303–308

Brasser, A. J., Barwacz, C. A., Dawson, D. V., Brogden, K. A., Drake, D. R., & Wertz, P. W. (2011). Presence of wax esters and squalene in human saliva. *Archives of Oral Biology*, *56*(6), 588–591

Carter, O., Peters, R., & Croteau, R. (2003). Monoterpene biosynthesis pathway construction in *Escherichia coli*. *Phytochemistry*, *64*(2), 425–433

Chaurasia, A. K., & Apte, S. K. (2011). Improved Eco-Friendly Recombinant *Anabaena* sp. Strain PCC7120 with Enhanced Nitrogen Biofertilizer Potential. *Applied and Environmental Microbiology*, *77*(2), 395–399

Domon, B., & Aebersold, R. (2010). Options and considerations when selecting a quantitative proteomics strategy. *Nature Biotechnology*, *28*(7), 710–721

Ducat, D. C., Way, J. C., & Silver, P. A. (2011). Engineering cyanobacteria to generate high-value products. *Trends in Biotechnology*, *29*(2), 95–103

Fischbach, R. J., Zimmer, W., & Schnitzler, J. P. (2001). Isolation and functional analysis of a cDNA encoding a myrcene synthase from holm oak (*Quercus ilex* L.). *European journal of biochemistry / FEBS*, *268*(21), 5633–5638.

Fujita, Y., Matsuoka, H., & Hirooka, K. (2007). Regulation of fatty acid metabolism in bacteria. *Molecular microbiology*, *66*(4), 829–839

Grossi, V., Yakimov, M. M., Al Ali, B., Tapilatu, Y., Cuny, P., Goutx, M., La Cono, V., Giuliano, L., & Tamburini, C. (2010). Hydrostatic pressure affects membrane and storage lipid compositions of the piezotolerant hydrocarbon-degrading *Marinobacter hydrocarbonoclasticus* strain #5. *Environmental microbiology*, *12*(7), 2020–2033

Gustafsson, C., Govindarajan, S., & Minshull, J. (2004). Codon bias and heterologous protein expression. *Trends in Biotechnology*, *22*(7), 346–353

Handke, P., Lynch, S. A., & Gill, R. T. (2011). Application and engineering of fatty acid biosynthesis in *Escherichia coli* for advanced fuels and chemicals. *Metabolic engineering*, *13*(1), 28–37

Harvey, B. G., Wright, M. E., & Quintana, R. L. (2010). High-Density Renewable Fuels Based on the Selective Dimerization of Pinenes. *Energy & Fuels*, *24*(1), 267–273

Hengge-Aronis, R. (1996). Back to log phase: sigma S as a global regulator in the osmotic control of gene expression in *Escherichia coli*. *Molecular microbiology*, *21*(5), 887–893.

Hernández, M. A., Mohn, W. W., Martínez, E., Rost, E., Alvarez, A. F., & Alvarez, H. M. (2008). Biosynthesis of storage compounds by *Rhodococcus jostii* RHA1 and global identification of genes involved in their metabolism. *BMC genomics*, *9*, 600

Hills, G. (2003). Industrial use of lipases to produce fatty acid esters. *European Journal of Lipid Science and Technology*, *105*(10), 601–607

Holtzapple, E., & Schmidt-Dannert, C. (2007). Biosynthesis of isoprenoid wax ester in *Marinobacter hydrocarbonoclasticus* DSM 8798: identification and characterization of isoprenoid coenzyme A synthetase and wax ester synthases. *Journal of bacteriology*, *189*(10), 3804-3812

Hong, S. H., Kim, J. S., Lee, S. Y., In, Y. H., Choi, S. S., Rih, J.-K., Kim, C. H., Jeong, H., Hur, C. G., & Kim, J. J. (2004). The genome sequence of the capnophilic rumen bacterium *Mannheimia succiniciproducens*. *Nature biotechnology*, *22*(10), 1275–1281

Huu, N. B., Denner, E. B., Ha, D. T., Wanner, G., & Stan-Lotter, H. (1999). *Marinobacter aquaeolei* sp. nov., a halophilic bacterium isolated from a Vietnamese oil-producing well. *International journal of systematic bacteriology*, *49 Pt 2*, 367-375

Ishige, T. (2003). Wax ester production by bacteria. *Current Opinion in Microbiology*, *6*(3), 244–250

Ishige, T., Tani, A., Sakai, Y., & Kato, N. (2000). Long-chain aldehyde dehydrogenase that participates in n-alkane utilization and wax ester synthesis in Acinetobacter sp. strain M-1. *Applied and environmental microbiology*, *66*(8), 3481–3486.

Ishige, T., Tani, A., Takabe, K., Kawasaki, K., Sakai, Y., & Kato, N. (2002). Wax ester production from n-alkanes by Acinetobacter sp. strain M-1: ultrastructure of cellular inclusions and role of acyl coenzyme A reductase. *Applied and environmental microbiology*, *68*(3), 1192–1195.

Jetter, R., & Kunst, L. (2008). Plant surface lipid biosynthetic pathways and their utility for metabolic engineering of waxes and hydrocarbon biofuels. *The Plant Journal*, *54*(4), 670–683

Jozefczuk, S., Klie, S., Catchpole, G., Szymanski, J., Cuadros-Inostroza, A., Steinhauser, D., Selbig, J., & Willmitzer, L. (2010). Metabolomic and transcriptomic stress response of *Escherichia coli*. *Molecular systems biology*, *6*, 364

Kalscheuer, R. (2010). Genetics of Wax Ester and Triacylglycerol Biosynthesis in Bacteria. In K. N. Timmis (Ed.), *Handbook of Hydrocarbon and Lipid Microbiology* (pp. 527–535). Springer Berlin Heidelberg.

Kalscheuer, R., & Steinbüchel, A. (2003). A novel bifunctional wax ester synthase/acyl-CoA:diacylglycerol acyltransferase mediates wax ester and triacylglycerol biosynthesis in *Acinetobacter calcoaceticus* ADP1. *The Journal of biological chemistry*, *278*(10), 8075–8082

Kalscheuer, R., Stölting, T., & Steinbüchel, A. (2006). Microdiesel: *Escherichia coli* engineered for fuel production. *Microbiology (Reading, England)*, *152*(Pt 9), 2529-2536

Kane, J. F. (1995). Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology*, *6*(5), 494–500

Kim, S. W., & Keasling, J. D. (2001). Metabolic engineering of the nonmevalonate isopentenyl diphosphate synthesis pathway in *Escherichia coli* enhances lycopene production. *Biotechnology and Bioengineering*, *72*(4), 408–415.

Kovach, M. E., Elzer, P. H., Hill, D. S., Robertson, G. T., Farris, M. A., Roop, R. M., & Peterson, K. M. (1995). Four new derivatives of the broad-host-range cloning vector pBBR1MCS, carrying different antibiotic-resistance cassettes. *Gene*, *166*(1), 175–176.

Kovach, M. E., Phillips, R. W., Elzer, P. H., Roop, R. M., & Peterson, K. M. (1994). pBBR1MCS: a broad-host-range cloning vector. *BioTechniques*, *16*(5), 800–802.

Kudla, G., Murray, A. W., Tollervey, D., & Plotkin, J. B. (2009). Coding-Sequence Determinants of Gene Expression in *Escherichia coli*. *Science*, *324*(5924), 255–258

Lee, C. H., Bhagwat, A., & Heffron, F. (1983). Identification of a transposon Tn3 sequence required for transposition immunity. *Proceedings of the National Academy of Sciences of the United States of America*, *80*(22), 6765–6769.

Lee, J. W., Lee, S. Y., Song, H., & Yoo, J.-S. (2006). The proteome of *Mannheimia succiniciproducens*, a capnophilic rumen bacterium. *Proteomics*, *6*(12), 3550–3566

Lee, P. C., Lee, S. Y., Hong, S. H., & Chang, H. N. (2002). Isolation and characterization of a new succinic acid-producing bacterium, *Mannheimia succiniciproducens* MBEL55E, from bovine rumen. *Applied microbiology and biotechnology*, *58*(5), 663–668

Lee, S. Y., Kim, J. M., Song, H., Lee, J. W., Kim, T. Y., & Jang, Y.-S. (2008). From genome sequence to integrated bioprocess for succinic acid production by *Mannheimia succiniciproducens*. *Applied microbiology and biotechnology*, *79*(1), 11–22

Leonard, E., Ajikumar, P. K., Thayer, K., Xiao, W.-H., Mo, J. D., Tidor, B., Stephanopoulos, G., & Prather, K. L. J. (2010). Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. *Proceedings of the National Academy of Sciences*, *107*(31), 13654–13659

Liu, X., Sheng, J., & Curtiss III, R. (2011). Fatty acid production in genetically modified cyanobacteria. *Proceedings of the National Academy of Sciences*, *108*(17), 6899–6904

Manilla-Pérez, E., Lange, A. B., Hetzler, S., & Steinbüchel, A. (2010). Occurrence, production, and export of lipophilic compounds by hydrocarbonoclastic marine bacteria and their potential use to produce bulk chemicals from hydrocarbons. *Applied microbiology and biotechnology*, *86*(6), 1693–1706

Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D., & Keasling, J. D. (2003). Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature Biotechnology*, *21*(7), 796–802

Martin, V. J., Yoshikuni, Y., & Keasling, J. D. (2001). The *in vivo* synthesis of plant sesquiterpenes by*Escherichia coli*. *Biotechnology and Bioengineering*, *75*(5), 497–503

McCaskill, D., & Croteau, R. (1997). Prospects for the bioengineering of isoprenoid biosynthesis. In R. G. Berger, W. Babel, H. W. Blanch, C. L. Cooney, S.-O. Enfors, K.-E. L. Eriksson, … (Eds.), *Advances in Biochemical Engineering/Biotechnology* (pp. 107–146). Springer Berlin Heidelberg.

McGarvey, D. J. (1995). Terpenoid Metabolism. *The Plant Cell Online*, *7*(7), 1015–1026

Minnikin, D. E., Dobson, G., Goodfellow, M., Magnusson, M., & Ridell, M. (1985). Distribution of some mycobacterial waxes based on the phthiocerol family. *Journal of general microbiology*, *131*(6), 1375–1381.

Newman, D. J., & Cragg, G. M. (2007). Natural Products as Sources of New Drugs over the Last 25 Years. *Journal of Natural Products*, *70*(3), 461–477

Oh, J. in, Kim, D. H., Oh, E. K., Lee, S. Y., & Lee, J. (2009). Optimization and scale-up of succinic acid production by *Mannheimia succiniciproducens* LPK7. *Journal of microbiology and biotechnology*, *19*(2), 167–171.

Ohnuma, S. i., Narita, K., Nakazawa, T., Ishida, C., Takeuchi, Y., Ohto, C., & Nishino, T. (1996). A role of the amino acid residue located on the fifth position before the first aspartate-rich motif of farnesyl diphosphate synthase on determination of the final product. *The Journal of biological chemistry*, *271*(48), 30748–30754.

Patel, S., Nelson, D. R., & Gibbs, A. G. (2001). Chemical and physical analyses of wax ester properties. *Journal of insect science (Online)*, *1*, 4.

Redding, A. M., Mukhopadhyay, A., Joyner, D. C., Hazen, T. C., & Keasling, J. D. (2006). Study of nitrate stress in *Desulfovibrio vulgaris* Hildenborough using iTRAQ proteomics. *Briefings in functional genomics & proteomics*, *5*(2), 133–143

Reiling, K. K., Yoshikuni, Y., Martin, V. J., Newman, J., Bohlmann, J., & Keasling, J. D. (2004). Mono and diterpene production in *Escherichia coli*. *Biotechnology and Bioengineering*, *87*(2), 200–212

Reiser, S., & Somerville, C. (1997). Isolation of mutants of *Acinetobacter calcoaceticus* deficient in wax ester synthesis and complementation of one mutation with a gene encoding a fatty acyl coenzyme A reductase. *Journal of bacteriology*, *179*(9), 2969–2975.

Rohmer, M. (1999). The discovery of a mevalonate-independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Natural Product Reports*, *16*(5), 565–574

Rontani, J. F. (2010). 34: Production of Wax Esters by Bacteria. In K. N. Timmis (Ed.), *Handbook of Hydrocarbon and Lipid Microbiology* (pp. 459-470). Berlin, Heidelberg: Springer Berlin Heidelberg.

Rontani, J. F., Bonin, P. C., & Volkman, J. K. (1999). Production of wax esters during aerobic growth of marine bacteria on isoprenoid compounds. *Applied and environmental microbiology*, *65*(1), 221-230.

Saltikov, C. W., & Newman, D. K. (2003). Genetic identification of a respiratory arsenate reductase. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(19), 10983–10988

Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, *3*(6), 1101–1108

Simpson, R. J. (2006). SDS-PAGE of Proteins. *Cold Spring Harbor Protocols*, *2006*(16), pdb.prot4313

Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S. B., & Keasling, J. D. (2010). Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, *463*(7280), 559–562

Stöveken, T., Kalscheuer, R., Malkus, U., Reichelt, R., & Steinbüchel, A. (2005). The wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase from Acinetobacter sp. strain ADP1: characterization of a novel type of acyltransferase. *Journal of bacteriology*, *187*(4), 1369–1376

Studier, F. W., & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of molecular biology*, *189*(1), 113–130.

Tani, A., Sakai, Y., Ishige, T., & Kato, N. (2000). Thermostable NADP(+)-dependent medium-chain alcohol dehydrogenase from Acinetobacter sp. strain M-1: purification and characterization and gene expression in *Escherichia coli*. *Applied and environmental microbiology*, *66*(12), 5231–5235.

Tani, A., Ishige, T., Sakai, Y., & Kato, N. (2002). Two acyl-CoA dehydrogenases of Acinetobacter sp. strain M-1 that uses very long-chain n-alkanes. *Journal of bioscience and bioengineering*, *94*(4), 326–329.

Teerawanichpan, P., & Qiu, X. (2010). Fatty Acyl-CoA Reductase and Wax Synthase from *Euglena gracilis* in the Biosynthesis of Medium-Chain Wax Esters. *Lipids*, *45*(3), 263–273

Timmis, K. N. (Ed.). (2010). *Handbook of Hydrocarbon and Lipid Microbiology*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Uthoff, S., Stöveken, T., Weber, N., Vosmann, K., Klein, E., Kalscheuer, R., & Steinbüchel, A. (2005). Thio wax ester biosynthesis utilizing the unspecific bifunctional wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase of Acinetobacter sp. strain ADP1. *Applied and environmental microbiology*, *71*(2), 790–796

van Vliet, A. H. M. (2010). Next generation sequencing of microbial transcriptomes: challenges and opportunities. *FEMS microbiology letters*, *302*(1), 1–7

Vaysse, P.-J. J., Prat, L., Mangenot, S., Cruveiller, S., Goulas, P., & Grimaud, R. (2009). Proteomic analysis of *Marinobacter hydrocarbonoclasticus* SP17 biofilm formation at the alkane-water interface reveals novel proteins and cellular processes involved in hexadecane assimilation. *Research in microbiology*, *160*(10), 829-837

Wahlen, B. D., Oswald, W. S., Seefeldt, L. C., & Barney, B. M. (2009). Purification, characterization, and potential bacterial wax production role of an NADPH-dependent fatty aldehyde reductase from *Marinobacter aquaeolei* VT8. *Applied and environmental microbiology*, *75*(9), 2758-2764

Wältermann, M., & Steinbüchel, A. (2005). Neutral lipid bodies in prokaryotes: recent insights into structure, formation, and relationship to eukaryotic lipid depots. *Journal of bacteriology*, *187*(11), 3607–3619

Wang, C. W., Oh, M. K., & Liao, J. C. (1999). Engineered isoprenoid pathway enhances astaxanthin production in *Escherichia coli*. *Biotechnology and Bioengineering*, *62*(2), 235–241.

Welch, M., Govindarajan, S., Ness, J. E., Villalobos, A., Gurney, A., Minshull, J., Gustafsson, C., & Kudla, G. (2009). Design Parameters to Control Synthetic Gene Expression in *Escherichia coli*. *PLoS ONE*, *4*(9), e7002

Wilson, W. A., Roach, P. J., Montero, M., Baroja-Fernández, E., Muñoz, F. J., Eydallin, G., Viale, A. M., & Pozueta-Romero, J. (2010). Regulation of glycogen metabolism in yeast and bacteria. *FEMS microbiology reviews*, *34*(6), 952–985

Zieske, L. R. (2006). A perspective on the use of iTRAQ reagent technology for protein complex and profiling studies. *Journal of experimental botany*, *57*(7), 1501–1508

Zinn, M., Witholt, B., & Egli, T. (2001). Occurrence, synthesis and medical application of bacterial polyhydroxyalkanoate. *Advanced drug delivery reviews*, *53*(1), 5–21.