Plant Phenylpropanoid Biosynthesis in *Escherichia coli*: Engineering Novel Pathways and Tools

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Sarah E. Bloch

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

Adviser: Dr. Claudia Schmidt-Dannert

October 2014

Acknowledgements

I would like to thank Claudia for the opportunity to explore exciting work in her lab and to gain a variety of experience and expertise that has equipped me well for my future career. I would also like to thank all of the Schmidt-Dannert lab members who gave me guidance and mentorship over the years. Specifically, I must thank Dr. Jacob Vick for teaching me the biochemistry and cloning basics and for guiding me through the lignan biosynthesis project. Dr. Mark Held provided invaluable guidance on publications and experimental design and help carrying out the Eut BMC encapsulation experiments, and for that I thank him. I am also thankful to Dr. Ethan Johnson, Dr. Fernando Lopez-Gallego, and Dr. Ilya Tikh for the training and advice I received from them. Dr. Grayson Wawrzyn was a valuable scientific and personal sounding board and support throughout my graduate career, and for that I am very grateful. In addition, I thank the undergraduates who worked with me on the BMC encapsulation project, Suzie Hsu and Kelsey Dahlgren, as well as all of the iGEM students whom I guided over the years, who made graduate school a lot more fun.

Finally, I am thankful to everyone else who kept me sane during the completion of this work and encouraged me to persevere: Monica Snyder, my parents, Mick Nyquist, Dr. Bree Hamman, Gregory J. Scott, Dr. Misha Golinskiy, Dr. Laura Mazzaferro, Dr. David Babson, and Dr. Sharon Murphy.

Dedication

I dedicate this thesis to the educators in my early life that fostered my passion for science: My mother and grandfather, who taught me notice and wonder at the living things around me; Mr. Greg Ulmer, an outstanding high school educator who sparked my fascination with the interface of biology and chemistry; Dr. Peter Goldman, who introduced me to the remarkable narrative of biological history; and Dr. Barbara Kramer, who showed me how exhilarating scientific research can be.

Abstract

Plant phenylpropanoid natural products are important in the discovery of safe and effective therapeutics. Most plant natural products cannot be economically mass produced *via* extraction from plant tissue or chemical synthesis. In recent decades, engineering microbes to carry out the biosynthesis of plant natural products has emerged as a powerful technology. The goal of this thesis was to expand the capabilities of microbial biosynthesis of plant phenylpropanoids in *Escherichia coli* through exploring novel biosynthetic pathways and metabolic engineering tools. I first explored the biosynthesis of valuable lignans in *E. coli*, establishing random oxidative radical coupling through overexpression of a laccase and attempting to show stereoselective coupling by a dirigent protein. I also designed and built a biosynthetic pathway for rosmarinic acid, a valuable hydroxycinnamic acid ester, showed pathway bottlenecks and limitations, and identified future optimization strategies. I have also begun a project to better understand cargo protein encapsulation within bacterial microcompartments and to develop their utility as a means of spatially organizing metabolic pathways. This work has contributed significantly to the field of microbial metabolic engineering and has laid the groundwork for future economically viable production platforms.

Table of Contents

List of Tables

List of Figures

Chapter 4. Exploring Engineered Metabolons: Investigation of Cargo Encapsulation within Eut Bacterial Microcompartments 94

List of Schemes

Chapter 1. Bioactive Phenylpropanoids: Biosynthesis in Plants and Strategies for Microbial Production

1. Background: the Need for Microbial Biosynthesis of Plant Natural Products

Plants are the sources of myriad natural products which enhance the daily lives of human beings. They have evolved the capability to synthesize these products for their own purposes – for protection from ultraviolet rays, to attract pollinators, and for defense against herbivores and pathogens. Yet over the millennia humans have found their own uses these same compounds, as flavors and colors that please our senses, as components of our most nutritious foods, and as medicines to reduce pain, promote healing and treat disease. Through the study of chemistry and biochemistry, humans have begun to understand on a molecular level the basis of the medicinal plants we have utilized for centuries, and much focus has been devoted to the discovery of valuable natural products for pharmaceutical use.

In recent decades, pharmaceutical discovery companies have shifted focus away from natural products due to limitations in their mass production. Source plants are often slow-growing, producing little biomass and low concentrations of the valuable products; additionally, extraction of the desired compound is arduous, often requiring separation from hundreds of other compounds.[1-2] Chemical synthesis provides a lower-cost, more rapid means of producing pharmaceuticals; however, most known bioactive natural products are highly complex and contain multiple chiral centers, making their

chemical synthesis economically infeasible.[3] Because of this, pharmaceutical companies have turned away from natural products and towards high-throughput screening of synthetic combinatorial chemical libraries for drug discovery.[4-5] Strikingly, the value of natural products in pharma is illustrated by the fact that over half of the drugs approved and brought to market over the last three decades have been natural products, despite the shift towards synthetic drug discovery.[6] In fact, the success rate for regulatory approval of drugs derived from natural products is fourfold greater than that of conventional synthetic drugs.[1] It is becoming clear that easily-synthesized compounds cannot replace natural products in the discovery and implementation of new pharmaceuticals, which necessitates the development of an economical production platform for complex plant natural products.

With the advent of recombinant DNA technology, a new means of synthesizing complex natural products has emerged and shown high potential for economical production of valuable compounds. After the discovery, cloning and characterization of genes and enzymes required for the biosynthesis of valuable compounds, microbes can be genetically reprogrammed to express non-native biosynthetic pathways and synthesize natural products through fermentation technology. Microbial fermentations benefit from fast biomass generation, a simplified extraction process compared to plant tissue extraction, and the ability to generate valuable compounds from inexpensive feedstocks.[1] The use of biosynthetic enzymes as *in vivo* catalysts makes this a stereo-selective process, and the recent sequencing revolution has provided us with an

enormous number of potential biosynthetic enzymes.[7] The implementation of this strategy has already led to the production of high value products in microbes, especially for plant-derived isoprenoids. For example, the biosynthesis of artemisinic acid, an important precursor to the anti-malarial drug artemisinin, has been achieved in *Saccharomyces cerevisiae* to a titer of 25 g/L and has led to an economically viable semi-synthetic production platform for the drug.[8] Likewise, key precursors to the anti-cancer drug paclitaxel have been achieved on the g/L scale in engineered strains in *Escherichia coli*.[9] Other medically-important plant isoprenoids, flavonoids, stilbenoids and alkaloids have also been produced in these hosts.[1, 10]

The goal of this thesis was to explore the microbial biosynthesis of plant-derived phenylpropanoid compounds with high medicinal value. While the biosynthesis of some plant phenylpropanoids has already been achieved in microbes, they have mostly been limited to the hydroxycinnamic acid, flavonoid, and stilbenoid compounds, and have so far overlooked some key value-added compound classes. Because of the previous success of the Schmidt-Dannert lab[11-12] and many other groups (reviewed in [1, 10]) in the use of *E. coli* as a host for flavonoid biosynthesis, I chose this highly-tractable microorganism as a host for phenylpropanoid biosynthetic pathways. I first explored its potential as a host for the biosynthesis of lignans as a first step toward the microbial biosynthesis of podophyllotoxin, a highly cytotoxic pro-drug whose derivatives are widely used in chemotherapy. I also developed a novel, chimeric biosynthetic pathway for an anti-cancer and neuroprotective hydroxycinnamic acid ester, rosmarinic acid, and

achieved its biosynthesis in *E. coli* for the first time. Finally, I have attempted to optimize the rosmarinic acid biosynthetic pathway through co-localization of two key enzymes to engineered bacterial microcompartments (BMCs), and have investigated the mechanism of protein encapsulation within BMC shells. This work has provided key insights into the biosynthesis of plant phenylpropanoids in *E. coli*, including its feasibility as a host for certain compound classes and the development of creative strategies to optimize and tailor pathways to the chosen host.

2. Thesis Overview

Chapter 1. This chapter provides a background for the importance of plant phenylpropanoid compounds and their biosynthesis in *E. coli*. I provide a summary of both plant and bacterial phenylpropanoid biosynthesis, and describe how groups have networked *E. coli*'s native L-tyrosine biosynthetic pathway with heterologous biosynthetic pathways to make value-added phenylpropanoids from simple feedstocks. I also discuss the importance of spatial organization in metabolic pathways and outline recently-developed strategies for improving pathway flux through enzyme colocalization.

Chapter 2. This chapter describes my early work towards developing a lignan biosynthetic platform in *E. coli*. Working in collaboration with Dr. Jacob Vick, I attempted the functional heterologous expression of two key proteins in the first committed step in lignan biosynthesis – a bacterial laccase and a plant-derived dirigent protein. While I

achieved expression of both proteins, the dirigent protein was found to be non-functional in *E. coli*, calling into question the utility of this organism as a host for lignan biosynthesis.

Chapter 3. This chapter describes the design and construction of a hydroxycinnamic acid biosynthetic pathway in *E. coli*. I conceptualized and implemented a chimeric biosynthetic pathway for rosmarinic acid distinct from that in plants, using both plant and bacterial enzymes. While initial titers from the pathway are insufficient for an economical production platform, I provide several strategies for future pathway optimization based on my investigation of pathway bottlenecks and diversions.

Chapter 4. This chapter discusses an ongoing investigation in the Schmidt-Dannert laboratory into the mechanism of enzyme encapsulation into engineered Eut BMCs in *E. coli*. I outline a strategy to optimize the rosmarinic acid biosynthetic pathway described in Chapter 3 through partial encapsulation within a BMC and show preliminary results of enzyme localization. I also describe an investigation of the interactions between BMC shell and cargo proteins that mediate encapsulation and show evidence of cognate shell/cargo protein pairs. For both of these studies, I propose future work to advance our understanding of enzyme encapsulation within BMCs. While I conceptualized this

work, some of the cloning, mutagenesis and microscopy was carried out by Suzie Hsu and Dr. Mark Held.

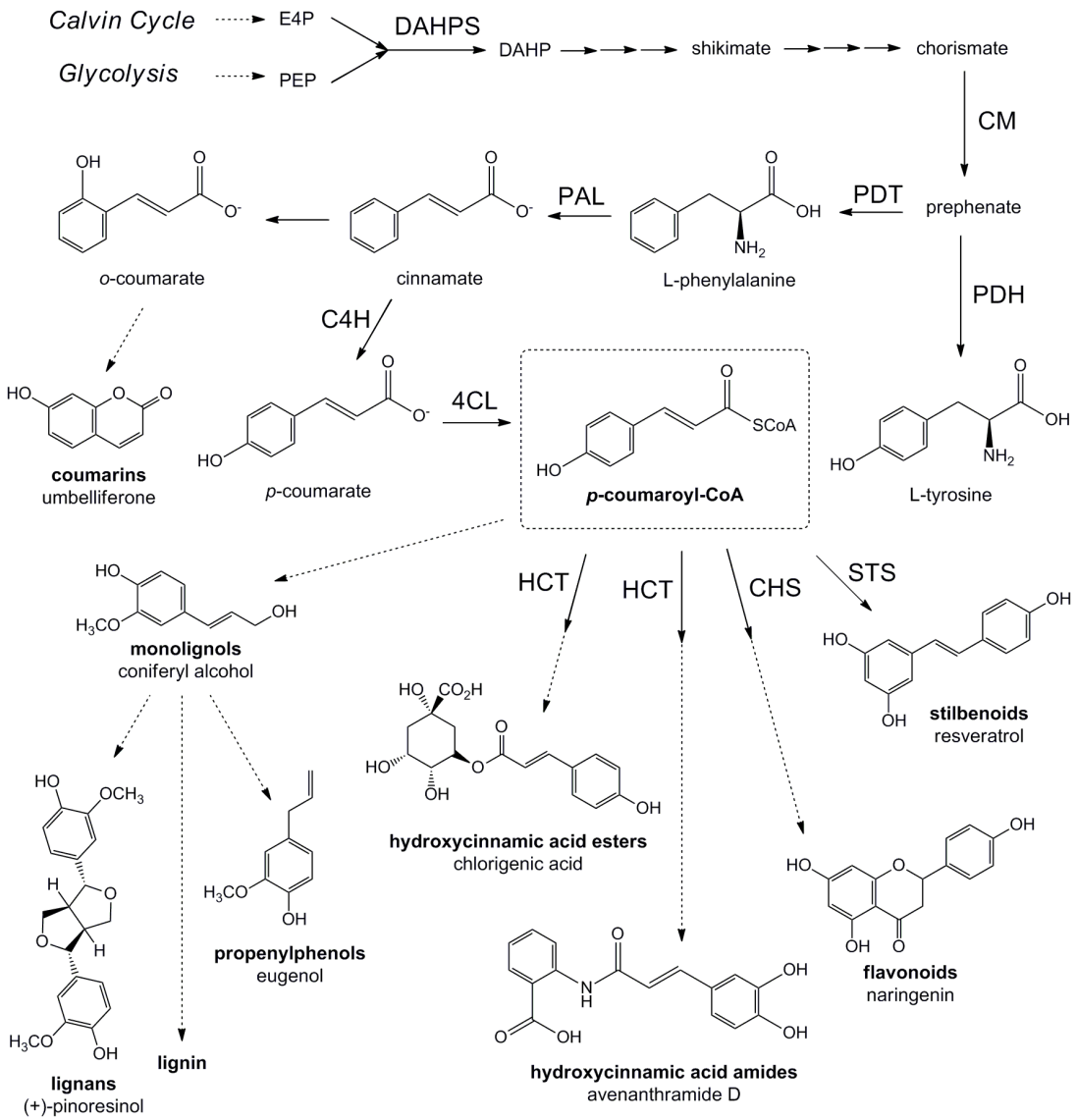
Chapter 5. This chapter summarizes the work described in the previous chapters and highlights its contribution to the field of microbial metabolic engineering.

3. Phenylpropanoid Biosynthesis in Plants and Bacteria

The biosynthesis of bioactive plant phenylpropanoids begins with the generation of aromatic amino acids *via* the shikimate pathway (Scheme 1.1).[13] The first step in shikimate biosynthesis is the condensation of phosphoenolpyruvate (PEP) from the glycolytic pathway and erythrose-4-phosphate (E4P) from the Calvin cycle and the pentose phosphate pathway to give 3-deoxy-D-arabinoheptulosonate-7-phosphate (DAHP) in a reaction catalyzed by DAHP synthase (DAHPS). DAHP is then converted *via* three enzymatic steps into shikimate, which is the key starting point in the aromatic amino acid biosynthesis (AAAB) pathway.[14] Another three enzymatic steps convert shikimate to chorismate, which is acted upon by chorismate mutase (CM) to produce prephenate. Prephenate serves as the branch point for L-phenylalanine and L-tyrosine biosynthesis, with prephenate dehydratase (PDT) and phenylpyruvate aminotransferase catalyzing its conversion to L-phenylalanine, and prephanate dehydrogenase (PDH) and 4-hydroxyphenylpyruvate aminotransferase catalyzing its conversion to L-tyrosine.[15]

For the most part, *E. coli* shares the same biosynthetic pathway for the aromatic amino acids, except that chorismate serves as the branch point for L-phenylalanine and L-tyrosine biosynthesis. Additionally, in *E. coli*, a single bi-functional enzyme, CM-PDT or CM-PDH, converts chorismate into either phenylpyruvate or 4-hydroxyphenylpyruvate, respectively (Scheme 1.2).[16] In plants, the activity of PDT and PDH is known to be subject to feedback inhibition by L-phenylalanine and L-tyrosine.[13, 15] Likewise, in *E. coli*, it has been shown that CM-PDT and CM-PDH are subject to feedback inhibition by L-phenylalanine and L-tyrosine, and that this is mediated by the PDH/PDT domain rather than the CM domain.[16-17]

The gateway from AAAB into bioactive phenylpropanoid biosynthesis is the generation of hydroxycinnamic acids. In plants, this is primarily carried out by the conversion of L-phenylalanine to cinnamic acid catalyzed by a phenylalanine ammonia lyase (PAL), followed by the hydroxylation of cinnamate to give *p*-coumarate. A few bacterial genomes contain genes encoding PALs and/or tyrosine ammonia lyases (TALs) and can therefore synthesize hydroxycinnamic acids and some further modified phenylpropanoid compounds (e.g. [18-20]). All vascular plants, on the other hand, are able to synthesize a diverse array of compounds from the hydroxycinnamic acid precursor coumarate (Scheme 1.1). Ortho-hydroxylated phenylpropanoids lead to the biosynthesis of the benzopyrone compounds known as coumarins, while para-hydroxylation of cinnamate to *p*-coumarate by a cinnamate-4-hydroxylase (C4H) and activation to *p*-coumaroyl-CoA by a 4-coumarate-CoA ligase (4CL) leads to the majority of



Scheme 1.1. Overview of plant phenylpropanoid biosynthesis. The key branch point leading to a wide array of plant phenylpropanoids, *p*-coumaroyl-CoA, is boxed. Solid arrows represent key biosynthetic enzymes; dashed arrows represent several known or unknown enzymatic steps. E4P, erythrose-4-phosphate; PEP, phosphoenolpyruvate; DAHP, 3-Deoxy-D-arabinoheptulosonate 7-phosphate; DAHPS, DAHP synthase; CM, chorismate mutase; PDH, prephenate dehydrogenase; PDT, prephenate dehydratase; PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumaroyl-CoA ligase; HCT, hydroxycinnamoyl transferase; CHS, chalcone synthase; STS, stilbene synthase.

characterized plant phenylpropanoids.[13] Conjugation of *p*-coumarate to another aromatic molecule such as shikimate, quinate, or anthranilate by a hydroxycinnamoyl transferase results in the formation of hydroxycinnamic acid esters and amides, which make up the most abundant category of bioavailable hydroxycinamic acids in human foods.[21-23] Addition of three malonyl-CoA units to *p*-coumarate by a type III polyketide synthase, such as chalcone synthase (CHS), leads to the biosynthesis of flavonoids and stilbenoids, a highly antioxidant group of compounds including tannins, anthocyanins, and aurones.[24] In vascular plants, *p*-coumarate is further hydroxylated, methoxylated and reduced to give monolignols, the monomeric units of the vascular cell wall polymer lignin.[25] Monolignols are oxidized by a single-electron oxidase and either undergo random radical coupling to form lignin[26] or regio- and stereo-selective coupling to give lignans, highly antioxidant and often cytotoxic compounds.[27] Monolignols may also undergo acetylation and further reduction, generating volatile phenylpropenes such as eugenol and chavicol.[28-29] From each of these major classes of compounds, valuable anti-oxidant, anti-inflammatory, anti-cancer, anti-bacterial, or otherwise useful compounds have been isolated and characterized.

4. Engineering Microbes for Plant Phenylpropanoid Biosynthesis

4.1 General principles of microbial metabolic engineering

Since its conception in the mid-1990s, metabolic engineering has emerged as a promising biotechnological means for the production of fuels, commodity chemicals and

high-value medicinal compounds.[30-31] Microbial metabolic engineering of plant biosynthetic pathways involves five major considerations: choice of host, choice of feedstock, selection of enzymes, construction of expression vectors, and pathway optimization. *E. coli* and *S. cerevisiae* are the two most commonly-used hosts, due to their genetic tractability and well-developed fermentation technology, and both have been used as hosts for plant natural product biosynthesis, including for phenylpropanoids.[32] An attractive feature of microbial biosynthesis is the ability for microbes to turn inexpensive feedstocks into value-added products, and much success has come from the engineering of the mevalonate pathway in *E. coli* and *S. cerevisiae* to allow for production of medicinal isoprenoids from simple sugars.[33] In some cases, however, the efficient conversion of sugars into natural product precursors has not been achieved, and feeding of inexpensive precursors downstream of the host's central metabolism provides an alternative means of production.[32] While the current sequencing revolution has caused the number of available enzymes for biosynthesis to skyrocket, care must be taken when selecting biocatalysts. Certain enzymes can be problematic in microbes, especially plant enzymes. For example, plant microsomal cytochrome P450s, which are key enzymes in the modification of bioactive products, require extensive expression optimization in microbial hosts.[34] Drawing enzymes from microbial sources, as well as "mix-and-match" combinatorial biochemistry using enzymes from several hosts in one pathway, can circumvent issues with problematic enzymes and even expand the available range of products from an engineered

pathway.[35] Additionally, advances in enzyme engineering have allowed the possibility of tailoring enzymes specifically for use in microbial heterologous biosynthetic pathways.[36] Once the desired enzymes have been identified, the construction of heterologous expression plasmids can be achieved through a number of recently-developed, high-throughput methods for DNA assembly techniques.[37] Finally, after building a biosynthetic pathway, production optimization can be carried out through the alteration of enzyme expression and/or precursor pathway optimization. One recently-developed method for pathway optimization is multivariate modular metabolic engineering (MMME), in which discrete synthetic operon modules are simultaneously varied in copy number, promoter, ribosome binding site, etc. and screened for optimal pathway flux.[38] Additionally, the importance of the spatial organization of biosynthetic enzymes is becoming apparent, and techniques are being developed to control the localization of heterologous enzymes in microbial hosts.[39] When some or all of these techniques have been streamlined, significant progress has been made toward the microbial biosynthesis of a number of valuable plant natural products.

4.2 Review of engineered plant phenylpropanoid pathways in microbes

The vast majority of reports on the microbial biosynthesis of plant phenylpropanoids have involved the biosynthesis of flavonoid and stilbenoid compounds. The first such study was published in 2003, in which *E. coli* overexpressing a plant PAL and CHS and a bacterial 4CL were fed L-tyrosine and L-phenylalanine and catalyzed their conversion to naringenin chalcone and pinocembrin chalcone, respecitvely.[40] Soon after, Watts et

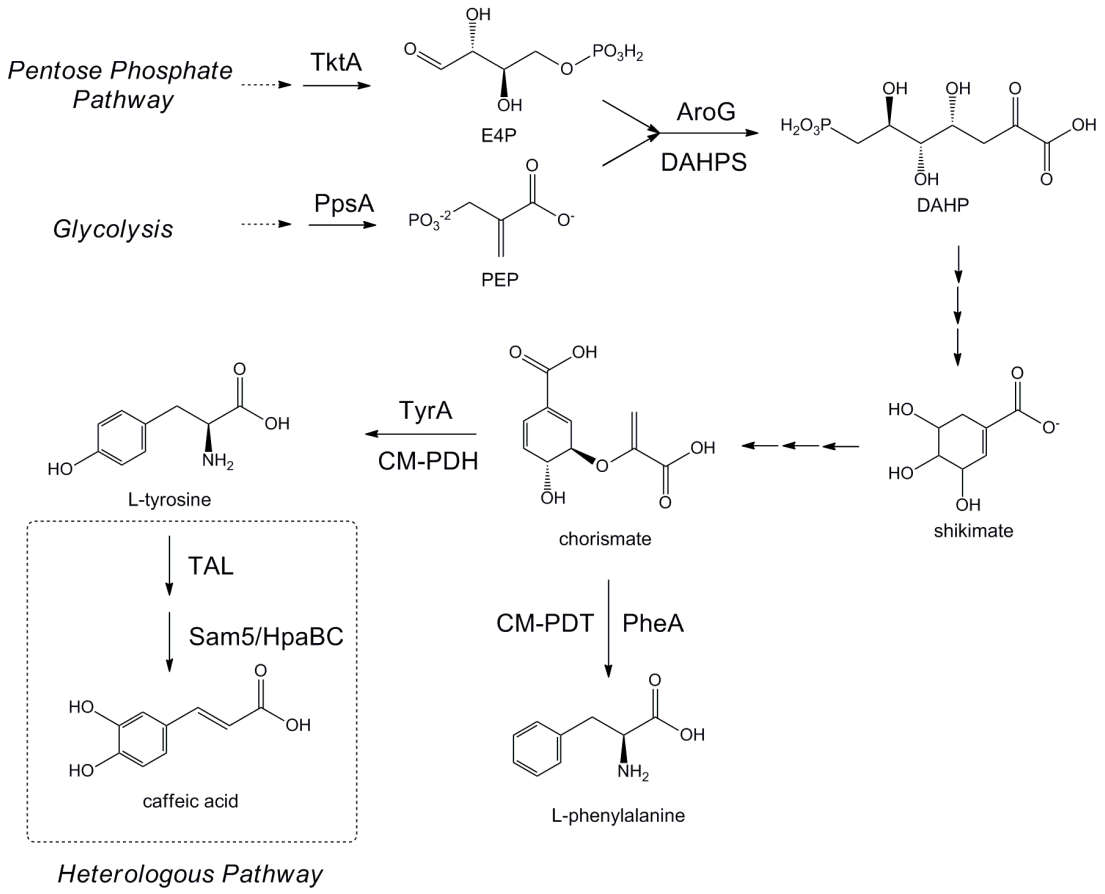
al. showed the production of narigenin chalcone from fed L-tyrosine in *E. coli* through the co-overexpression of a bacterial TAL and a plant 4CL and CHS.[11] Since then, many studies have been published showing the production of naringenin (Scheme 1.1) using similar strategies in both *E. coli* and *S. cerevisiae*, mostly through the feeding of phenylpropanoid precursors (reviewed in [32]). The same strategy has also been used for the production of stilbenoids such as resveratrol (Scheme 1.1), by replacing the CHS with a stilbene synthase (STS).[12, 41] Through the course of these studies, it has become apparent that the limiting factor in microbial flavonoid biosynthesis, particularly in *E. coli*, is the limited malonyl-CoA pool within the cell.[32] To circumvent this, some studies have increased the malonyl-CoA pool through overexpression of acetyl-CoA carboxylase and acetyl-CoA synthase and knockout of competing pathway enzymes, leading to a significant increase in flavonoid production.[42-43] Still, the need to feed phenolic precursors such as aromatic amino acids or hydroxycinnamic acids is a limiting factor for flavonoid biosynthesis. Recently, two studies have reported the *de novo* biosynthesis of flavanoids from glucose in *E. coli* and *S. cerevisiae* engineered for increased L-tyrosine metabolism, reaching up to ~300 mg/L naringenin in the yeast system.[44-45] While these levels are not yet sufficient for industrial applications, the last decade has seen great strides in improvement of stilbenoid and flavonoid biosynthesis in microbes.

Relatively recently, there has been an increased interest in the microbial biosynthesis of hydroxycinnamic acids and their esters and amides, due to the evidence that this class of compounds makes up a large percentage of dietary anti-oxidant and

anti-inflammatory compounds.[21-23, 46] The *de novo* biosynthesis of caffeic acid has been achieved in *E. coli* through overexpression of a bacterial TAL and a bacterial flavin monooxygenase Sam5 from *Saccharothrix espanaensis* capable of carrying out the 3-hydroxylation of *p*-coumarate (Scheme 1.2).[47-48] The additional co-expression of a caffeic acid *o*-methyltransferase from *Arabidopsis thaliana* in this system led to the biosynthesis of ferulic acid, an important precursor to monolignol biosynthesis.[47, 49] It was also recently shown that *E. coli*'s native hydroxyphenylacetate-3-hydroxylase complex, HpaBC, is able to hydroxylate *p*-coumarate at the 3 position, and that its co-overexpression with a TAL in *E. coli* leads to the *de novo* biosynthesis of caffeic acid (Scheme 1.2).[50] Because most bioavailable plant hydroxycinnamic acids exist as esters and amides, and because of the potential pharmaceutical applications of some of these conjugates, the microbial biosynthesis of such compounds has also been pursued.[23, 46] The production of hydroxycinnamoyl shikimates and quinates, including chlorigenic acid (Scheme 1.1), the ester of caffeic acid and quinate, and hydroxycinnamoyl amide tyramine derivatives was first achieved in *E. coli* through feeding hydroxycinnamic acid precursors and overexpression of plant 4CL and hydroxycinnamoyl transferase enzymes.[51-53] The more recent advances in *de novo* hydroxycinnamic acid biosynthesis in *E. coli* have allowed the possibility of synthesizing hydroxycinnamoyl esters and amides from glucose. Eudes et al. recently showed the *de novo* biosynthesis of the hydroxycinnamoyl anthranilate avenanthramide D (Scheme 1.1) to 77 mg/L in an L-tyrosine-overproducing strain of *E. coli*.[54] On a similar scale, the *de novo* biosynthesis

of chlorigenic acid was recently achieved to 78 mg/L in *E. coli* through similar strategies.[55] While still in its early stages, the development of microbial biosynthetic platforms for hydroxycinnamic acids and their esters and amides shows promise, and has benefited from previous research on microbial flavonoid and stilbenoid biosynthetic platforms.

Within just the last two years, the plant phenylpropanoid biosynthetic capabilities of *E. coli* have been further expanded to include additional compound classes. For example, the biosynthesis of medically-important plant-specific coumarin compounds (Scheme 1.1) has been shown through the co-overexpression of caffeic acid biosynthetic enzymes with a plant caffeic acid ortho-hydroxylase.[56-57] Additionally, co-overexpression of genes from creosote bush encoding a cinnamoyl alcohol acyltransferase with either an allylphenol synthase or a propenylphenol synthase led to the production of eugenol (Scheme 1.1) or isoeugenol, important flavorants and antiseptics, from fed monolignol precursors.[58] Finally, the bioconversion of the lignan (+)-pinoresinol to matairesinol, a precursor to the highly-valuable cancer pro-drug podophyllotoxin, has been achieved in *E. coli* through the overexpression of a fusion of the pinoresinol-lariciresinol reductase and secoisolariciresinol dehydrogenase (PLR-SDH) enzymes from *Podophyllum pleianthum*, a plant that naturally produces podophyllotoxin.[59] Clearly, the last decade of research on the biosynthesis of plant phenylpropanoids has laid the groundwork for a plethora of valuable compounds to eventually be synthesized from simple sugars in industrially-relevant organisms.



Scheme 1.2. Aromatic amino acid biosynthesis (AAAB) in *E. coli* as a precursor to plant phenylpropanoid production. AroG and TyrA are sensitive to feedback inhibition by L-tyrosine. An example of a recombinant biosynthetic pathway for a plant-specific phenylpropanoid is boxed. L-tyrosine overproduction has been achieved by overexpression of PpsA (PEP synthase) and TktA (transketolase), overexpression of feedback resistant mutants of AroG and TyrA, and knockout of the competing pathway enzyme PheA. E4P, erythrose-4-phosphate; PEP, phosphoenolpyruvate; DAHP, 3-Deoxy-D-arabinoheptulosonate 7-phosphate; DAHPS, DAHP synthase; CM, chorismate mutase; PDH, prephenate dehydrogenase; PDT, prephenate dehydratase; TAL, tyrosine ammonia lyase.

4.3 Metabolic engineering of the shikimate pathway for increased phenylpropanoid production

In the last decade, alongside advancements in the microbial biosynthesis of phenylpropanoids through heterologous expression of plant pathways, researchers have discovered a number of ways to alter the native shikimate pathway of microbes for increased production of critical phenylpropanoid precursors, especially L-tyrosine (reviewed in [16]). The most prevalent method has been to deregulate the AAAB pathway, which is tightly controlled *via* transcriptional regulation by TyrR and feedback inhibition of DAHP synthase (AroG in *E. coli*) and the PDT domain of CM-PDT (TyrA in *E. coli*) by L-tyrosine. Knockout of the gene encoding TyrR and/or the overexpression of feedback-resistant mutants or chimeras of AroG and TyrA has consistently been shown to increase production of L-tyrosine and downstream phenylpropanoid products (Scheme 1.2).[17, 45, 48-49, 55, 60-61] One group took a unique approach to AAAB deregulation by knocking out the post-transcriptional carbon storage regulator CsrA, which down-regulates PEP carboxykinase and PEP synthetase translation, leading to an increase in L-phenylalanine production.[62] In addition to AAAB deregulation, many studies have increased phenylpropanoid production through the overexpression of PEP synthetase (PpsA) and transketolase (TktA), the enzymes giving rise to PEP and E4P and feeding the shikimate pathway (Scheme 1.2).[50, 54, 56, 60-61] Because the PEP:carbohydrate phosphotransferase system (PTS) competes with the shikimate pathway for available PEP, some studies have used PTS^- strains for the production of

aromatic compounds.[61, 63] In an alternative approach to this problem, one group has evolved a 2-keto-3-deoxy-6-phosphogalactonate aldolase to convert pyruvate and E4P into DAHP, therefore eliminating the need to compete with PTS for available PEP.[64] In some cases, when L-tyrosine is specifically needed as a phenylpropanoid precursor, the competing pathway towards L-phenylalanine has been blocked through knockout of the gene encoding CM-PDH (PheA) in *E. coli*, leading to an increase in L-tyrosine derived products (Scheme 1.2).[45, 48, 55] In addition to the above rational approaches, random and combinatorial genetic approaches have been used to boost L-tyrosine biosynthesis through the development of a high-throughput screen for L-tyrosine.[65-67] These techniques have collectively proven to be powerful tools for the development of economically-viable production of aromatic amino acids and their derivatives, with L-tyrosine production levels reaching as high as 13.8 g/L.[67] One notable example was the recent *de novo* biosynthesis of hydroxycinnamoyl anthranilates in *E. coli* by Eudes et al. (2013).[54] AAAB deregulation through TyrR knockout and overexpression of feedback-resistant AroG and TyrA, as well as overexpression of rate-limiting enzymes in the shikimate pathway led to an increase of more than 100-fold in avenanthramide D biosynthesis.[54] These advances have illustrated the utility of *E. coli* as a host for heterologous plant phenylpropanoid biosynthetic pathways and have made significant strides toward the application of these platforms on an industrial scale.

5. Metabolons in Microbial Natural Product Biosynthesis

5.1 The importance of spatial organization in metabolic pathways

In the past, scientists have simplified metabolism by looking at cells as "bags of enzymes," and much of the study of biochemistry and enzymology has mimicked this notion through *in vitro* enzyme characterization. However, the natural cellular environment paints a very different picture of metabolism in which highly-ordered spatial organization of enzymatic pathways leads to fine-tuned, efficient biocatalysis. Eukaryotes are well-known to organize metabolic pathways within organelles or tethered to lipid membranes. Even the cells of prokaryotes such as *E. coli*, which we think of as having no organelles, has been shown to be 20 – 30% macromolecules by volume, [68] which significantly slows the rate of small molecule and metabolic enzyme diffusion.[69] In such an environment, colocalization of enzymatic pathways into so-called "metabolons" confers the advantages of metabolic channeling of pathway intermediates.[39] Nature has taken a number of routes to evolving metabolons, including the creation of multi-functional enzymes, formation of multi-enzyme complexes, and localization to a sub-cellular organelle or compartment.[39] Having enzyme active sites in close proximity reduces the required fusion distance of intermediates, making them more likely to undergo a reaction before diffusing away. In the case of the *Salmonella typhimurium* L-tryptophan synthetase channel, this limitation of indole diffusion results in a one or two order of magnitude increase in catalytic efficiency.[70] In addition to keeping local concentrations of intermediates high, metabolic channeling keeps cellular

concentrations low, which can prevent toxicity of harmful intermediates. This is a distinct advantage conferred by bacterial microcompartments (BMCs) which encapsulate enzymes and toxic aldehyde intermediates involved in ethanolamine and 1,2-propanediol utilization in a number of enteric bacteria.[71] Furthermore, metabolons often provide a means of metabolic regulation and the prevention of cross-talk with other pathways. For example, clusters of purine biosynthetic enzymes assemble into purinosomes in mammalian cells, but dissipate when the demand for purines decreases.[72] It is believed that secondary metabolism is also dependent on the controlled spatial arrangement of enzymes, usually involving dynamic, low-affinity protein-protein interactions. There is strong evidence for metabolic channeling in plant natural product biosynthesis, including phenylpropanoid biosynthesis, which involves the interaction of several enzymes with membrane-bound cytochrome P450s.[73] Based on this knowledge, it is crucial to understand and, if possible, manipulate the spatial arrangement of enzymes when building heterologous biosynthetic pathways in microbes.

5.2 Synthetic metabolons for microbial production platforms

Since early in the history of microbial metabolic engineering, researchers have attempted to increase the flux through biosynthetic pathways by creating synthetic metabolons. The most common way to do this has been through the creation of fusion proteins. In most cases, the fusion of two or more biosynthetic enzymes has resulted in an increase in catalytic efficiency and/or metabolic flux through a pathway (reviewed in [39]). In a recent example, the percent bioconversion of pinoresinol to matairesinol by *E.*

coli expressing PLR and SDH enzymes from *P. pleianthum* was three fold higher when the enzymes were expressed as a PLR-SDH fusion.[59] Interestingly, the creation of fusion proteins has also been the result of a non-rational strain engineering endeavor. Meynial et al. evolved *E. coli* containing a plasmid with two open reading frames for glycerol biosynthesis, and the final selected high-titer production strain showed a deletion between the ORFs resulting in an in-frame fusion of the two proteins.[74] While creating a fusion protein is a simple way to bring enzymes into close proximity, there are a number of limitations with this approach. First, most biosynthetic enzymes are active as multimeric proteins, and creation of a protein fusion can interfere with oligomer formation.[39] Second, the folding of large multi-domain heterologous proteins can be problematic in microbes and is especially inefficient in bacterial hosts.[75] In fact, while Kuo et al. discovered that the PLR-SDH fusion increased lignan bioconversion, the fusion in the opposite conformation, with SDH at the N-terminus (SDH-PLR), only exhibited PLR activity.[59] In light of these setbacks, there is a need for a means of enzyme colocalization that can allow enzymes to take their native conformations as much as possible.

In recent years, a new, post-translational approach to creating synthetic metabolons has emerged based on the model of bacterial cellulosomes. Bacterial cellulosomes consist of a long scaffold containing several cohesin domains and several cellulase enzymes that bind to the scaffold *via* dockerin domains. Engineered cellulosomes in *E. coli*, using distinct cohesin-dockerin domain pairs from different

species, leads to an increase in the degradation of recalcitrant cellulose over free cellulase enzymes (reviewed in [39]). Additionally, it was shown that different geometric arrangements of enzymes on the scaffold affects cellulose degradation, suggesting that tight control of enzyme colocalization on synthetic scaffolds could be a valuable tool in biocatalysis.[76] In 2009, this approach was expanded with the development of a completely synthetic scaffold system in *E. coli*, utilizing eukaryotic protein interaction domains to control the number and arrangement of enzymes in a metabolon.[77] Using the three enzymes in the *S. cerevisiae* mevalonate pathway as a proof of concept, Dueber et al. showed that synthetic scaffolding significantly increased pathway flux, and that altering the architecture and enzyme stoichiometry of the metabolon influenced pathway productivity.[77] Since then, studies have used this scaffolding system to optimize the biosynthesis of glucaric acid in *E. coli*[78] and the production of resveratrol in *S. cerevisiae*.[79] Meanwhile, another group developed an alternative means of protein scaffolding, by equipping metabolic enzymes with zinc finger domains designed to bind to specific sequences of DNA. Through overexpression of zinc finger-containing metabolic enzymes in the same cell as a scaffold sequence-containing plasmid, Conrado et al. showed an increase in violacein biosynthesis in *E. coli*.[80] Additionally, Delebecque et al. have developed an RNA-based scaffolding system, using RNA molecules designed with specific aptamer sequences to bind metabolic enzymes, leading to an increase in efficiency of a hydrogen production pathway in *E. coli*.[81] The

use of these scaffold-based engineered metabolons shows great promise for application in the microbial biosynthesis of plant natural products.

An alternative approach to engineering metabolons in bacterial systems is the encapsulation of metabolic enzymes in protein-based shells. In nature, bacteria have evolved the encapsulins and bacterial microcompartments (BMCs) to achieve this. Encapsulins are small compartments made up of a single shell protein that encapsulate ten to twenty copies of a single oxidative stress-related enzyme, such as a peroxidase or ferritin-like protein.[82] A C-terminal cargo-protein motif was identified that is necessary for encapsulation, although it has never been tested with heterologous enzymes.[82] However, other groups have engineered simple charged tags on the interior of the encapsulin shell protein and the C-terminus of the cargo protein and achieved encapsulation of both eGFP[83] and HIV protease.[84] While engineered encapsulins are useful for sequestering single enzymes, particularly enzymes like HIV protease which are toxic to *E. coli*, their applications are limited in metabolic engineering due to the small number of enzymes encapsulated within each.

BMCs, on the other hand, are much larger and have evolved in nature to sequester a large number and variety of enzymes. The most well-known forms of BMCs are the carboxysomes, which sequester the dark reactions in cyanobacteria, and the 1,2-propanediol utilization (Pdu) and ethanolamine utilization (Eut) microcompartments, which encapsulate the enzymes required for catabolism of these compounds in enteric bacteria.[71] BMC shells are made up of multiple proteins containing a conserved BMC

domain (PF00936) which self-assemble into sheets of hexamers that make up an icosahedral shell. Recombinant expression of suites of shell proteins from the *S. enterica* Pdu BMCs,[85] the *Halothiobacillus neapolitaus* CsoS carboxysomes,[86] and the *S. enterica* Eut BMCs [87] has led to the formation of empty BMC shells in *E. coli*. Additionally, short, N-terminal signal peptides have been shown to be required for localization of enzymes into Pdu BMCs,[88-89] and Choudhary et al. have identified a signal peptide both necessary and sufficient for encapsulating recombinant proteins to Eut BMC shells in *E. coli*.[87] Interestingly, it was found that in the Eut BMC system, overexpression of only one BMC domain shell protein, EutS, is required for recombinant shell formation,[87] making Eut BMCs an attractive platform for future engineered metabolons. An interesting potential advantage of BMCs over scaffold-based systems is the fact that different BMC-domain protein hexamers form isotopically distinct pores in the center, opening up the possibility to engineer the selective permeability of BMC shells.[90-91] While BMCs have yet to be utilized successfully in the construction of an engineered metabolon containing multiple enzymes, the use of this system shows great promise for the *in vivo* colocalization of heterologous biosynthetic pathways.

6. Thesis Motivation and Preliminary Work

I was drawn to Dr. Schmidt-Dannert's laboratory by two main interests: plant natural products biosynthesis and microbial engineering. I chose to pursue work related to the bacterial production of bioactive plant phenylpropanoids, drawing from previous work in the lab to establish the biosynthesis of flavonoid and stilbene production in *E. coli*.[11-12] I

attempted to expand the repertoire of plant phenylpropanoids that can be produced in plants by incorporating heterologous biosynthetic pathways for two classes of products that had not yet been produced in *E. coli*. Additionally, I worked to develop recombinant BMCs as a means of directing the spatial organization of enzymes involved in heterologous biosynthetic pathways.

6.1 Lignan biosynthesis in *E. coli*

In 2009, I began work on a project in collaboration with Dr. Jacob Vick to engineer the first microbial strain capable of producing lignans from an inexpensive feedstock. Lignans are valuable for their anti-oxidant and phytoestrogen properties. We were particularly interested in the biosynthesis of podophyllotoxin, an extremely cytotoxic compound and an important precursor to widely-used chemotherapy drugs.[92] At the time this work was begun, none of the intermediates in plant lignan biosynthesis beyond *p*-coumaric acid had been produced in *E. coli*. The ultimate goal of this work was to establish the bioconversion of a cheap feedstock, ferulic acid, to matairesinol, the latest precursor to podophyllotoxin whose biosynthesis has been fully elucidated.[93] The work was broken down into three parts. First, Dr. Vick worked to establish the bioconversion of ferlulic acid into the monolignol coniferyl alcohol through the overexpression of three enzymes from *A. thaliana*. Second, I worked to establish the regio- and stereo-selective coupling of coniferyl alcohol to (+)-pinoresinol in *E. coli* through the expression of a bacterial laccase and a dirigent protein. Finally, I planned to establish the conversion of (+)-pinoresinol by matairesinol through overexpression of genes encoding PLR and SDH

from *A. thaliana*. Because of the difficultly of dirigent protein expression in *E. coli* and the publication of the bioconversion of (+)-pinoresinol to matairesinol in *E. coli* before the completion of this work,[59] I have only reported the study of dirigent-mediated oxidative radical coupling in this thesis (Chapter 2).

6.2 Hydroxycinnamic acid ester biosynthesis in *E. coli*

Because of the consistent difficult of establishing regio- and stereo-specific coupling for lignan biosynthesis in *E. coli*, I decided to pursue a second project focusing on a different class of plant phenylpropanoids: the hydroxycinnamic acid esters (HCEs). These are important dietary anti-oxidant and anti-inflammatory compounds, and one HCE in particular, rosmarinic acid (RA), has been shown in *in vitro* and animal studies to have potent anti-cancer and neuroprotective activities.[23, 94] While the bioconversion of hydroxycinnamoyl amides from fed hydroxycinnamic acid precursors had previously been established in *E. coli*,[52-53] the production of HCEs had not yet been published. In plants, the meta-hydroxylation of the phenylpropanoid moieties in RA biosynthesis is carried out by microsomal cytochrome P450s after ester formation,[23] posing a significant obstacle to the *de novo* biosynthesis of RA in plants. However, the beginning of this work in late 2011 coincided with the first report of caffeic acid biosynthesis in *E. coli*,[47] which allowed me to redesign the biosynthetic pathway to carry out meta-hydroxylation prior to ester formation using a bacterial enzyme. I used this approach to design and assemble a *de novo* biosynthetic pathway for RA specifically tailored to *E. coli*, utilizing bacterial enzymes wherever possible and shortening the biosynthetic route.

6.3 Study of enzyme encapsulation within recombinant Eut BMCs

After the laboratory's publication of recombinant protein encapsulation within recombinant Eut BMC shells,[87] Dr. Schmidt-Dannert's group has focused on understanding recombinant Eut BMC biogenesis and developing tools for the creation of synthetic metabolons within this system. I discovered cross-talk leading to undesired products in my engineered RA biosynthetic pathway and devised a model for how partial encapsulation of the pathway within a BMC shell could prevent such cross-talk. Therefore, I carried out preliminary experiments to attempt localization of RA biosynthetic enzymes to engineered Eut BMCs using the $EutC^{1-19}$ signal peptide discovered by Choudhary et al.[87] This led me to conceptualize and begin work on a study of the mechanism of enzyme encapsulation within recombinant Eut BMCs, with the ultimate goal of understanding how to fine-tune control of multiple enzyme encapsulation. This work is being completed in collaboration between myself, Dr. Mark Held and Suzie Hsu, and will be continued by Dr. Sarah Perdue and Kelsey Dahlgren upon my graduation.

Chapter 2. Towards Microbial Lignan Biosynthesis: Exploring Oxidative Radical Coupling in *Escherichia coli*.

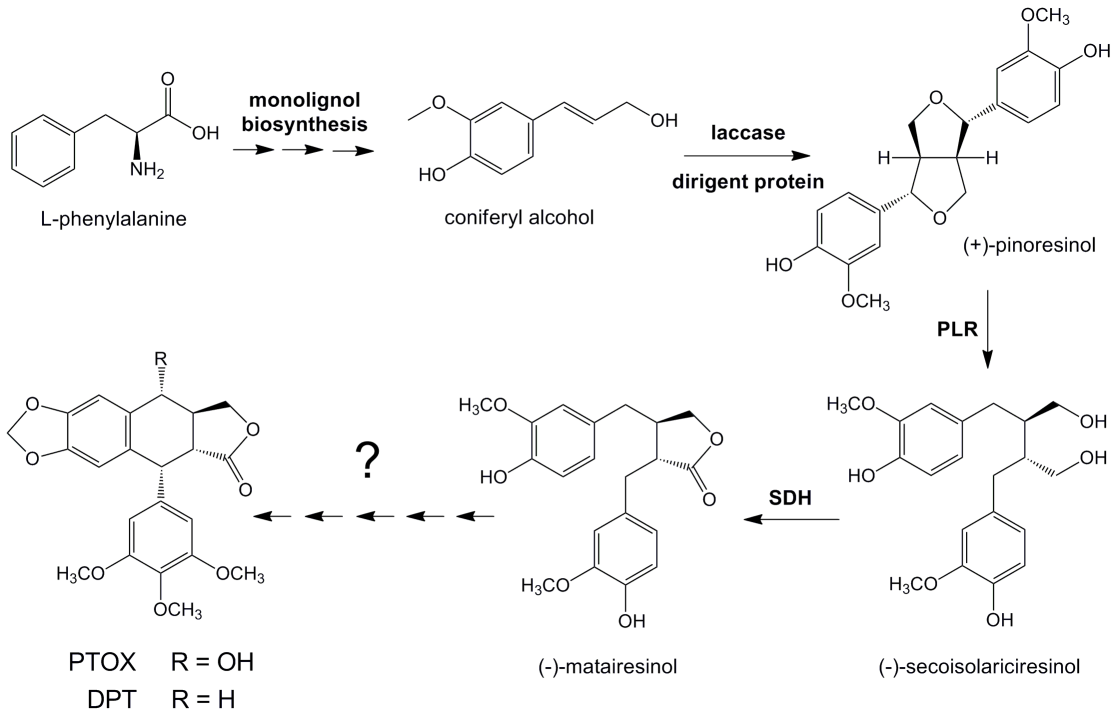
Summary

Lignans are bioactive plant phenylpropanoids derived from monolignols. The most commercially important lignan, podophyllotoxin (PTOX), is in high demand as a chemotherapy pro-drug. Due to the depletion of source plant populations and the infeasibility of chemical synthesis, a microbial production platform would be extremely valuable. Lignan biosynthesis begins with the regio- and stereo-selective coupling of two coniferyl alcohol radicals to form (+)-pinoresinol, which is driven by molecular chaperones called dirigent proteins. In this work, I aimed to establish (+)-pinoresinol production in *E. coli* by coexpressing a dirigent protein and a laccase that can generate coniferyl alcohol radicals. I showed that SLAC, a small laccase from *Streptomyces coelicolor*, is able to oxidize coniferyl alcohol and give racemic pinoresinol in *E. coli*. I was also able to establish the soluble expression of a well-characterized dirigent protein from *Forsythia intermedia*, FiDRP1, through co-overexpression with *E. coli* protein-folding chaperones. Ultimately, I was unable to show function of the *E. coli*-expressed FiDRP1. Meanwhile, another group has shown that post-translational modifications are essential for dirigent protein function. This suggests that *E. coli* is not a viable host for lignan biosynthesis, and that a eukaryotic host may be more feasible.

1. Introduction

Lignans are class of polyphenolic plant natural products known for their potent biological activities. Derived from monolignols, the same monomeric units that make up lignin, lignans are widespread among vascular plants and make up a significant portion of important dietary antioxidants.[95] Lignans are also phytoestrogens – food-derived compounds whose metabolites mimic estrogens in the human body – which have been shown to be important as preventatives of cardiovascular disease based on their antioxidant and anti-inflammatory properties, as well as their positive effects on vascular smooth muscle, endothelial cells, and extracellular matrix.[96-97] Some lignans and their derivatives are also important pharmaceuticals. Podophyllotoxin (PTOX), a lignan found in the roots of plants the genus *Podophyllum* (may apples), and its close analog deoxypodophyllotoxin (DPT) are microtubule inhibitors that exhibit strong anti-mitotic activity, making them potent anti-tumor and anti-viral treatments (Scheme 2.1).[93, 98] While PTOX is too cytotoxic to be used for chemotherapy, three of its derivatives – etoposide, teniposide and etopophos – are widely used in the chemotherapeutic treatment of several cancers such as leukemia, lymphoma, testicular cancer, and small-cell lung cancer.[93]

The primary means of producing PTOX remains the same as it has for centuries: extraction from plants which naturally produce the compound. Unfortunately, plants of the species *Podophyllum* are not easily cultivated, and wild sources of the plants – especially the population of *P. hexandrum* in the Himalayas – are being depleted.[99]



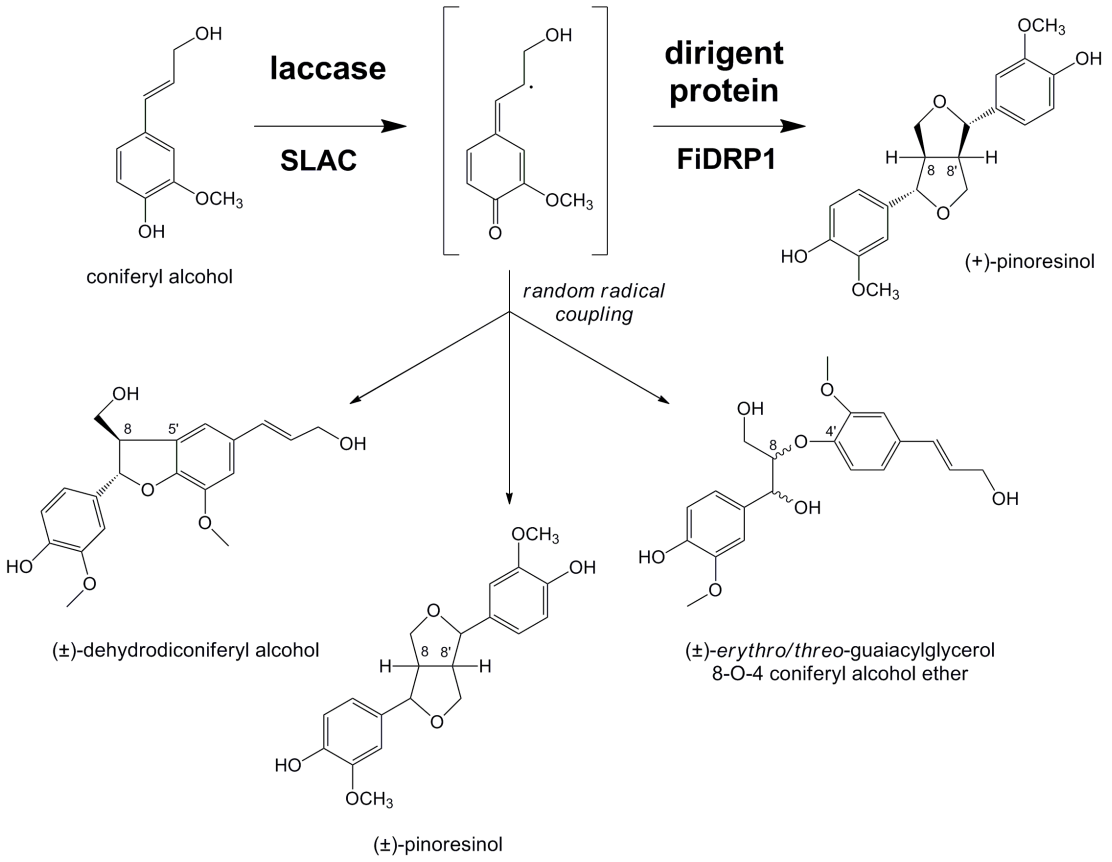
Scheme 2.1. Biosynthesis of valuable lignans PTOX and DPT in plants. The monolignol coniferyl alcohol is derived from L-phenylalanine through a well-characterized biosynthetic pathway. A single electron oxidase, likely a laccase, and a molecular chaperone called a dirigent protein catalyze the regio- and stereo-specific coupling of two coniferyl alcohol radicals to (+)-pinoresinol. Two enzymes, pinoresinol-lariciresinol reductase (PLR) and secoisolariciresinol dehydrogenase (SDH) catalyze the conversion of (+)-pinoresinol to (-)-matairesinol. Finally, several undiscovered enzymatic conversions lead to the biosynthesis of the valuable lignans podophyllotoxin (PTOX) and deoxypodophyllotoxin (DPT).

Meanwhile, the steadily increasing demand for PTOX necessitates the development of a sustainable means of mass production. Four chiral centers in the PTOX molecule make synthesis economically impossible due to low yields.[92] Some groups have reported the production of PTOX in plant cell and tissue cultures; however, the yields are insufficient for this method to be implemented on an industrial scale.[99] Given all of these setbacks, a microbial production platform for PTOX and related lignans could be an attractive

option, allowing for the inexpensive, renewable production of these value-added lignans with the desired stereochemistry and precise regulatory control. While the biosynthetic pathway of PTOX has not yet been fully elucidated in plants,[92] enough is understood to begin the lay the groundwork for a heterologous biosynthetic pathway in microbes.

The biosynthesis of PTOX in plants begins with the biosynthesis of the monolignol coniferyl alcohol *via* the shikimate and aromatic amino acid pathways (Scheme 2.1). In lignin biosynthesis, coniferyl alcohol and other monolignols undergo single-electron oxidation by a laccase or peroxidase and are randomly incorporated into growing lignin polymers through oxidative radical coupling. The first step in lignan biosynthesis, however, requires the controlled regio- and stereo-selective coupling of two monolignol radicals – in the case of PTOX, two coniferyl alcohol radicals are linked at the 8 – 8' positions to form (+)-pinoresinol. The mechanism of this selective oxidative radical coupling was first elucidated by Davin et al., who showed that this process depends on a molecular chaperone called a dirigent protein to guide the proper alignment of two monolignol radicals and give a specific product (Scheme 2.2).[100] Dirigent proteins have no catalytic center and are unable to oxidize monolignols; rather, they are thought to bind free monolignol radicals in a specific conformation to facilitate controlled coupling.[101-102] After oxidative radical coupling, the enzymes pinoresinol-lariciresinol reductase (PLR) and secoisolariciresinol dehydrogenase (SDH) convert (+)-pinoresinol stereoselectively to (-)-matairesinol, which is subsequently converted to PTOX in four or five unknown enzymatic steps (Scheme 2.1).

The most critical step in the creation of a microbial biosynthetic platform for valuable (+)-pinoresinol-derived lignans such as PTOX is to establish controlled oxidative radical coupling in a widely-used microbial host. To this end, we have explored the possibility of engineering the regio- and stereo-selective coupling of coniferyl alcohol



Scheme 2.2. Mechanism of stereo- and regio-specific oxidative radical coupling. After oxidation by a laccase or peroxidase, in the absence of a dirigent protein, coniferyl alcohol undergoes random coupling to give racemic mixtures of 8-5', 8-8' and 8-O-4' linked dimers. The dirigent protein binds free coniferyl alcohol radicals, orients them for stereospecific 8-8' linkage, and leads to the production of an enantiomeric excess of (+)-pinoresinol. SLAC, small laccase from *Streptomyces coelicolor*; FiDRP1, characterized (+)-pinoresinol-forming dirigent protein from *Forsythia intermedia*.

to (+)-pinoresinol in *Escherichia coli*. We first established the expression and *in vivo* function of a bacterial single electron oxidase capable of generating coniferyl alcohol radicals. Then, we explored the expression and function of two dirigent proteins in *E. coli*, one dirigent from *Forsythia intermedia* well-characterized for (+)-pinoresinol formation and another putative dirigent protein from *A. thaliana*. We found the expression of dirigent proteins in *E. coli* to be extremely problematic, and were unable to establish the function of the heterologously-expressed (+)-pinoresinol-forming dirigent protein. Ultimately, this and recent work published by other groups have illustrated the impossibility of establishing dirigent protein function in *E. coli*. For future engineering of lignan biosynthetic pathways in microbes, we recommend the use of a well-understood eukaryotic host instead.

2. Results

2.1 Oxidative coupling of coniferyl alcohol to racemic pinoresinol in *E. coli* by SLAC, a small laccase from *Streptomyces coelicolor*

The first component required to achieve stereoselective oxidative radical coupling is an enzyme capable of single-electron oxidation of monolignol precursors. In plants, this is believed to be carried out by one or more nonspecific laccase or peroxidase enzymes.[27, 100] Laccases are multi-copper single electron oxidases that can oxidize a broad range of phenolic compounds, and have received much attention in recent years for their current and potential applications in the textile industry,[103] bioremediation[104-105] and

biocatalysis,[106] with fungal laccases being the most abundant and well-characterized.[107] However, because the expression of fungal enzymes can be problematic in a bacterial host, we desired a bacterial laccase for use in biosynthetic applications in *E. coli*. A small laccase from *Streptomyces coelicolor*, called SLAC, had previously been expressed in *E. coli* and shown to have remarkable stability and high activity at neutral pH, unlike many fungal laccases.[108] We hypothesized that SLAC would be able to oxidize coniferyl alcohol and lead to the production of racemic pinoresinol, among other di- and polyphenol compounds. We therefore amplified the gene encoding SLAC from *S. coelicolor* genomic DNA and cloned it into the plasmid pUCmodRBS, a high-copy plasmid with a modified constitutive *lac* promoter, in frame with a C-terminal His-tag. Transformation of *E. coli* BW27784 cells with pUCmodRBS-SLAC led to the overexpression of the SLAC enzyme, which was easily enriched *via* Ni^{2+} affinity purification (Figure 2.1). Two predominant bands are visible in the SLAC purification: a band at ~37 kDA, corresponding to the predicted size of the SLAC monomer, and a band at ~75 kDA, which is roughly the size of dimerized SLAC. Indeed, the previous characterization of this enzyme showed the presence of active, dimeric SLAC under the denaturing conditions of SDS-PAGE, speaking to the extreme stability of this enzyme.[108] Incorporation of copper ions was facilitated by dialysis of the purified enzyme in a copper-containing buffer.

Having purified SLAC, we initially tested it for activity with coniferyl alcohol through an *in vitro* assay. SLAC was incubated with 4 mM coniferyl alcohol at 30 °C,

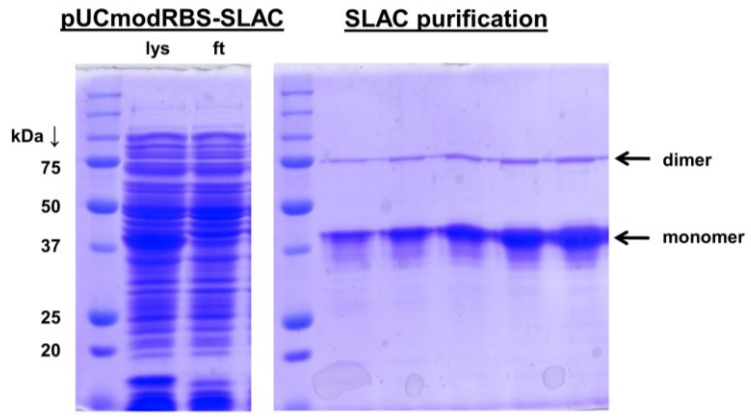


Figure 2.1. Expression and purification of SLAC from *E. coli* BW27784 cells. Cells were transformed with pUCmodRBS-SLAC, cultured to mid-log phase at 37 °C, and moved to a 20 °C shaker and incubated for 20 h before harvesting and lysis. Purification was carried out *via* FPLC on a 5-mL Ni^{2+} column. Lys, lysate; ft, flowthrough.

and samples were taken over time, extracted with ethyl acetate and analyzed *via* HPLC. We observed the depletion of coniferyl alcohol from the reaction over time, as well as the appearance of a peak corresponding to pinoresinol within 20 min, suggesting that SLAC is indeed able to oxidize coniferyl alcohol (Figure 2.2). Several other peaks, which likely correspond to other di- and poly-lignol coupling products, were observed. Coniferyl alcohol completely disappeared within 120 min reaction time; however, the abundance of pinoresinol and the other products remains stable between 40 and 60 min, and even seems to have decreased at 120 min. This is likely due to the non-specific nature of laccases – SLAC may have been further oxidizing the dimeric lignan products further to higher-order polyphenols, which were then lost in the extraction process. In fact, we observed a white solid that remains at the interface of the aqueous and organic phases

during ethyl acetate extraction which appears to increase in abundance over time, which was likely due to the formation of large, polyphenol molecules.

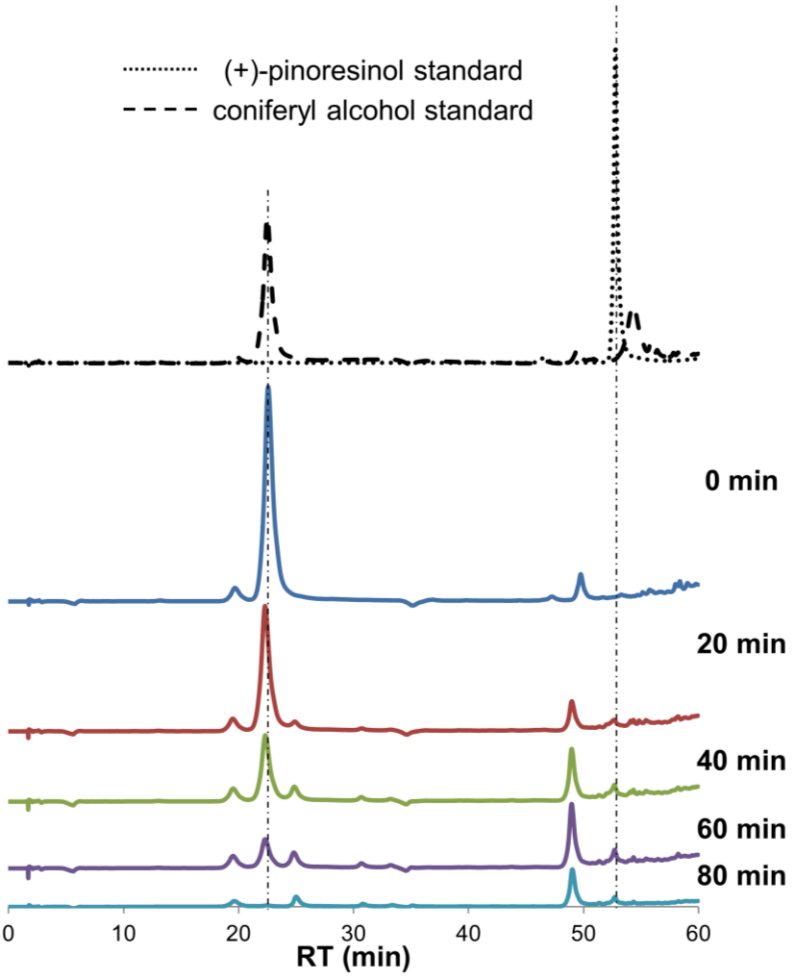


Figure 2.2. In vitro oxidation of coniferyl alcohol to pinoresinol and other products by SLAC. Purified SLAC (20 µg/mL) was incubated with 4 mM coniferyl alcohol in 100 mM Tris pH 8.0 at 30 °C with shaking. Every 20 min, samples were taken and immediately extracted with ethyl acetate for subsequent HPLC analysis. Vertical dashed lines indicate the retention times of coniferyl alcohol and pinoresinol peaks, as compared to analytical standards.

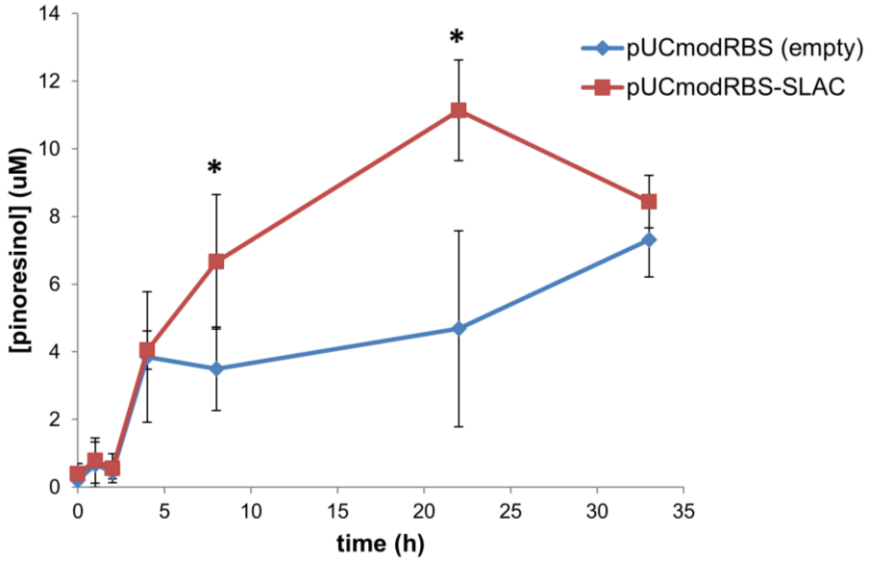


Figure 2.3. Oxidation of coniferyl alcohol to pinoresinol by *E. coli* BW27784 cells expressing SLAC. *E. coli* BW27784 cells were transformed with pUCmodRBS (empty) or pUCmodRBS-SLAC. Transformants were cultured in modified minimal media at 30 °C; 3 h after inoculation, cultures were supplemented with 100 µM coniferyl alcohol and 100 µM copper sulfate. Culture samples were taken over time, cleared of cells, and extracted with ethyl acetate for HPLC analysis. An analytical standard of pinoresinol was used to confirm peak identity and quantification. Asterisks denote statistical significance at $p < 0.05$

Next, to gauge SLAC's potential utility in microbial lignan biosynthesis, we tested whether heterologous expression of SLAC allowed *E. coli* to convert exogenous coniferyl alcohol to pinoresinol. *E. coli* BW27784 were transformed with pUCmodRBS-SLAC or empty pUCmodRBS as a negative control and cultured in modified minimal media supplement 100 µM coniferyl alcohol. Because initial screens using crude *E. coli* cell lysates containing SLAC gave little detectable coniferyl alcohol oxidation (data not shown), these cultures were also supplemented with copper sulfate to ensure

incorporation of copper into the active site of SLAC. Ethyl acetate extraction and HPLC analysis of the culture media showed the accumulation of up to 11 ± 1 μM pinoresinol in cells expressing SLAC (Figure 2.3). Surprisingly, a peak at the retention time of pinoresinol (53 min) was also detected in cells transformed with empty vector. However, at 8 h and 24 h, cultures transformed with pUCmodRBS-SLAC exhibit a significantly higher apparent concentration of pinoresinol than cultures transformed with empty vector ($p < 0.05$), which suggests that SLAC expression leads to the oxidation of coniferyl alcohol and the formation of pinoresinol in *E. coli* (Figure 2.3).

2.2 Cloning and expression optimization of a characterized (+)-pinoresinol-forming dirigent protein from *Forsythia intermedia*

The second essential component for lignan bisoynthesis in *E. coli* is the heterologous expression of a functional (+)-pinoresinol-forming dirigent protein. Genes encoding dirigent and dirigent-like proteins are widespread among plants,[102, 109-110] but very few have been characterized biochemically.[111-113] For the specific coupling of coniferyl alcohol to (+)-pinoresinol in our system, we chose the first and most well-characterized (+)-pinoresinol-forming dirigent protein from *Forsythia intermedia*, FiDRP1.[100] An *E. coli* codon-optimized gene encoding FiDRP1 was assembled *via* overlap extension PCR of several synthetic DNA oligomers, excluding an N-terminal cell-wall targeting sequence present in the native protein. This gene was cloned into pUCmodRBS with a C-terminal His-tag and transformed into *E. coli* BL21 (DE3) cells. In an initial screen, expression of

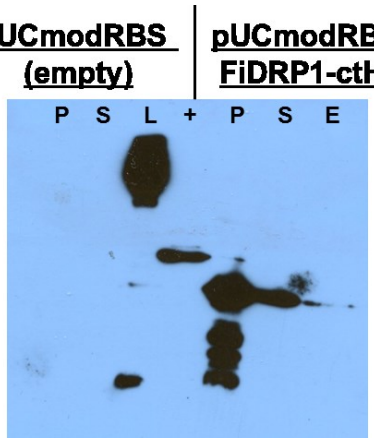


Figure 2.4. Degradation of FiDRP1 with a C-terminal His-tag in *E. coli* BL21 (DE3). Cells were transformed with pUCmodRBS (empty) or pUCmodRBS-FiDRP1-ctH6; transformants were cultured overnight at 30 °C, lysed by sonication, and subjected to anti-His-tag western blotting. The lane marked "L" is a molecular weight ladder; the lane marked "+" contains a 25-kDa His-tagged protein as a positive control. P, pellet; S, soluble fraction; E, eluent from a small-scale batch purification with Talon™ metal affinity resin.

FiDRP1 was not visible *via* SDS-PAGE; however, anti-His-tag western blotting revealed the presence of a band the approximate predicted size of FiDRP1 (~19 kDa) (Figure 2.4). The majority of the observed protein was present in the insoluble fraction, with some protein in the soluble fraction, and only a very faint band in the eluent from a small-scale batch purification with metal affinity resin. Interestingly, there appears to be many smaller bands in the insoluble fraction, suggesting that the insoluble FiDRP1 is being degraded (Figure 2.4). This apparent degradation and low binding to a metal affinity resin suggested misfolding of FiDRP1. At the time of this study, there were no crystal structures or structural models of dirigent proteins; however, we noted a weak homology between dirigent proteins and allene oxide cyclases,[114] whose C-terminus is

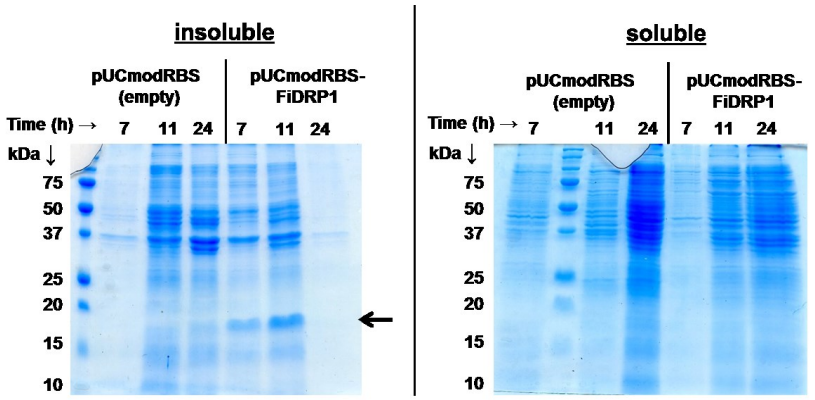


Figure 2.5. Insoluble expression of FiDRP1 with no His-tag in *E. coli* BL21 (DE3). Cells were transformed with pUCmodRBS (empty) or pUCmodRBS-FiDRP1, transformants were cultured at 30 °C, sampled periodically, lysed by sonication, and subjected to SDS-PAGE. The black arrow indicates the band corresponding to FiDRP1 (predicted size: ~18 kDa).

involved in a salt-bridge between protein monomers.[115] Therefore, we thought the C-terminal His-tag of FiDRP1 may be interfering with folding and/or quaternary structure formation and began exploring other expression options.

We first re-cloned FiDRP1 into pUCmodRBS with no His-tag and transformed the construct into *E. coli* BL21 (DE3) cells. While cultures of these transformants exhibited some noticeable expression of FiDRP1, protein was only visible in the insoluble fraction after lysis (Figure 2.5). To achieve higher expression levels, and for ease of future protein purification, FiDRP1 was also cloned into pET15b, which contains an in-frame N-terminal His-tag and thrombin cleavage site. Transformation of *E. coli* with pET15b-FiDRP1 resulted in the very high expression of FiDRP1, but again, all protein was found

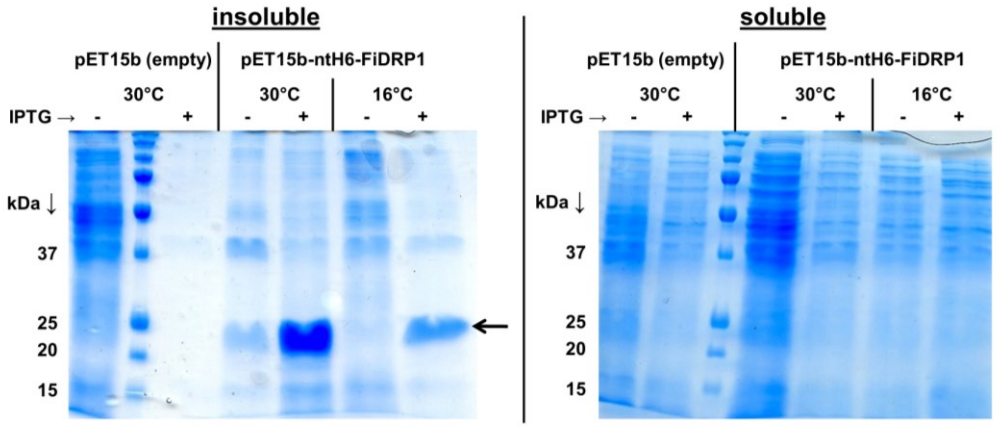


Figure 2.6. Insoluble expression of FiDRP1 with an N-terminal His-tag in *E. coli* BL21 (DE3). Cells were transformed with pET15b (empty) or pET15b-FiDRP1, and transformants were cultured to mid-log phase at 37 °C; then, half of the cultures were induced with IPTG (1mM), and all were moved to 30 °C or 16 °C shakers for 4 h induction. Cells were lysed by sonication, and subjected to SDS-PAGE. The black arrow indicates the band corresponding to FiDRP1 with the His-tag and thrombin cleavage site present in the pET15b vector (predicted size: ~21 kDa).

in the insoluble fraction, even at low temperatures (Figure 2.6). In an attempt to increase the solubility of FiDRP1, we cotransformed pET15b-FiDRP1 with three plasmids for the arabinose-inducible overexpression of native *E. coli* protein folding chaperones to increase the protein-folding capacity of the host cells. One plasmid, encoded the expression of GroEL and GroES, another DnaK, DnaJ and GrpE, and a third Tf (trigger factor). When small-scale cultures of the transformants reached mid-log phase, FiDRP1 and chaperone expression were induced by the addition of IPTG and arabinose, respectively. After 2.5 h induction at 30 °C, only cotransformation with pKJE7, encoding DnaK, DnaJ and GrpE, led to some soluble expression of FiDRP1 that was visible *via* SDS-PAGE (Figure 2.7). Scale up to a 1-L culture and induction at 37 °C

for 2.5 h led to high levels of soluble FiDRP1 expression (Figure 2.8); however, when this cell lysate was used to attempt the purification of FiDRP1, the His-tagged protein repeatedly failed to bind to a metal affinity resin.

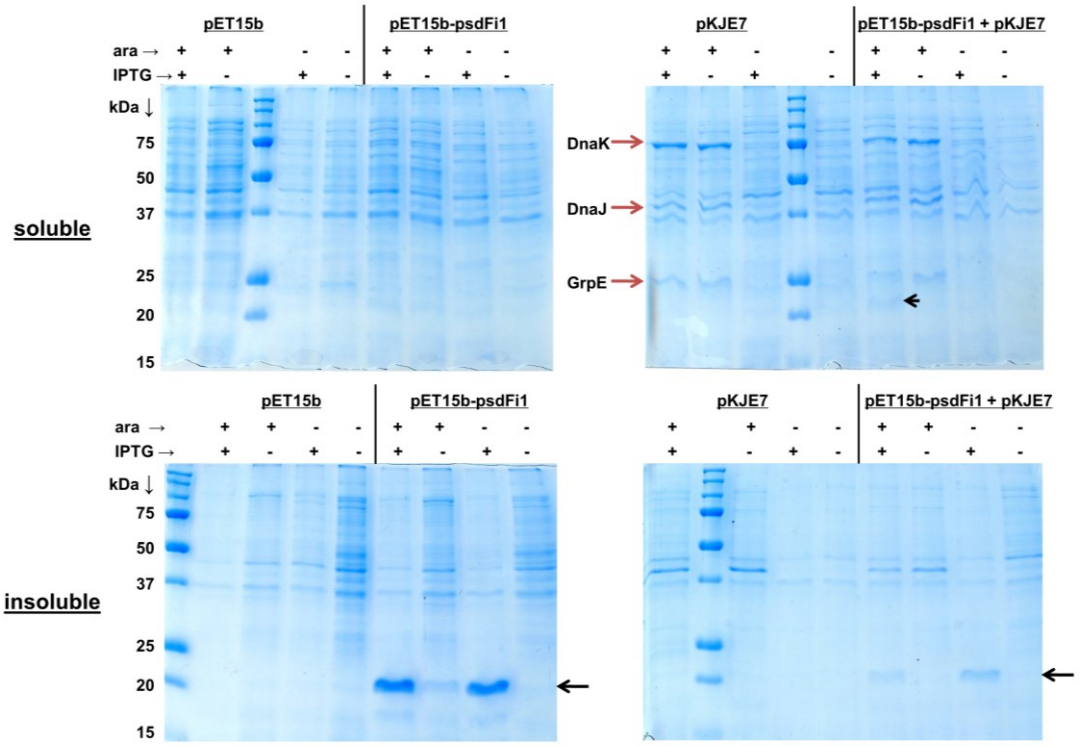


Figure 2.7. Cotransformation *E. coli* BL21 (DE3) with pKJE7 and pET15b-FiDRP1 leads to the soluble expression of FiDRP1. Cells were transformed with pET15b (empty), pKJE7 alone, pET15b-FiDRP1 or pET15b-FiDRP1 and pET15b. Transformants were cultured to mid-log phase at 37 °C, at which point they were induced with either IPTG (1 mM), arabinose (0.2 %w/v), neither, or both, and all were moved to a 30 °C shaker for 2.5 h induction. Cells were lysed by sonication, and subjected to SDS-PAGE. Red arrows indicate the expression of the overexpressed protein-folding chaperones; black arrows indicate bands corresponding to FiDRP1 with the His-tag and thrombin cleavage site present in the pET15b vector (predicted size: ~21 kDa).

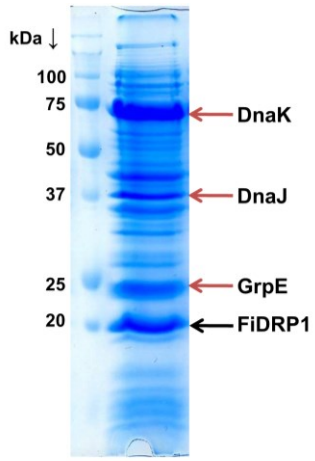


Figure 2.8. High levels of soluble FiDRP1 expression in *E. coli* BL21 (DE3). Transformants of pET15b-FiDRP1 and pKJE7 were cultured to mid-log phase in 1 L LB media at 37 °C, at which point they were induced with both IPTG (1 mM) and arabinose (0.2 %w/v) and incubated at 37 °C for a further 2.5 h. Cells were lysed by sonication, and samples were subjected to SDS-PAGE. Red arrows indicate the expression of the overexpressed protein-folding chaperones; black arrow indicates the band corresponding to FiDRP1 with the His-tag and thrombin cleavage site present in the pET15b vector (predicted size: ~21 kDa).

Although we were unable to purify the His-tagged FiDRP1 protein, we decided to conduct a preliminary test of dirigent function using a crude cell lysate containing soluble FiDRP1. Previous studies of dirigent proteins had shown that, when assaying dirigent function, it is important to achieve a slow rate of monolignol oxidation, such that the random coupling of two radicals does not outpace the binding of the radicals to the dirigent protein.[100-101] Therefore, we adapted an assay for dirigent function used previously by other groups, using flavin mononucleotide (FMN) as a non-specific oxidant of coniferyl alcohol. *E. coli* BL21 (DE3) cell lysate containing soluble FiDRP1 (Figure 2.8) was incubated with coniferyl alcohol and FMN for 4 h at 30 °C, after which the

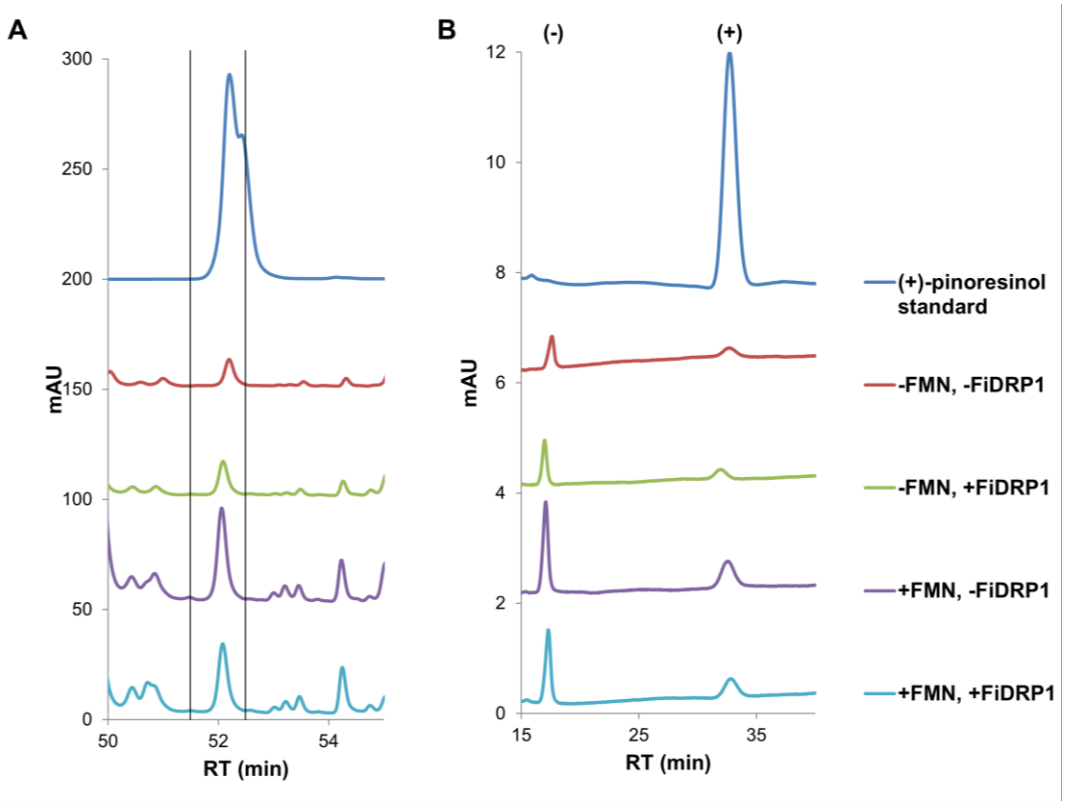


Figure 2.9. Preliminary test of FiDRP1 dirigent function using crude cell lysates. A crude cell lysate containing high levels of soluble FiDRP1 (Figure 8) was incubated with 4 mM coniferyl alcohol and 1 mM FMN for 4 h in HEPES-MES-acetate buffer (pH 5.0) with 75 mM Na_2SO_4 at 30 °C. Then the reaction was immediately extracted with ethyl acetate and prepared for reverse phase and chiral HPLC. A) Reverse-phase separation of pinoresinol from other media components. Vertical lines indicate the fraction collected for subsequent concentration and chiral separation. B) Chiral separation of pinoresinol enantiomers.

reaction was immediately extracted with ethyl acetate and prepared for HPLC. First, samples were run on a C18 column for the separation of pinoresinol from other media components (Figure 2.9A); the pinoresinol collection from this run was then collected, concentrated, and subjected to chiral separation for the detection of (+)- and (-)-pinoresinol enantiomers (Figure 2.9B). In the presence of the cell lysate containing

FiDRP1, we were unable to detect any relative increase in formation of the (+)-pinoresinol enantiomer, or any decrease in the formation of the (-)-pinoresinol enantiomer, suggesting that FiDRP1 is non-functional in *E. coli.* Oddly, we detected the formation of some racemic pinoresinol in the samples even when FMN was not present (Figure 2.9); we suspect this may be due to some other nonspecific oxidant present in the cell lysate, which made up a large portion of the total reaction volume in the assay.

2.3 Cloning and attempted expression of putative dirigent protein from *Arabidopsis thaliana* in *E. coli*

While putative dirigent-protein-encoding genes are widespread in vascular plants, at the time this work was begun, only one dirigent protein, FiDRP1, had been cloned and characterized. We therefore attempted to clone and express three putative dirigent proteins from *Arabidopsis thaliana* – AtDRP2, AtDRP3, and AtDRP5 – for future biochemical characterization. Both truncated (omitting the N-terminal signal peptide) open reading frames of the three genes were amplified from *A. thaliana* cDNA and cloned into pUCmodRBS with C-terminal His-tags. Colonies of some of the resulting plasmids showed an unusual, translucent phenotype on agar plates, suggesting to us that the genes may be toxic to *E. coli* when expressed constitutively. Therefore, the open reading frames (including His-tags) were re-cloned into pBADmod1, which contains an arabinose-inducible promoter. When *E. coli* BW27784 was transformed with these constructs, cultured to mid log-phase, and then supplemented with arabinose for expression induction, only one putative dirigent protein, AtDRP5, showed expression *via*

anti-His-tag western blotting (Figure 2.10). Interestingly, as overall protein levels in the cell appeared to increase over time, a higher percentage of that protein appeared to aggregate and remain in the insoluble fraction during sample processing. In fact, while overall expression levels were lowest 2 h after induction, all protein was soluble; however, at 24 h, very little soluble protein can be seen (Figure 2.10). This clearly illustrated in AtDRP5 the same tendency to form insoluble aggregates that we and others had observed with FiDRP1,[116] and would later be observed by other groups characterizing AtDRPs.[117]

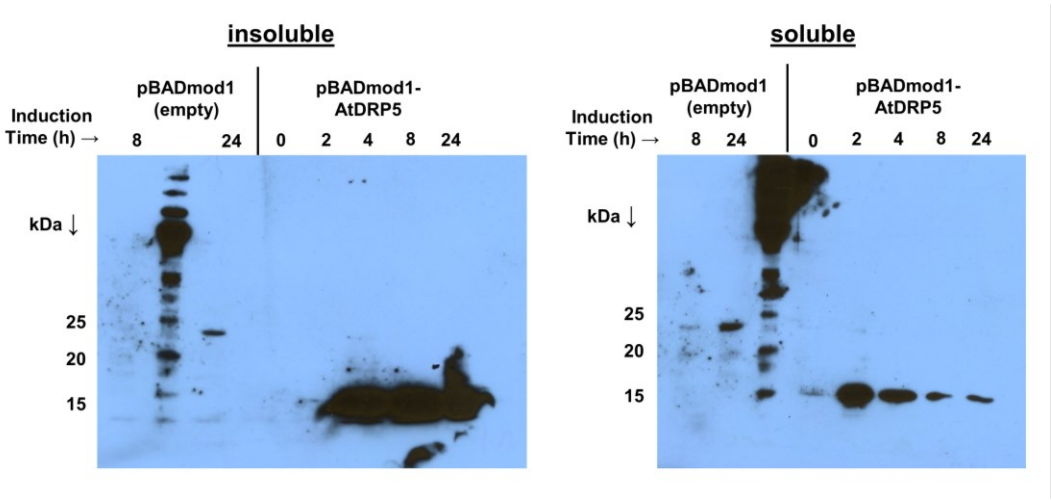


Figure 2.10. Expression and aggregation of AtDRP5 in *E. coli* BW27784. Cells were transformed with pBADmod1 (empty) or pBADmod1-AtDRP5-ctH6, cultured overnight at 37 °C, and then moved to an 18 °C shaker after induction with arabinose (0.2 % w/v). Cells were periodically harvested, lysed *via* sonication, and subjected to anti-His-tag western blotting.

3. Discussion

3.1 SLAC as a single electron oxidase for lignan biosynthesis in *E. coli*

We chose SLAC as a single electron oxidase for ease of expression in a bacterial host, in addition to its reported high activity. We do in fact observe high levels of SLAC expression in *E. coli* with the same high stability reported in the literature (Figure 2.1).[108] We were also able to show that purified SLAC can oxidize coniferyl alcohol *in vitro* (Figure 2.2). With supplementation of copper to the media, *E. coli* cells expressing SLAC were also able to oxidize coniferyl alcohol to pinoresinol at higher levels than cells transformed with empty plasmid (Figure 2.3). In both *in vitro* and *in vivo* assays, the coniferyl alcohol was not stoichiometrically converted into pinoresinol. This is to be expected, as single-electron oxidation of coniferyl alcohol in the absence of a dirigent protein has been shown to result in a mixture of racemic pinoresinol, dehydro-diconiferyl alcohol, and erythro/threo-guaiacylglycerol coniferyl alcohol ethers.[27, 102] However, levels of pinoresinol remained quite low and eventually decreased over time in both assays, even as coniferyl alcohol continued to be oxidized (Figures 2.2 and 2.3). Based on the widespread reported promiscuity of laccases,[106-107] this is likely due to the further oxidation and crosslinking of pinoresinol and other dimeric products, forming larger polymers which are not visible *via* our analytical methods. SLAC's promiscuous activity could be problematic for use in heterologous phenylpropanoid biosynthetic pathways, as any downstream product retaining a phenolic ring may be subject to further oxidation and crosslinking. Ideally, an oxidase with a high specificity for coniferyl would

circumvent this problem; unfortunately, to our knowledge, no such enzyme has been reported in the literature. However, we note that in the *in vitro* assay, a decrease in pinoresinol levels does not occur until late in the assay, when all of the coniferyl alcohol has been depleted (Figure 2.2). This suggests that the presence of high concentrations of a competing substrate may competitively slow the oxidation of lignan products. Likewise, decreasing the level of SLAC enzyme present by slowing transcription/translation rates could prevent the oxidation of downstream products. However, both of these approaches would also cause a decrease in coniferyl alcohol oxidation. Because of the lack of specific single-electron oxidases for monolignols, a laccase such as SLAC is the best option for heterologous biosynthesis of lignans; however, this approach has an inherent flaw in the promiscuous nature of laccases.

3.2 Dirigent proteins are inherently problematic in bacterial hosts.

We encountered extreme difficulty in attempting the soluble expression of dirigent proteins in *E. coli*. FiDRP1 was only solubly expressed at useful levels when *E. coli* protein folding chaperones were also overexpressed (Figure 2.7 and 2.8), suggesting that protein folding is a significant hindrance to the expression of dirigent proteins in bacteria. Both FiDRP1 (Figures 2.4-2.6) and AtDRP5 (Figure 2.10) exhibited a tendency to aggregate in *E. coli*, suggesting that this may be a common problem among dirigent proteins, rather than a problem specific to one of them. We are currently unable to explain the inability of soluble N-terminally His-tagged FiDRP1 to bind to a metal affinity resin; this could be due to burial of the His-tag in the quaternary structure of the protein,

or potentially aggregation of FiDRP1 during the purification process. In the absence of purified protein, an assay using a crude cell lysate with high levels of FiDRP1 gave no evidence of dirigent function (Figure 2.9).

During the completion of this work, two studies were published which validated our difficulties with functional dirigent protein expression in *E. coli*. First, Pickel et al. published a structural study and a molecular model of a dirigent protein from *A. thaliana*, AtDIR6, using a crystal structure of an allene oxide cyclase as a template. This model established AtDIR6 and FiDRP1 as homodimeric, all-barrel proteins, showed the presence of a disulfide bridge linking the N- and C-termini of AtDIR6, and identified two potential N-glycosylation sites.[118] As this disulfide bond may be necessary for proper protein folding and stability, the inability of *E. coli* to form disulfide bonds was likely a significant contributor to the difficulty in expressing soluble dirigent proteins. Shortly thereafter, the same group published an optimized expression system for AtDIR6 in *Pichia pastoris* and showed that deglycosylation of the purified protein led to loss of dirigent function and aggregation.[117] Again, the inability of *E. coli* to perform N-glycosylation similar to the native organism most likely was a major factor in the aggregation of dirigent proteins in our system. Moreover, the necessity of the N-glycosylation for dirigent function makes it impossible for a functional dirigent protein to be expressed in *E. coli*.

3.3 The potential of a microbial production platform for lignans.

The original goal of this work was to establish the first step towards a platform for lignan biosynthesis in *E. coli*. Very recently, Kuo et al. published the bioconversion of pinoresinol to matairesinol, a precursor to podophyllotoxin in *E. coli*.[59] In this study, a strain expressing a fusion of two plant enzymes, pinoresinol-lariciresinol reductase and secoisolariciresinol reductase (PLR-SDH), was fed exogenous pinoresinol and catalyzed its conversion to matairesinol.[59] An ideal biosynthetic platform for lignans could network *in vivo* stereospecific oxidative radical coupling with the use of downstream biosynthetic enzymes. However, taken together, our findings of the difficulty of dirigent protein expression in *E. coli* combined with the recent studies on the necessity of post-translational modifications for dirigent function demonstrate the infeasibility of achieving stereoselective oxidative radical coupling in a bacterial host. The establishment of an overexpression system for functional dirigent proteins in *P. pastoris*[117] points to the use of eukaryotic microbes as a more suitable heterologous host for lignan biosynthesis. Granted that a suitable single electron oxidase could be identified, the *P. pastoris* dirigent protein expression strain could also be equipped with upstream biosynthetic enzymes as well as the PLR-SDH fusion created by Kuo et al.[59] for the efficient bioconversion of cheap phenolic feedstocks, such as ferulic acid, to valuable lignans, such as matairesinol.[117]

4. Conclusions

In order to create a microbial biosynthetic platform for valuable lignan compounds, we established the expression in *E. coli* of two proteins necessary for stereo- and regio-selective oxidative radical coupling of coniferyl alcohol to (+)-pinoresinol. The single electron oxidase SLAC was successfully expressed, cloned, purified and shown to oxidize coniferyl alcohol to pinoresinol both *in vitro* and *in vivo*. The characterized (+)-pinoresinol-forming dirigent protein FiDRP1 was solubly expressed, after extensive expression optimization, but was not found to be functional in *E. coli*. Finally, AtDRP5 was expressed at low levels and shown to have the same propensity as FiDRP1 to aggregate. Studies published during the completion of this work have illustrated the infeasibility of using *E. coli* as a heterologous host for lignan biosynthesis, due to the requirement of post-translational modifications for dirigent protein function. By contrast, the use of a eukaryotic microbe, such as *P. pastoris*, shows much greater promise for the achievement of our original goal of microbial conversion of cheap feedstocks to value-added lignans.

5. Materials and Methods

5.1 Chemicals and Enzymes

Chemicals for this work were procured as follows: (+)-pinoresinol (Arbo Nova, Turku, Finland); coniferyl alcohol, HEPES, MES (Sigma Aldrich, St. Louis, MO, USA); L-(+)-arabinose (ICN Biomedical, Inc., Aurora, OH, USA); IPTG (G. Biosciences, St. Louis,

MO, USA); Tris (Fisher Scientific, Waltham, MA, USA); cupric sulfate pentahydrate (Baker Analyzed, Phillipsburg, NJ, USA); potassium acetate, sodium sulfate (Mallinckrodt, St. Louis, MO, USA). The following enzymes and kits were used for DNA manipulation and protein visualization according to the manufacturers' protocols: Wizard® Genomic DNA Purification Kit (Promega, Madison, WI, USA); QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany); Taq and Pfu DNA polymerases (Life Technologies, Carlsbad, CA, USA); restriction enzymes, T4 DNA ligase (New England Biolabs, Ipswich, MA, USA); Monoclonal Mouse IgG1 Anti-poly-His Antibody (R&D Systems, Minneapolis, MN, USA); SuperSignal West Pico Mouse IgG Detection Kit (Thermo Scientific, Rockford, IL, USA). Oligomers used for the assembly of the codon-optimized gene encoding FiDRP1 were obtained from Integrated DNA Technologies (Coralville, IA, USA).

5.2 Strains, Growth Conditions and Expression Studies

E. coli DH5α and *E. coli* JM109 were used for all plasmid manipulation and cloning and were cultured at 37 °C in Luria-Burtani (LB) broth with ampicillin (100 μg/mL) as needed. All protein expression experiments were carried out using *E. coli* BW27784 (*E. coli* Genetic Stock Center, New Haven, CT), which overexpressed the *araE* gene encoding an arabinose transporter,[119] or *E. coli* BL21 (DE3), which expresses a T7 RNA polymerase. A list of strains used in this study is listed in Table 2.1. For the *in vivo* SLAC activity assay, *E. coli* BW27784 cells transformed with pUCmodRBS-SLAC were cultured in modified M9 supplemented with 1.25 g/L yeast extract and 5 g/L glycerol

instead of glucose. For protein expression analysis, 100 mL cultures of cells containing overexpression plasmids were grown in LB media at various temperatures as noted in the main text and figure captions. At various timepoints, 5-mL samples of the cultures were harvested and cells pelleted for 5 min at 4000 RPM and 4 °C. The media was discarded and cell pellets were resuspended in 1 mL 100 mM Tris pH 8.0 with 10 mM NaCL and lysed by 1 min sonication (5 s on, 10 s off). The lysate was centifuged for 10 min at 13,200 RPM at and 4 °C; the resulting supernatant was removed and saved, and the pellet was resuspended in 1 mL 100 mM Tris pH 8.0 with 10 mM NaCL. These samples were subjected to SDS-PAGE and anti-His-tag western blotting.

Table 2.1. Strains used in this study.

Strain	**Description**	**Source**
Escherichia coli DH5α	*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--*Tet^S*)2 [dcm] R(zgb-210::Tn10--*Tet^S*) endA1 Δ(mcrC-mrr)114::IS10*	New England Biolabs, Ipswich, MA, USA
Escherichia coli JM109	*endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB$^+$ Δ(lac-proAB) e14- [F' traD36 proAB$^+$ lacI^q lacZΔM15] hsdR17(r$_K^-$m$_K^+$)*	New England Biolabs, Ipswich, MA, USA
Escherichia coli BW27794	*lacI*q *rrnB3* D*lacZ4787 hsdR514* DE(*araBAD*)*567* DE(*rhaBAD*)*568* DE(*araFGH*) U(D*araEp* PCP")±*araE*)	[119]
Escherichia coli BL21 (DE3)	F$^-$ ompT gal dcm lon hsdS$_B$(r$_B^-$ m$_B^-$) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])	New England Biolabs, Ipswich, MA, USA

5.3 Gene amplification and cloning

The gene encoding SLAC (GenBank: CAB45586.1) was amplified from *Streptomyces coelicolor* genomic DNA with a forward primer containing an NdeI site at the 5' end and a reverse primer containing an in-frame XhoI site and no stop codon at the 5' end. This was digested and ligated into pUCmodRBS cut with the same restriction sites, such that the C-terminal His-tag contained on the plasmid was in frame with the SLAC open reading frame. A truncated version of the gene encoding FiDRP1 (GenBank: AAF25357.1) without the N-terminal 24-residue plant cell wall targeting sequence (MVSKTQIVALFLCFLTSTSSATYG) was assembled through overlap extension PCR of eight 82-88 nucleotide synthetic DNA oligomers. The oligomers were mixed in equimolar amounts and subjected to PCR using a forward primer containing and NdeI site and a reverse primer containing an XhoI site either with or without a stop codon. The amplified DNA was digested and ligated into pUCmodRBS cut with the same restriction sites. In the case of the amplicon without the stop codon, the C-terminal His6 tag contained on the plasmid was in frame with the FiDRP1 open reading frame; in the case of the amplicon with the stop codon, the resulting protein remained non-His-tagged. Subsequently, pUCmodRBS-FiDRP1 (no His-tag) was digested with NdeI and XhoI, and the resulting fragment was ligated into pET15b digested with the same enzymes such that the N-terminal His-tag and thrombin cleavage site contained on the plasmid was in frame with the FiDRP1 open reading frame. The putative dirigent proteins from *A. thaliana*, AtDRP2 (GenBank: AEE33205.1; TAIR: AT1G5521), AtDRP3 (GenBank:

AEC07124.1; TAIR: AT2G21110) and AtDRP5 (GenBank: AEE34203.1; TAIR: AT1G64160), were amplified from cDNA clones obtained from the Arabidopsis Biological Resource Center. Amplification excluded the N-terminal plant cell wall targeting sequence for each protein (AtDRP2, MAKLIFFLAVQI; AtDRP3, MGKNLGLVVSFY; AtDRP5, MVGQMKSFLFLFVFLV) and was performed with a forward primer containing an NdeI site and a reverse primer containing an XhoI site. The amplified DNA was digested and ligated into pUCmodRBS cut with the same restriction sites such that the C-terminal His-tag contained on the plasmid was in frame with the gene. Subsequently, using the pUCmodRBS clones as templates, these genes were re-amplified using forward primers containing AscI sites and reverse primers containing PacI sites, such that the C-terminal His-tag was amplified with the gene. The amplified DNA was then ligated into pBADmod1 digested with the same enzymes. A list of plasmids used in this study is given in Table 2.2.

Table 2.2. Plasmids used in this study.

Plasmid	Description	Source
Standard Plasmid Backbones		
pUCmodRBS	pMB1 (colE1); Amp^r $P_{lacP'}$	[120]
pET15b	pBR322; $Amp^r P_{T7}$; N-terminal 6xHis-tag and thrombin cleavage site (*ntH6*)	Novagen (EMD Millipore)
pBADmod1	pMB1 (colE1); Amp^r P_{araBAD}	[11]
Expression Plasmids		
pUCmodRBS-SLAC	pUCmodRBS:: $P_{lacP'}$*SLAC*ctH6	This study
pUCmodRBS-FiDRP1-ctH6	pUCmodRBS:: $P_{lacP'}$*fiDRP1*ctH6	This study
pUCmodRBS-FiDRP1	pUCmodRBS:: $P_{lacP'}$*fiDRP1*	This study
pET15b-FiDRP1	pET15b:: $P_{lacP'}$*fiDRP1*	This study
pUCmodRBS-AtDRP2-ctH6	pUCmodRBS:: $P_{lacP'}$*atDRP2*ctH6	This study
pUCmodRBS-AtDRP3-ctH6	pUCmodRBS:: $P_{lacP'}$*atDRP3*ctH6	This study
pUCmodRBS-AtDRP5-ctH6	pUCmodRBS:: $P_{lacP'}$*atDRP5*ctH6	This study
pBADmod1-AtDRP2-ctH6	pBADmod1:: P_{araBAD}*atDRP2*ctH6	This study
pBADmod1-AtDRP3-ctH6	pBADmod1:: P_{araBAD}*atDRP3*ctH6	This study
pBADmod1-AtDRP5-ctH6	pBADmod1:: P_{araBAD}*atDRP5*ctH6	This study

5.4 HPLC analysis of oxidative radical coupling assays

Coniferyl alcohol oxidation assays were performed using the conditions described in figure legends (Figure 2.2, Figure 2.3, Figure 2.9). Samples from assays were extracted twice in 2x volume ethyl acetate, and the organic fraction was dried to completion and

resuspended in methanol for a 10x concentration of the original sample. For detection of coniferyl alcohol and racemic pinoresinol, samples were separated *via* reverse-phase chromatography on a Zorax SB-C18 column (5 uM, 4.6 X 250 mm) with a 1 mL/min flow rate using the following timetable: 5% acetonitrile isocratic for 5 min; gradient to 10% acetonitrile for 15 min; gradient to 20% acetonitrile for 25 min; gradient to 50% acetonitrile for 15 min; post-run at 5% acetonitrile for 10 min. Buffers (acetonitrile and water) contained 3% acetic acid. Coniferyl alcohol was detected at 264 nm; pinoresinol was detected at 280 nm. Analytical standards were used to confirm peak identity. For subsequent chiral separation of racemic pinoresinol, the fractions were collected from retention time 51.5 min to 52.5 min, corresponding to the retention time of pinoresinol. The collected fractions were lyophilized, resuspended in methanol, and separated on a Chiralcel OD-H chiral column (250 x 4.6 mm) with a flow rate of 0.5 mL/min and an isocratic elution with 1:1 hexanes:ethanol. Pinoresinol was detected at 280 nm, and a (+)-pinoresinol standard was used to confirm the identity of the (+) enantiomer peak.

Chapter 3. Design and Creation of a Novel Biosynthetic Pathway for Hydroxycinnamic Acid Esters in *Escherichia coli*

Summary

Hydroxycinnamic acid esters (HCEs) are widely-distributed phenylpropanoid-derived plant natural products. Rosmarinic acid (RA **7**), the most well-known HCE, shows

promise as a treatment for cancer and neurological disorders. In contrast to extraction from plant material or plant cell culture, a microbial production platform for HCEs could provide a sustainable, controlled means of production. Through the overexpression of a six-enzyme chimeric bacterial and plant pathway, we show the *de novo* biosynthesis of RA **7** and the related HCE isorinic acid (IA **8**) in *E. coli*. Probing the pathway through precursor supplementation showed several potential pathway bottlenecks. We show HCE biosynthesis using three plant RAS orthologs exhibiting different levels of HCE biosynthesis, but the same ratio of IA **8** to RA **7** produced. This work serves as a proof of concept for a microbial production platform for HCEs using a modular biosynthetic approach to access diverse natural and non-natural HCEs.

*This work was accepted to ChemBioChem and published online on 9 September 2014.[121] The manuscript has been adapted here with minor formatting revisions.

1. Introduction

Hydroxycinnamic acid esters (HCEs) are a diverse class of phenolic plant natural products consisting of a hydroxycinnamoyl moiety, typically caffeic acid **6**, conjugated with another aromatic molecule, such as phenyllactate, shikimate, or quinate.[23, 46] HCEs have received attention recently as potential nutraceutical supplements due to their anti-

oxidant and anti-inflammatory properties. The most well-known HCE, rosmarinic acid (RA **7**), is the ester of caffeic acid **6** and 3,4-dihydroxyphenyllactic acid (3,4-DHPL **3**), and shows promise for a broad range of nutraceutical and pharmaceutical applications (Scheme 3.1).[23, 46] Of particular interest is its demonstrated anti-carcinogenic and anti-tumorogenic activities and utility in cancer prevention and treatment in cell culture [78] and animal models.[122-125] Additionally, RA **7** exhibits neuroprotective [94, 126-128] and cognition-enhancing effects [129], and has shown promise for use in the prevention and treatment of diverse neurological disorders including Alzheimer's,[130] Parkinson's disease,[131] depression [132] and post-traumatic stress disorder.[133]

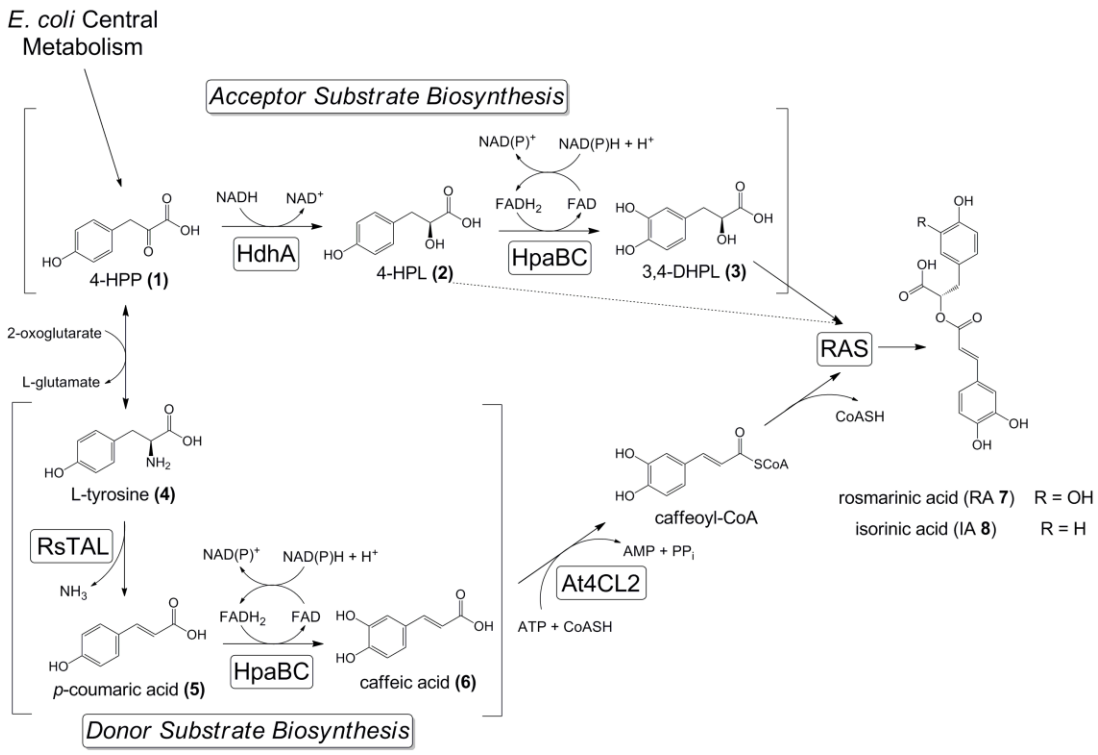
Currently, RA **7** and other HCEs are produced primarily by extraction from plant tissue, which is a costly endeavor limited by the availability of plant material and the low concentration of HCEs in plant tissues. There are scant reports on the total [134-135] and chemoenzymatic [136] synthesis of RA **7** and some derivatives;[137-138] however, these methods are inefficient for economical commercial production, requiring the use of harsh chemicals and giving low yields due to a chiral center in RA **7**. Cell culturing methods could provide a lower-cost means of mass-production, and many groups have reported high production titers for RA **7** from plant cell cultures.[139-144] However, none of these cell culturing methods have yet been applied on an industrial scale, perhaps due to the lack of understanding of regulatory networks that cause a decline in RA **7** biosynthesis in long-term plant cell culture conditions.[23] By contrast, the development of a microbial production platform for RA **7** and other HCEs could allow for a more controlled and

stable means of producing these compounds. Additionally, the use of a microbial host could allow for a modular production system and, through combinatorial biochemistry, could lead to the biosynthesis of novel HCEs with valuable biological activities.

In plants, HCE biosynthesis relies on the convergence of two pathways – the synthesis of the hydroxycinnamoyl-CoA ester donor substrate and the conjugate acceptor substrate, which are subsequently ester-linked by a hydroxycinnamoyl transferase (Scheme 3.1). In the RA **7** biosynthetic pathway, which has been completely elucidated in plants, the hydroxycinnamoyl transferase rosmarinic acid synthase (RAS) conjugates 4-hydroxyphenyllactic acid (4-HPL **2**) and *p*-coumaroyl-CoA, and the resulting ester is then hydroxylated at the 3 and 3' positions by cytochrome P450s to give RA **7**.[46, 145] The native plant biosynthetic pathway for RA **7** is impractical for use in a microbial host, requiring the co-overexpression of several plant enzymes including microsomal cytochrome P450s. Because they require localization to the endoplasmic reticulum for optimal activity and an associated NADPH-cytochrome P450 reductase for electron transfer, plant microsomal P450 expression is notoriously problematic in microorganisms commonly used in biotechnology, such as *E. coli* and *S. cerevisiae*.[34] Fortunately, however, many hydroxycinnamoyl transferases, including the three RASs previously characterized, show broad substrate flexibility and readily accept 3,4-hydroxylated acids and CoA esters as substrates.[146-155] This could allow for the design of a biosynthetic pathway in which the 3-hydroxylation of the donor and acceptor

substrate precursors could be carried out by a non-cytochrome P450 bacterial monooxygenase prior to ester formation.

Recent studies have reported the bioconversion of phenolic precursors to hydroxycinnamoyl amides,[52-53] shikimates and quinates [51] and the biosynthesis of hydroxycinnamoyl anthranilates from glucose [54] in *E. coli*. In this study, we designed and constructed a novel, modular biosynthetic pathway for the biosynthesis of RA **7** from endogenous L-tyrosine and 4-hydroxyphenylpyruvate (4-HPP **1**) (Scheme 3.1). Through the overexpression of this engineered pathway, we observe *de novo* biosynthesis of RA **7** and isorinic acid (IA **8**), another hydroxycinnamoyl-hydroxyphenyllactate conjugate, in *E. coli* using three orthologous RAS enzymes. This work serves as a proof of concept for a microbial production platform for RA **7** and other HCEs. Advances in metabolic engineering of the shikimate pathway in *E. coli* have shown the utility of this approach for the mass production of aromatic compounds from simple feedstocks.[16, 49, 61, 63, 67, 156] This pathway, used in a tyrosine-overproducing strain and with future optimization of pathway enzyme activity and expression, could lead to an inexpensive production platform for HCEs.



Scheme 3.1. Redesigned chimeric RA biosynthetic pathway. The D-hydroxyisocaproate dehydrogenase HdhA from *L. delbrueckii* subsp. *bulgaricus* catalyzes the conversion of endogenous 4-HPP **1** to 4-HPL **2**, which is subsequently meta-hydroxylated by hydroxylase complex HpaBC cloned from *E. coli* to yield 3,4-DHPL **3**, the intended acceptor substrate for rosmarinic acid synthase (RAS). Donor substrate biosynthesis is achieved through the conversion of L-tyrosine **4** to *p*-coumaric acid by a tyrosine ammonia lyase from *R. sphaeroides* (RsTAL) and hydroxylation of *p*-coumaric acid **5** to caffeic acid **6** by HpaBC. A 4-coumaroyl-CoA ligase from *Arabidopsis thaliana* (At4CL2) generates the CoA-ester of the donor substrate, allowing ester formation catalyzed by RAS. Both 3,4-DHPL **3** and 4-HPL **2** may be accepted by RAS to form both RA **7** and IA **8**, respectively.

2.1 Design of a chimeric RA biosynthetic pathway for *E. coli*

In order to design an effective system for microbial production of RA **7** and minimize the number of plant enzymes required, we designed a novel chimeric bacterial-plant

pathway in which the meta-hydroxylation of the donor and acceptor substrates takes place before condensation by RAS (Scheme 3.1). In this pathway, caffeic acid **6** and 3,4-DHPL **3** serve as the donor and acceptor substrates, respectively, rather than *p*-coumaric acid **5** and 4-HPL **2** as in the plant pathway. All pathway genes were initially cloned into pUCBB, one of our group's custom set of BioBrick™-compatible vectors, for subsequent assembly of multi-gene plasmids to create the biosynthetic pathway modules described later.[157] Monocistronic "bricks" containing gene expression cassettes allow for "stacking" using the standard BioBrick™ restriction sites [158] that can be reused repeatedly.

The biosynthesis of caffeic acid **6** has been demonstrated previously in *E. coli*;[47-50, 54] however, at the time of this work, biosynthesis of 3,4-DHPL **3** had not been shown in *E. coli*. To achieve 3,4-DHPL **3** biosynthesis, we first chose a 2-hydroxyacid dehydrogenase from *Lactobacillus delbrueckii* subsp.*bulgarcius*, HdhA, that was originally characterized as a D-isomer specific 2-hydroxyisocaproate dehydrogenase and was shown to have some activity against phenylpyruvic acid.[159] We hypothesized that the overexpression of the gene *hdhA* would lead to the reduction of endogenous 4-HPP **1** to 4-HPL **2** as a first step in 3,4-DHPL **3** biosynthesis (Scheme 3.1). Therefore, we amplified the gene encoding HdhA from *L. delbrueckii* genomic DNA and created the plasmid pUCBB-hdhA, which contains a constitutive expression cassette for *hdhA* (Figure 3.1). *E. coli* BW27784 transformed with pUCBB-hdhA were cultured in minimal media and samples were taken over time, cleared of cells, and extracted with ethyl

acetate for HPLC analysis. We observed production of 4-HPL **2** to 39 ± 6 mg/L without the addition of any precursors to the culture media (Figure 3.2), suggesting that that HdhA reduces endogenous 4-HPP **1** to 4-HPL **2**. No 4-HPL **2** biosynthesis was observed in transformants of the empty pUCBB plasmid.

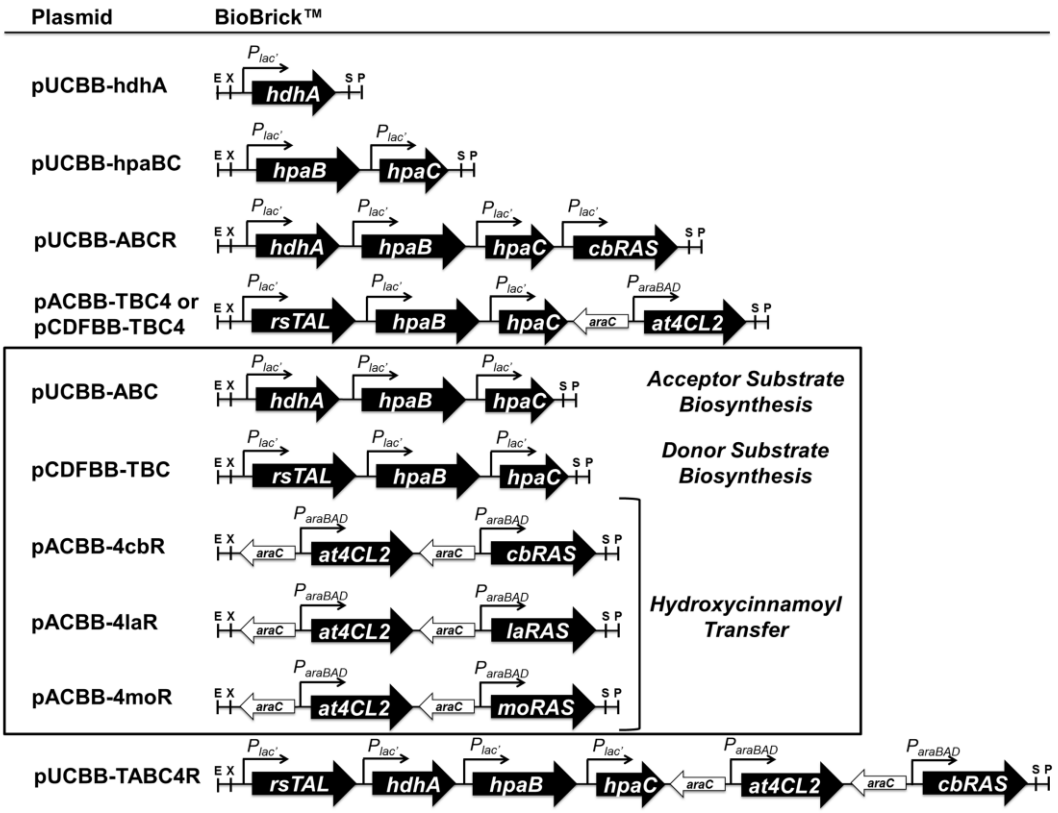


Figure 3.1. Plasmids assembled for donor substrate, acceptor substrate and HCE biosynthesis. All plasmids were assembled using the BioBrick™ system previously designed and built by our lab.[155] $P_{lac'}$, a modified constitutive *lac* promoter; P_{araBAD}, arabinose inducible promoter; *araC*, gene encoding the repressor of P_{araBAD}; pUCBB, high copy number plasmid backbone; pCDFBB, medium copy number plasmid backbone; pACBB, low copy number plasmid backbone. Plasmids that constitute the functional three-module RA **7** biosynthetic pathway are boxed.

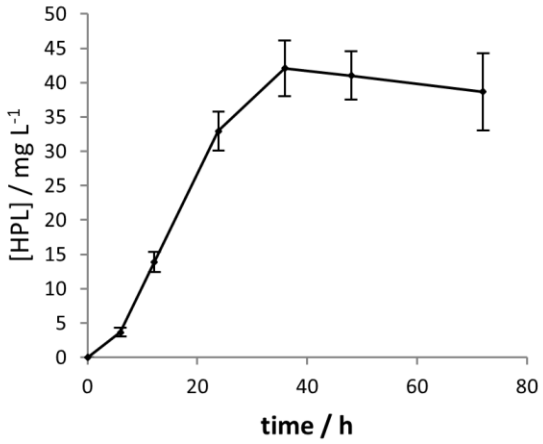


Figure 3.2. De novo biosynthesis of 4-HPL by *E. coli* BW27784 cells transformed with pUCBB-hdhA. Transformants were cultured in minimal media at 30 °C; culture samples were taken over time, cleared of cells, and extracted with ethyl acetate for HPLC analysis. An analytical standard of 4-HPL **2** was used to confirm peak identity and quantification. No 4-HPL **2** was detected in cultures of cells transformed with the empty pUCBB vector.

The next challenge in 3,4-DHPL **3** biosynthesis was the meta-hydroxylation of 4-HPL **2** to 3,4-DHPL **3** (Scheme 3.1). Two flavin-dependent monooxygenases, Sam5 from *Saccharothrix espanaensis* and the *E. coli* 4-hydroxyphenylacetate 3-hydroxylase complex, HpaBC, have been shown to carry out the meta-hydroxylation of *p*-coumaric acid **5** and facilitate biosynthesis of caffeic acid **6** in *E. coli* when co-overexpressed with a tyrosine ammonia lyase (TAL).[47-50] Due to the similarity in structure between *p*-coumaric acid **5** and 4-HPL **2**, we hypothesized that one of these two monooxygenases may be able to accept 4-HPL **2** as a substrate and produce 3,4-DHPL **3**. We amplified

the gene encoding Sam5 from *S. espanaensis* genomic DNA, amplified the genes encoding HpaB and HpaC individually from *E. coli* BL21 genomic DNA, and created the constitutive overexpression plasmids pUCBB-sam5 and pUCBB-hpaBC (Figure 3.1). *E. coli* BW27784 transformed with pUCBB-hpaBC were cultured in minimal media in the presence of 1 mM *p*-coumaric acid **5** or 1 mM 4-HPL **2** (Scheme 3.1); culture samples were taken over time, cleared of cells, extracted and analyzed *via* HPLC. The overexpression of *sam5* led to the complete conversion of *p*-coumaric acid **5** to caffeic acid **6** within 24 h, as had been shown previously; however, no conversion of 4-HPL **2** to 3,4-DHPL **3** could be observed. By contrast, the overexpression of *hpaBC* led to the complete conversion of *p*-coumaric acid **5** to caffeic acid **6** within 24 h and the detection of 3,4-DHPL **3** after 24 h (Figure 3.3). Additionally, because we chose to employ the same enzyme to carry out two simultaneous conversions in the RA biosynthetic pathway, we incubated transformants of pUCBB-hpaBC with both 1 mM 4-HPL **2** and 1 mM *p*-coumaric acid **5** to determine if the presence of one substrate would inhibit the activity of the enzyme against the other. We did not observe any significant effect of feeding both substrates simultaneously on enzyme activity with either substrate (Figure 3.S1). We therefore decided to use HpaBC for the meta-hydroxylation of 4-HPL **2** as well as for caffeic acid **6** biosynthesis.

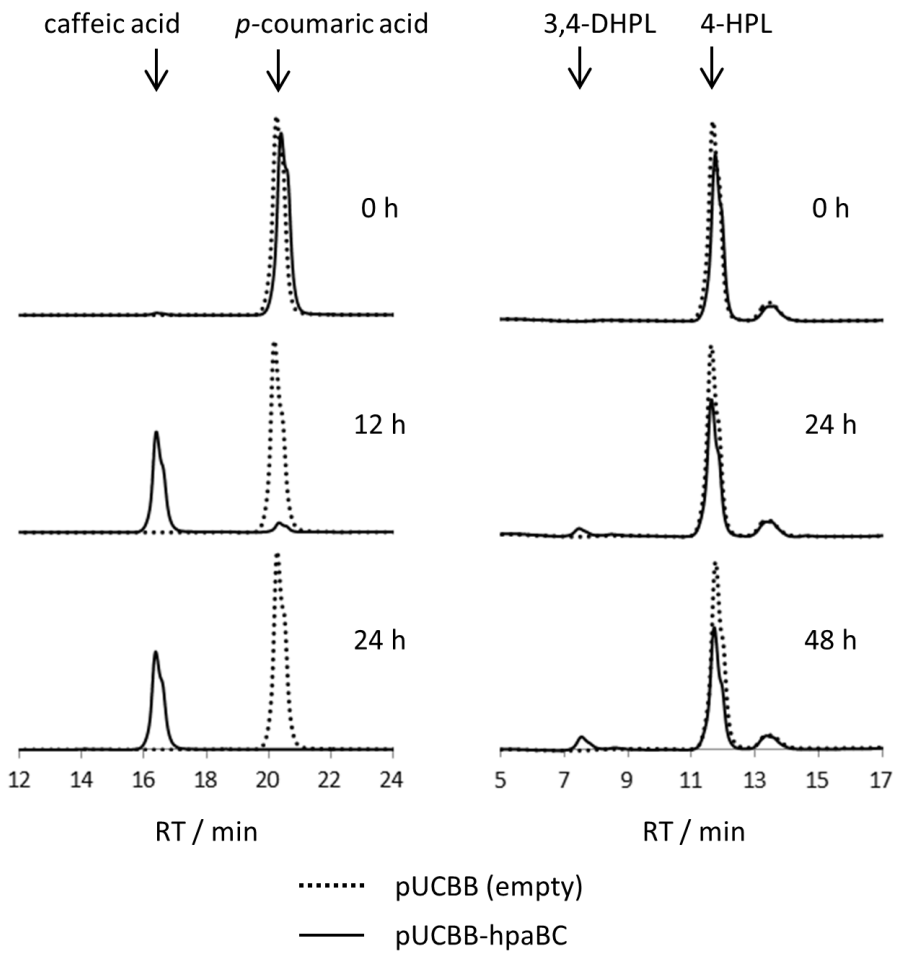


Figure 3.3. Bioconversion of p-coumaric acid and 4-HPL to caffeic acid and 3,4-DHPL, respectively, by *E. coli* BW27784 transformants of pUCBB-hpaBC. Transformants were cultured in minimal media at 30 °C, supplemented with 1 mM *p*-coumaric acid or 1 mM 4-HPL **2** culture and samples were taken over time and processed for HPLC analysis. Samples were compared to standard compounds in order to identify compound peaks. Solid trace, pUCBB-hpaBC transformants; dotted trace, empty pUCBB transformants (control).

After confirming the activities of HdhA and HpaBC with 4-HPP **1** and 4-HPL **2**, respectively, as key RA **7** precursors, we cloned genes encoding enzymes with known activities for the remaining steps of the chimeric pathway as shown in Scheme 3.1. In our chimeric pathway design, acceptor substrate biosynthesis combines the activities of HdhA and HpaBC for the production of 3,4-DHPL **3**. For donor substrate biosynthesis, we chose a highly active TAL from *Rhodobacter sphaeroides* (RsTAL) that we have previously shown to convert L-tyrosine into *p*-coumaric acid **5**.[11] RsTAL and HpaBC together catalyze the conversion of endogenous L-tyrosine **4** to caffeic acid **6**. For CoA activation of the donor substrate caffeic acid **6**, we selected 4-coumarate-CoA ligase 2 from *Arabidopsis thaliana* (At4CL2) due to the reported preference of this isoenzyme for caffeic acid **6** as a substrate [160] and cloned the gene from *A. thaliana* cDNA. Finally, to carry out the condensation of 3,4-DHPL **3** and caffeic acid **6** to form RA **7**, we chose the first characterized RAS from *Coleus blumei* (CbRAS) [153, 161] and cloned the gene from cDNA synthesized from RNA extracted from *C. blumei* plant tissue.

2.3 *De novo* biosynthesis of RA and IA using a modular biosynthetic pathway

One of the biggest advantages of using a heterologous system for the assembly and expression of biosynthetic pathways is the relative ease with which pathway design and precursor input can be manipulated once a working pathway has been established. The biosynthesis of HCEs lends itself well to combinatorial biochemistry, because the hydroxycinnamoyl transferases that carry out condensation of the donor and acceptor substrates have been shown to have relatively broad substrate specificity and catalyze

the condensation of a range of esters and amides [152-154]. The construction of a modular HCE biosynthetic pathway would therefore allow for the interchange of modules for the biosynthesis of different acceptor and donor substrates and the condensation of these to synthesize a range of known or novel HCEs. To this end, we assembled and tested three modular plasmid systems for the production of RA **7** (Figure 3.1) using our previously designed, compatible BioBrick™ plasmids for pathway assembly in *E. coli*.[157]

Because the conversion of 4-HPL **2** to 3,4-DHPL **3** by HpaBC appeared to be relatively slow (Figure 3.3), we assembled the genes for 3,4-DHPL **3** biosynthesis onto the high copy number vector pUCBB. We first attempted a two-plasmid pathway setup that consisted of 3,4-DHPL **3** (acceptor substrate) biosynthetic genes and *cbRAS* being expressed on the plasmid pUCBB-ABCR and the donor substrate biosynthetic genes on the lower-copy plasmids pCDFBB-TBC4 or pACBB-TBC4 (Figure 3.1). Because we have found co-expression of 4CL enzymes with TALs to be problematic, perhaps due to the depletion of CoA pools within the cell, *at4CL2* was expressed under the control of an arabinose inducible promoter (P_{araBAD}) while all other genes were expressed constitutively from a modified $P_{lac'}$ promoter.[157] When *E. coli* BW27784 was co-transformed with the acceptor substrate and *cbRAS* plasmid pUCBB-ABCR and the donor substrate plasmids pCDFBB-TBC4 or pACBB-TBC4 and expression of *at4CL2* was induced 24 h after inoculation, no RA **7** could be detected in the media.

Next, we assembled a three-plasmid modular system, consisting of the acceptor substrate module pUCBB-ABC, containing *hdhA* and *hpaBC*; the donor substrate

module pCDFBB-TBC, containing *rsTAL* and *hpaBC*; and the hydroxycinnamoyl transfer module, pACBB-4cbR, containing *at4CL2* and *cbRAS* (Figure 3.1). Because 3,4-DHPL **3** production by HpaBC was first seen 24 h after the addition of 4-HPL **2** (Figure 3.3), in this three-module system we expressed both *at4CL2* and *cbRAS* under the control of the arabinose inducible promoter P_{araBAD}, while all other genes were expressed constitutively from $P_{lac'}$. Transformants harboring these three plasmids were grown in minimal media, and *at4CL2* and *cbRAS* expression was induced 24 h after inoculation. This led to the detection of RA **7** in the culture media 48 h after induction and up to 0.33 ± 0.05 μM RA **7** 72 h after the induction of *at4CL2* and *cbRAS* expression (Figure 3.4). *E. coli* cells harboring pUCBB-ABC, pCDFBB-TBC and empty pACBB or all three empty plasmid backbones were tested as negative controls; no RA **7** biosynthesis was observed in these cultures. We were concerned that replicating three plasmids may put an unnecessary metabolic burden on the *E. coli* cells and that expressing *hpaBC* at high levels from both pUCBB and pCDFBB may overwhelm the host's protein folding machinery and lead to aggregation of insoluble recombinant proteins in the cell. Therefore, we attempted one non-modular plasmid, pUCBB-TABC4R, with *cbRAS* and *at4CL2* under the control of the arabinose promoter and *rsTAL*, *hdhA* and *hpaBC* expressed constitutively (Figure 3.1). When transformants of this plasmid were grown in minimal media and *at4CL2* and *cbRAS* expression was induced 24 h after inoculation, no production of RA **7** was detected.

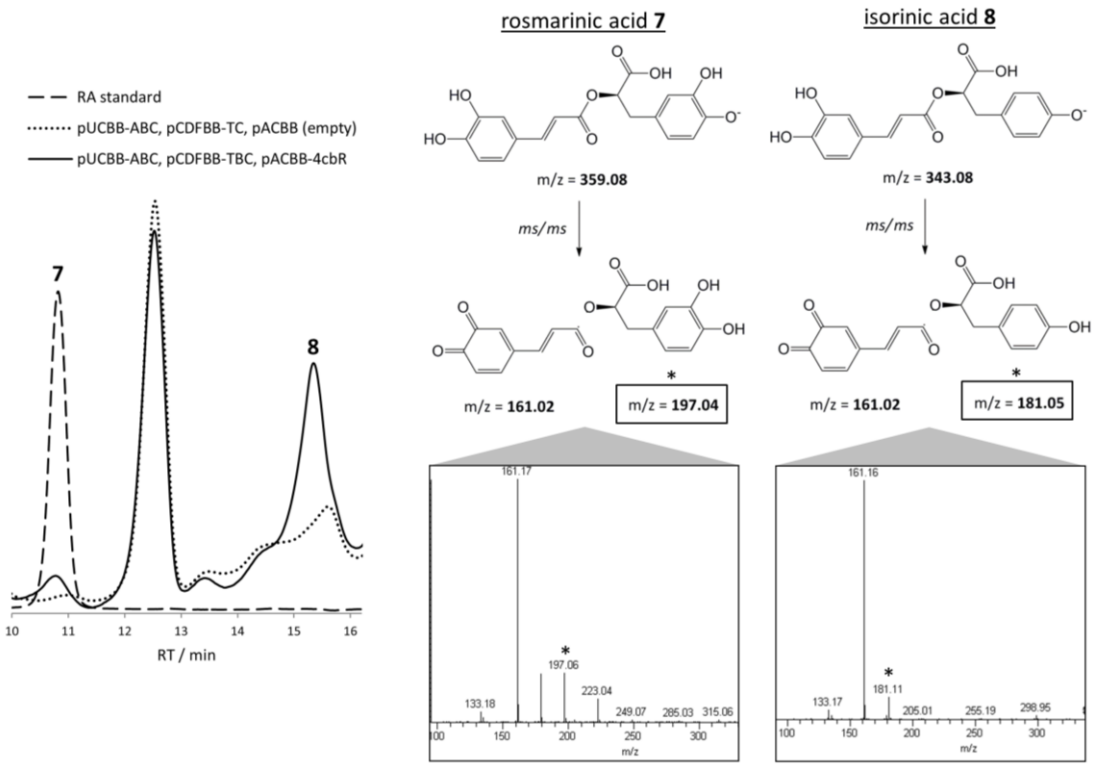


Figure 3.4. De novo biosynthesis of HCEs in *E. coli*. *E. coli* BW27784 was co-transformed with pUCBB-ABC containing the acceptor substrate module, pCDFBB-TBC4 containing the acceptor substrate module, and pACBB-4cbR containing the hydroxycinnamoyl transfer module (Figure 1). Transformants were cultured in minimal media at 30 °C, and *at4CL2* and *cbRAS* expression was induced after 24 h growth. Culture samples were taken over time and processed for HPLC and LC-MS analysis. An analytical standard of RA **7** was used to confirm peak identity and mass fragmentation of RA **7**, and MS/MS mass spectra confirm the presence of both RA **7** and IA **8** in culture media. Asterisks denote key mass fragments in the identification of the acceptor substrate moieties of the products.

Extracts of culture media of cells expressing the three-plasmid pathway also contained an HPLC peak with a different retention time from RA **7** but a nearly identical UV spectrum (Figure 3.S2). Analysis *via* LCMS revealed this peak to be IA **8**, the ester

of 4-HPL **2** and caffeic acid **6**, based on mass shift of the fragment corresponding to 3,4-DHPL **3** from 197 to 181 m/z, and the mass fragment corresponding to caffeic acid **6** (161 m/z) remaining the same (Figure 3.4). The biosynthesis of IA **8** alongside RA **7** can be explained by the previously reported slight preference of CbRAS to bind 4-HPL **2** over 3,4-DHPL **3** [153] and the apparently slow turnover of 4-HPL **2** to 3,4-DHPL **3** by HpaBC (Figure 3.3).

2.4 Feeding cultures with pathway precursors shows potential bottlenecks and a loss of acceptor substrates

Because of the relatively low concentration of HCEs produced and the detection of both RA **7** and IA **8** in the media, we fed the *E. coli* cells expressing the engineered RA **7** pathway modules selected precursors and intermediates and measured the impact on HCE production in order to gain insight into potential pathway bottlenecks. Cultures of *E. coli* BW27784 harboring all three RA **7** pathway modules (pUCBB-ABC, pCDFBB-TBC and pACBB-4cbR) were supplemented with either 1 mM L-tyrosine **4**, 1 mM *p*-coumaric acid **5**, 1 mM caffeic acid **6**, or 1 mM caffeic acid **6** and 1 mM 4-HPL **2** 24 h after inoculation and at the same time of induction of *at4CL2* and *cbRAS* expression, and the levels of HCEs and pathway intermediates were measured *via* HPLC 48 h after induction (Figure 3.5A).

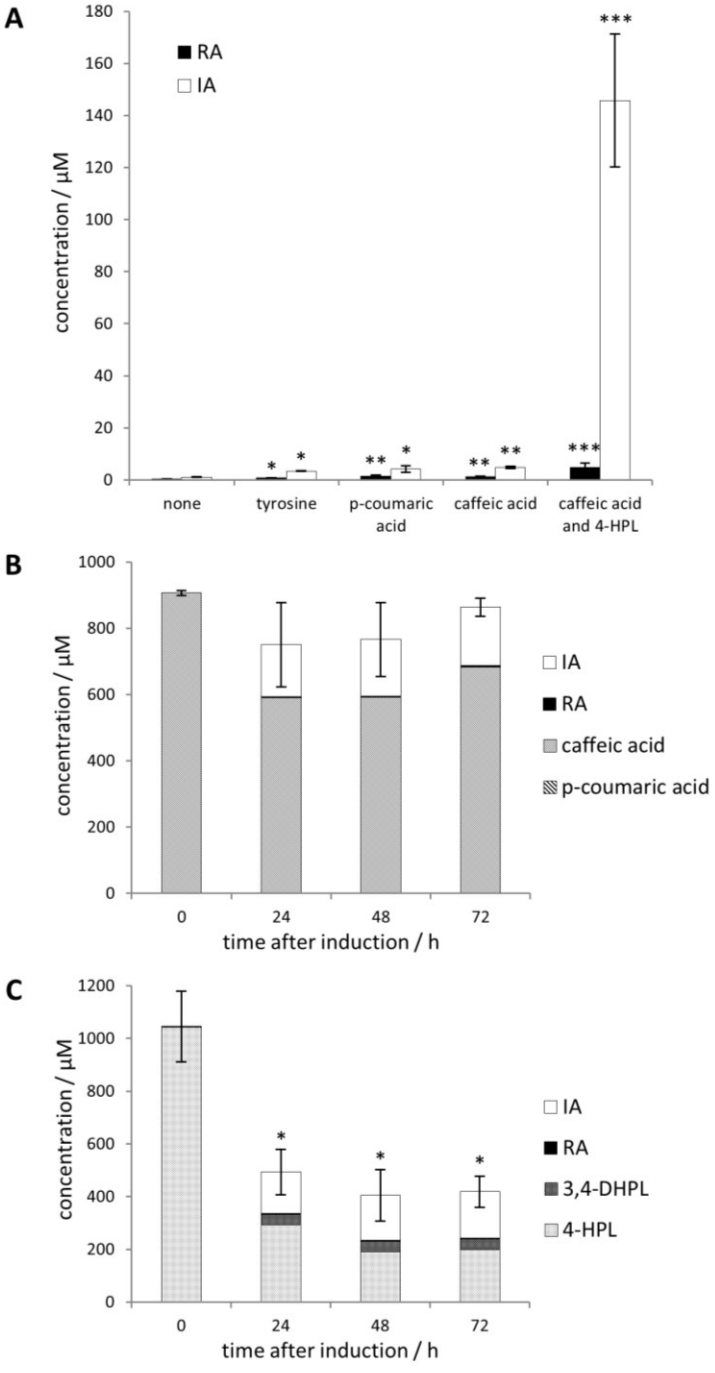


Figure 3.5. Limiting factors and leaks in the *E. coli* HCE biosynthetic pathway. *E. coli* BW27784 containing the modular HCE biosynthesis plasmids were cultured in minimal media at 30 °C; after 24h, *at4CL2* and *cbRAS* expression was induced and various supplements were added to a final concentration of 1 mM. Culture samples were taken over time and analyzed by HPLC. Asterisks denote significantly different concentrations at $p < 0.05$. A) HCE production is increased in the presence of different feeding precursors 48h after induction of *at4CL2* and *cbRAS*. B) The total concentrations of donor substrates and HCE products remain stable over time. C) The total concentrations of acceptor substrates and HCE products significantly decreases after 24 h.

The addition of 1 mM L-tyrosine **4** yielded a significant increase in both RA **7** and IA **8** production ($p < 0.001$), suggesting that precursor availability is a significant bottleneck to the productivity of this pathway. Additionally, supplementing 1 mM *p*-coumaric acid **5** or 1 mM caffeic acid **6** resulted in a further increase in RA **7** production ($p < 0.05$), but only 1 mM caffeic acid **6** led to a significantly greater increase in IA **8** biosynthesis than feeding of 1 mM L-tyrosine ($p < 0.01$). These results suggest that, while the availability of L-tyrosine is a limiting factor in the donor substrate biosynthetic pathway, the conversion of L-tyrosine to *p*-coumarate by RsTAL is also a bottleneck, as has been shown in several other engineered hydroxycinnamic acid biosynthetic pathways.[48-50, 54]

Supplementation of 1 mM caffeic acid **6** and 1 mM 4-HPL **2** led to significantly higher RA **7** and IA **8** biosynthesis than any of the other added pathway intermediates ($p < 0.05$), suggesting that acceptor substrate availability and therefore the conversion of 4-HPP **1** to 4-HPL **2** by HdhA may also be a limiting factor. Interestingly, none of the supplemented cultures showed a significant difference in the ratio of IA **8** to RA **7** compared to the unfed culture, except when 1 mM caffeic acid **6** and 1 mM 4-HPL **2** were added, which increased the ratio of IA **8** to RA **7** 10-fold (Table 3.S1). This increase in the ratio of IA **8** to RA **7** due to the addition of 4-HPL **2** illustrates the potential of this or related HCE biosynthetic pathways to be pushed toward production of the desired product by the addition of pathway precursors.

In order to investigate in more detail the flux through the donor and acceptor substrate branches of the engineered RA **7** pathway, we measured the concentrations of pathway intermediates when supplementing cultures with donor or acceptor substrates and precursors over time. When 1 mM caffeic acid **6** and 1 mM 4-HPL **2** were added to cultures, the sum of the molar concentrations of *p*-coumaric acid **5**, caffeic acid **6** and HCE products remained stable over time (Figure 3.5B). By contrast, the sum of the molar concentrations of acceptor substrates (4-HPL **2** and 3,4-DHPL **3**) and HCE products decreased significantly within the first 24 h (Figure 3.5C). This suggests that a leak exists in the acceptor substrate biosynthetic pathway. One possible explanation is that acceptor substrates are being lost to polyphenol formation. Melanin production has been well-documented in *E. coli* through the overexpression of tyrosinase enzymes which catalyze the 3-hydroxylatioin of L-tyrosine **4** to L-DOPA and dopaquinone.[162] Indeed, we observe the accumulation of a dark brown pigment on agar plates or liquid cultures of cells overexpressing *hpaBC* (Figure 3.S3), presumably due to the conversion of endogenous L-tyrosine **4** to L-DOPA.[50, 163-164] Similarly, 3,4-DHPL **3** may be further oxidized to 3,4-dihydroxyphenyllactoquinone, non-enzymatically polymerized to a polyphenol and thus removed from the precursor pool of the HCE biosynthetic pathway.

2.5 Three orthologous RASs give different levels of RA and IA production, but a consistent ratio of IA to RA

Because of the low concentration of HCEs detected in the media and the detection of a high ratio of IA **8** to RA **7** in the media, we sought to compare the *in vivo* activity and

product profile of CbRAS with other RAS orthologs. The DNA encoding two other characterized RAS genes, from *Lavandula angustifolia* (*laRAS*) [152] and *Melissa officinalis* (*moRAS*),[154] was synthesized based on their respective cDNA sequences, and the genes were assembled into the hydroxycinnamoyl transfer module of the HCE biosynthetic pathway, giving pACBB-4laR and pACBB-4moR (Figure 3.1). *E. coli* BW27784 were transformed with pUCBB-ABC, pCDFB-TBC and either pACBB-4cbR, pACBB-4laR, pACBB-4moR or empty pACBB as a control,[157] transformants were cultured in minimal media, and *at4CL2* and RAS-encoding gene expression was induced 24 h after inoculation. At 72 h after induction, there was a significant difference in the concentrations of RA **7** and IA **8** detected with the three different RAS enzymes (Figure 3.6): LaRAS led to significantly higher HCE production than CbRAS ($p < 0.05$), up to 0.9 ± 0.2 μM RA **7** and 3.2 ± 0.4 μM IA **8** in the media; MoRAS led to significantly higher production than both LaRAS and CbRAS ($p < 0.05$), up to 1.8 ± 0.3 μM RA **7** and 5.3 ± 0.7 μM IA **8** in the media. An expression study of these enzymes did not show a difference in protein levels that correlated with HCE production, suggesting that the difference in HCE production levels may be due to differences in catalytic activity rather than expression (Figure 3.S4). Interestingly, the ratio of IA **8** to RA **7** detected did not significantly differ between orthologs ($p \geq 0.2$); the average ratio of IA **8** to RA **7** in all samples was 4.0 ± 0.8 (Table 3.S2). The lack of characterized RAS genes that favor production of RA **7** over IA **8** suggests that an enzyme engineering approach would be

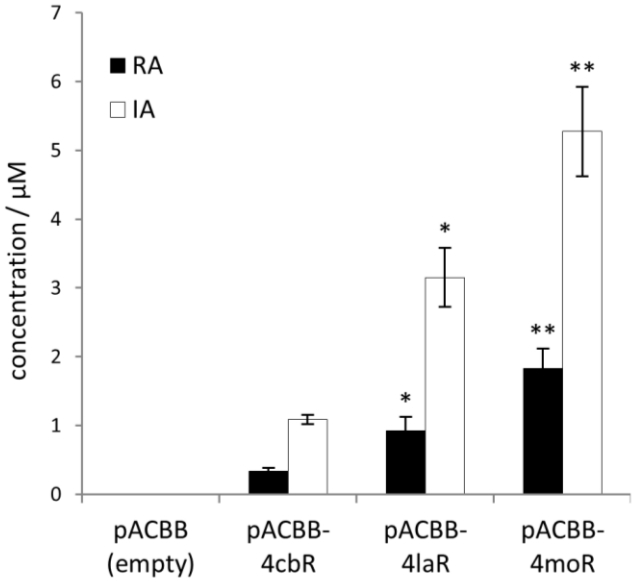


Figure 3.6. HCE biosynthesis in *E. coli* cultures expressing three orthologous RAS enzymes in the context of the redesigned biosynthetic pathway. Two new hydroxycinnamoyl transfer module plasmids were constructed containing *laRAS* and *moRAS*. *E. coli* BW27784 were co-transformed with empty pACBB, pACBB-4cbR, pACBB-4laR or pACBB-4moR, pUCBB-ABC and pCDFBB-TBC and analyzed for HCE production 72 h after induction of the *at4CL2* and RAS-encoding genes. Asterisks denote significantly different concentrations at $p < 0.01$.

necessary to increase RAS substrate specificity toward 3,4-DHPL **3** and thus direct the pathway flux toward the RA **7** product.

3. Conclusions

In this work, we have successfully designed a chimeric biosynthetic pathway tailored for a microbial host that allows heterologous production of the plant-specific HCEs, RA **7**

and IA **8**. To accomplish this, we implemented bacterial enzymes with known biosynthetic utility for hydroxycinnamic acid production, identified new substrates for known bacterial enzymes to establish 4-HPL **2** and 3,4-DHPL **3** biosynthesis in *E. coli*, and incorporated the plant enzymes At4CL2 and RAS to carry out hydroxycinnamoyl transfer. Incorporation of three plant RAS orthologs showed a significant variation in overall HCE biosynthesis, but not in the product profile of the HCEs observed, which remained about 4:1 IA **8** to RA **7** for all three RASs. *E. coli* cultures expressing the RA **7** pathway containing the RAS from *M. officinalis* produced the highest levels of HCEs without precursor supplementation, up to 1.8 ± 0.3 μM RA **7** and 5.3 ± 0.7 μM IA **8**. Because of the modular pathway design, the RA **7** pathway could easily be altered for the biosynthesis of different HCEs through swapping the donor substrate, acceptor substrate and/or the hydroxycinnamoyl transfer modules. Feeding of alternative precursor compounds likewise would allow biosynthesis of structurally diverse HCEs; their economic production, however, would depend on access to inexpensive precursor sources. For example, 4-HPL **2** and 3,4-DHPL **3** are infeasible feeding precursors for industrial applications given their cost (about $370/g and $1850/g, respectively).

Although we achieved RA **7** and IA **8** biosynthesis, the current production levels are far less than sufficient for industrial application. One potential means of optimizing the HCE biosynthetic pathway would be to increase flux through the shikimate and L-tyrosine biosynthetic pathways of the host strain. Indeed, the levels of HCEs we observe are comparable to previously reported levels of the hydroxycinnamic acid amide

avenanthramide D produced from glucose in *E. coli* prior to optimizing L-tyrosine and shikimate pathway flux.[54] In that work, a more than 100 fold increase in avenanthramide D production was achieved through co-overexpression of rate-limiting enzymes and feedback-resistant mutant enzymes to increase flux and decrease feedback inhibition of the shikimate and L-tyrosine pathways.[54] Introduction of our engineered HCE pathway into a tyrosine- or shikimate-overproducing strain would likely have a similar effect on production levels.

During the completion of this work, Yao et al. showed the production of 3,4-DHPL **3** (also known as salvianic acid A or danshensu) in *E. coli* through the overexpression of a mutant D-lactate dehydrogenase from *Lactobacillus pentosus* and HpaBC.[165] The key mutations they used had been shown previously to convert the enzyme from a D-lactate dehydrogenase to a D-hydroxyisocaproate dehydrogenase [166] and increase activity against phenylpyruvic acid.[167] While Yao et al. do not report on the level of *de novo* production of 4-HPL **2** or 3,4-DHPL **3** without precursor feeding or engineering of upstream pathways for L-tyrosine overproduction, they were able to show up to 7.1 g/L 3,4-DHPL **3** with their tyrosine-overproducing strain in a batch-fed process.[165] The implementation of our 3,4-DHPL **3** biosynthetic pathway in a tyrosine-overproducing strain could therefore lead to similarly high titers of 3,4-DHPL **3**.

Several downstream pathway steps are also prime targets for optimization, such as TAL activity, which has been shown before to limit *E. coli* hydroxycinnamic acid biosynthetic pathways.[48-50, 54] Another optimization strategy must address the problem

of acceptor substrate diversion. This could be achieved by co-localization of the acceptor substrate biosynthetic pathway and the RAS, which has proved to be a successful strategy to increase the flux through heterologous biosynthetic pathways.[77] Besides an increase in production levels, the ability to eliminate byproducts and tune the engineered pathway towards the biosynthesis of RA **7** alone, rather than RA **7** and IA **8**, would increase the value of this pathway. For example, the ratio of RA **7** to IA **8** could be increased either by increasing the catalytic efficiency of 4-HPL **2** hydroxylation to 3,4-DHPL **3** and thus decreasing the pool of 4-HPL **2** in the cell or by altering a RAS to more specifically bind 3,4-DHPL **3**. In summary, given the numerous options for pathway optimization, this work provides the basis for the development of a valuable microbial production platform for RA **7** and other HCEs.

4. Experimental Section

4.1 Chemicals and Enzymes

Chemicals used in this study were procured as follows: L-tyrosine **4**, caffeic acid **6**, 4-hydroxyphenylpyruvic acid **1**, DL-4-hydroxyphenyllactic acid, rosmarinic acid **7** (Sigma Aldrich, St. Louis, MO, USA), *p*-coumaric acid **5** (MP Biomedical, LLC, Solon, OH, USA), sodium danshensu (3,4-dihydroxyphenyllactic acid, Select Chemicals, Houston, TX, USA), L-(+)-arabinose (ICN Biomedical, Inc., Aurora, OH, USA). The following enzymes and kits were used for DNA manipulation according to the manufacturers' protocols: Wizard® Genomic DNA Purification Kit (Promega, Madison, WI, USA), RNeasy Plant

Mini Kit (Qiagen, Hilden, Germany), SuperScript® III Reverse Transcriptase, Zero Blunt® TOPO® PCR Cloning Kit, (Life Technologies, Carlsbad, CA, USA), restriction enzymes, T4 DNA ligase, Phusion® High Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA).

4.2 Strains, Media and Growth Conditions

E. coli C2566 (New England Biolabs, Ipswich, MA, USA) was used for plasmid manipulation and propagation and was cultivated in Luria-Bertani (LB) medium at 37 °C. *L. delbrueckii* subsp*. bulgaricus* (DSM-20081) was cultivated anaerobically in MRS broth (DSMZ Medium 11) at 37 °C. *S. espanaensis* (DSM-44229) was cultivated on GYM agar plates (DSMZ Medium 65) at 28 °C. All biosynthesis experiments were carried using *E. coli* BW27784 (*E. coli* Genetic Stock Center, New Haven, CT). [119] Strains used in this study are listed in Table 3.S3.

For production experiments, seed cultures of *E. coli* BW27784 containing pathway plasmids were cultivated from single colonies overnight and grown in LB media at 37 °C with the appropriate antibiotics: ampicillin (100 µg/mL), chloramphenicol (50 µg/mL), and/or streptomycin (50 µg/mL). After overnight growth these seed cultures were used to inoculate cultures in modified M9 media containing yeast extract (1.25 g/L) and glycerol (5 g/L) instead of glucose and supplemented with the appropriate antibiotics. For cultures to test the biosynthesis of 4-HPL **2**, the bioconversion of *p*-coumaric acid **5** and 4-HPL **2** to caffeic acid **6** and 3,4-DHPL **3**, and the relative

expression levels of three RAS orthologs, flasks cultures (100 mL) were inoculated with the seed culture (1 mL); for all other experiments, test tube cultures (10 mL) were inoculated with the seed culture (100 µL). Where appropriate, the expression of *at4CL2* and RAS-encoding genes was induced 24 h after culture inoculation by the addition of arabinose to a final concentration of 0.2 %w/v. For precursor supplementation experiments, *p*-coumaric acid **5**, caffeic acid **6**, or caffeic acid **6** plus 4-HPL **2** (50 µL each) were added to the cultures from 200 mM stocks in methanol to a final concentration of 1 mM after 24 h at the same time as induction; L-tyrosine was added to a final concentration of 1 mM by the addition of dry powder (2 mg) at the time of induction. All biosynthesis experiments were carried out in three biological replicates.

4.3 Gene Cloning, Plasmid Construction and Enzyme Expression

All gene cloning and pathway assembly was carried out using our custom set of BioBrick™-compatible vectors [157] which are available from Addgene (www.addgene.org). The open reading frame (ORF) encoding cbRAS (GenBank: CAK55166.1) was amplified from *C. blumei* cDNA, which was synthesized from RNA extracted from plant tissue donated by the University of Minnesota Horticulture Club. An ORF encoding At4CL2 (GenBank: AAD47192.1) was amplified from *A. thaliana* cDNA which was a generous gift from the Olszewski laboratory at the University of Minnesota. The gene encoding D-2-hydroxyisocaproate dehydrogenase HdhA (GenBank: CAI97812.1) was amplified from *L. delbrueckii* subsp. *bulgaricus* genomic DNA in two pieces using mutagenic primers designed to remove an EcoRI site in the ORF. The

resulting PCR products were amplified in an overlap-extension PCR to give the full-length sequence with a silent mutation eliminating the EcoRI site. The gene encoding Sam5 (GenBank: CCH33123.1) was amplified from *S. espanaensis* genomic DNA. The genes encoding HpaB (GenBank: CAQ34705.1) and HpaC (GenBank: AAZ90964.1) were amplified individually from *E. coli* BL21 (DE3) genomic DNA. The gene encoding RsTAL from *R. sphaeroides* (GenBank: ABA81174.1) was amplified from pUC-TAL [11] in two pieces with mutagenic primers to remove the EcoRI site, as described above for HdhA. ORFs encoding LaRAS (GenBank: ABI48360.1) and MoRAS (GenBank: CBW35684.1) were synthesized with their native codon usage (Life Technologies, Carlsbad, CA, USA). All genes were amplified with forward primers containing either a BglII or a NdeI site at the 5' end, and a reverse primer containing either an NsiI or NotI site and a stop codon at the 5' end. The genes for At4CL2 and CbRAS were first TOPO-cloned using the Zero Blunt® TOPO® PCR Cloning Kit according to the manufacturer's protocol. Plasmids TOPO-cbRAS and TOPO-4CL2 were used for future sub-cloning into pUCBB. All other amplified genes were digested with the appropriate restriction enzymes and ligated into pUCBB digested with the same enzymes. The resulting pUCBB-gene plasmids were used for subsequent subcloning and pathway assembly. Full primer sequences are listed in Table 3.S4.

An initial screen of *at4CL2* and *cbRAS* expression did not show visible expression on an SDS-PAGE gel (Figure 3.S5A). Therefore, an N-terminal His-tag was added to each of these enzymes by amplification of the genes with forward primers

containing NdeI sites and the reverse primers containing NotI sites and subsequent digestion and cloning into pUCBB-ntH6 [157], which encodes a His-tag and thrombin cleavage site just upstream of the NdeI site. Subsequent expression studies showed that N-terminally His-tagged At4CL2 protein was visible *via* SDS-PAGE (Figure 3.S5B). Additionally, an anti-His-tag Western blot showed the presence of both N-terminally-His-tagged At4CL2 and CbRAS (Figure 3.S5C). Because of the apparent increase in At4CL2 protein levels and the confirmed presence of CbRAS protein, the N-terminally-tagged versions of *at4CL2* and *cbRAS* were used for all subsequent subcloning and cultivations. Additionally, *laRAS* and *moRAS* were cloned into pUCBB-ntH6 before further subcloning, such that all RAS enzymes tested carried the same N-terminal His-tag.

The assembly of plasmids containing multiple enzymes was carried out *via* the BioBrick™ prefix-stacking strategy as described in [157]. Briefly, for example, after cloning *hpaB* and *hpaC* individually into the pUCBB backbone, giving pUCBB-hpaB and pUCBB-hpaC; pUCBB-hpaC was then digested with EcoRI and XbaI, while pUCBB-hpaB was digested with EcoRI and SpeI. The resulting DNA fragment containing the modified constitutive lac promoter $P_{lac'}$, [157] the ribosome binding site and the hpaB ORF was subsequently ligated into the digested pUCBB-hpaC, to give pUCBB-hpaBC, a plasmid containing a BioBrick™ comprised of a distinct expression cassette for each gene. All plasmids containing multiple genes were constructed in this way by prefix-stacking of the appropriate cassette into the appropriate plasmid with the pUCBB backbone. Once the multi-enzyme BioBrick™ was constructed in the pUCBB backbone, if needed, it was

subsequently moved to the pACBB or pCDFBB backbones [157] through digestion with EcoRI and SpeI and ligation to the desired backbone digested with the same enzymes. A list of plasmids created for this study and their descriptions is given in Table 3.S5.

4.4 HPLC and LC-MS Analysis of Pathway Intermediates and Products

For biosynthesis and feeding experiments, cultures were sampled at various timepoints by removing a sample of the culture (1 mL), pelleting the cells, and storing the supernatant at -20 °C. Samples were later thawed at room temperature, acidific with 1N hydrochloric acid (50 µL) and extracted twice with ethyl acetate (2 mL). The ethyl acetate fraction was dried to completion and resuspended in methanol (100 µL). For the detection and quantification of pathway intermediates, samples were separated on a Zorbax SB-C18 column (5 uM, 4.6 X 150 mm) on an Agilent 1100 Series system. A flow rate of 1 mL/min according was used according to the following timetable: 5 min at 10% methanol, a gradient to 50% methanol for 20 min, and 10 min post-run at 10% methanol. For the detection and quantification of HCEs, samples were separated on a Zorbax SB-C18 column (5 uM, 4.6 X 250 mm) at a flow rate of 1 mL/min according the following timetable: 5 min at 45% methanol, a gradient to 55% methanol for 10 min, and 5 min post-run at 45% methanol. All buffers contained trifluoroacetic acid (0.1 % v/v). RA 7, IA 8 and caffeic acid 6 were detected at 330 nm; 4-HPL 2 and 3,4-DHPL 3 were detected at 223nm; p-coumaric acid 5 was detected at 310 nm. The identities and quantities of all intermediates and products except IA 8 were verified by the comparison of retention times and UV spectra to authentic standards. Because IA 8 is not commercially

available, and due to the similarity of the UV spectra of IA 8 and RA 7 (Figure 3.S2), IA 8 was quantified using the RA 7 standard curve. For LC-MS analysis of HCEs, the HPLC method was transferred to a Thermo Finnigan Spectra System HPLC in line with a Thermo Finnigan LCQ ion trap. The same HPLC method was used as the one listed for HCE detection above, except that for elution the trifluoroacetic acid was replaced with glacial acetic acid (0.1 % v/v). Negative-mode ESI was used for ionization, and ms/ms scanning events were set up for parent ion masses for RA 7 (359.2 ± 2 m/z) and the mass of dehydroxylated RA 7 (343.2 ± 2 m/z) using 30% ionization energy for fragmentation.

5. Supporting Information

Table 3.S1. Product profile of HCE biosynthesis in cultures supplemented with pathway precursors. Measurements made were taken 48 h after induction of 4CL2 and cbRAS and supplementation. Values reported are mean ± 95% confidence intervals; t-tests were carried out for two-tailed distributions assuming unequal variance.

supplement(s) added (1 mM)	HCE production			p-value *vs.* none
	[RA] / µM	[IA] / µM	IA : RA	
none	0.14 ± 0.05	0.67 ± 0.01	5.0 ± 1.5	-
L-tyrosine	0.76 ± 0.04	3.0 ± 0.2	4.0 ± 0.2	0.33
p-coumaric acid	1.4 ± 0.2	4.2 ± 1.3	3.0 ± 0.5	0.11
caffeic acid	1.2 ± 0.2	5.5 ± 0.3	4.8 ± 0.5	0.81
caffeic acid and 4-HPL	3.6 ± 0.9	170 ± 50	48 ± 3	0.0002

Table 3.S2. Product profile of HCE biosynthesis in cells expressing three RAS orthologs 72 h after RAS and at4CL2 induction. Values reported are mean ± 95% confidence intervals; t-tests were carried out for two-tailed distributions assuming unequal variance.

	HCE production			p-value *vs.*		
	[RA] / µM	[IA] / µM	IA : RA	CbRAS	LaRAS	MoRAS
CbRAS	0.3 ± 0.1	1.1 ± 0.1	3.3 ± 0.5	-		
CaRAS	0.9 ± 0.2	3.2 ± 0.5	3.5 ± 0.4	0.64	-	
MoRAS	1.8 ± 0.3	5.2 ± 0.6	2.9 ± 0.6	0.39	0.20	-

Table 3.S3. Strains used in this study.

Strain	**Description**	**Source**
Escherichia coli C2566	*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--*Tet^S*)2 [dcm] R(zgb-210::Tn10--*Tet^S*) endA1 Δ(mcrC-mrr)114::IS10*	New England Biolabs, Ipswich, MA, USA
Escherichia coli BW27794	*lacI*q *rrnB3* D*lacZ4787 hsdR514* DE(*araBAD*)*567* DE(*rhaBAD*)*568* DE(*araFGH*) U(D*araEp* PCP")±*araE*)	[119]
Lactobacillus delbrueckii subsp. *bulgaricus*	LMG6901 = NCIB11778; isolated from Bulgarian yoghurt	DSM-20081, [159]
Saccharothrix espanaensis	Type strain	DSM-44229, [18]

Table 3.S4. Primers used for gene amplification

Target Gene	Primer	RE Site	Sequence
hdhA	Forward	NdeI	5’ – GGATCCCATATGACTAAAATTGCCATG – 3’
	Reverse	NotI	5’ – ATATATGCGGCCGCTTACAGGTTAACGATGC – 3’
	Mutagenic Forward	-	5’ – GCAAGATCGGGGAGTTCCGCTACCGTATGG – 3’
	Mutagenic Reverse	-	5’ – CCATACGGTAGCGGAACTCCCCGATCTTGC – 3’
rsTAL	Forward	BglII	5’ – GCCCAGATCTATGCTCGCCATGAGCCCC – 3’
	Reverse	XhoI	5’ – CTCGAGCTCGAGTCAGACGGGAGATTGC – 3’
	Mutagenic Forward	-	5’ – CCCGCCGGGTTAAATTCCGGCTTCATGG – 3’
	Mutagenic Reverse	-	5’ – CCATGAAGCCGGAATTTAACCCGGCGGG – 3’
sam5	Forward	BglII	5’ – GCCCAGATCTATGACCATCACGTCACCTGCGCCG – 3’
	Reverse	NotI	5’ – AGATCTGCGGCCGCTCAGGTGCCGGGGTTGATCA – 3’
hpaB	Forward	NdeI	5’ – GAAGGAGGAGATCTGGATCCATAT GAAACCAGAAGATTTCCGCGCCAGTACCC – 3’
	Reverse	NsiI	5’ – GCTCGAGGCGGCCGCCATGGATGC ATTTATTTCAGCAGCTTATCCAGCATGTTG – 3’
hpaC	Forward	NdeI	5’ – GTAGAAGGAGGAGATCTGGATCCA TATGATGCAATTAGATGAACAACGCCTGCGC – 3’

	Reverse	NsiI	5’ – GCTCGAGGCGGCCGCCATGGATGCA TTTAAATCGCAGCTTCCATTTCCAGCATC – 3’
at4CL2	Forward	BglII	5’ – GCCCAGATCTATGACGACACAAGATGTG – 3’
	Forward	NdeI	5’ – CATAGGCATATGACGACACAAGATGTG – 3’
	Reverse	NotI	5’ – AGATGTGCGGCCGCCTAGTTCATTAATC – 3’
cbRAS	Forward	BglII	5’ – GCCCAGATCTATGAAGATAGAAGTCAAAGACTCG – 3’
	Forward	NdeI	5’ – CATAGGCATATGAAGATAGAAGTCAAAGACTC – 3’
	Reverse	NotI	5’ – AGATATGCGGCCGCTCAAATCTCATAAAACAAC – 3’

Table 3.S5. Plasmids used in this study.

Plasmid	**Description**	**Source**
Standard Plasmid Backbones		
pUCBB	pMB1 (colE1); Amp^r $P_{lacP'}$	[157]
pACBB	p15A; Cm^r $P_{lacP'}$	[157]
pCDFBB	CloDF13; $Strep^r$ $P_{lacP'}$	[157]
pUCBB-pBAD	pMB1 (colE1); Amp^r P_{araBAD}	[157]
pUCBB-ntH6	pMB1 (colE1); Amp^r P_{lacP}; N-terminal 6xHis-tag and thrombin cleavage site (*ntH6*)	[157]
pCR™-Blunt II-TOPO®	pMB1 (colE1); Kan^r P_{lac}	Life Technologies, Carlsbad, CA, USA

Partial Pathway Plasmids		
pUCBB-rsTAL	pUCBB:: P_{lacP}-*rsTAL*	This study
pUCBB-hdhA	pUCBB:: P_{lacP}-*hdhA*	This study
pUCBB-sam5	pUCBB:: P_{lacP}-*sam5*	This study
pUCBB-hpaBC	pUCBB:: P_{lacP}-*hpaB*- P_{lacP}-*hpaC*	This study
pUCBB-at4CL2	pUCBB:: P_{lacP}-*at4CL2*	This study
pUCBB-ntH6-at4CL2	pUCBB:: P_{lacP}-*ntH6-at4CL2*	This study
pUCBB-pBAD-at4CL2	pUCBB:: P_{araBAD}*ntH6-at4CL2*	This study
pUCBB-cbRAS	pUCBB:: P_{lacP}-*cbRAS*	This study
pUCBB-ntH6-cbRAS	pUCBB:: P_{lacP}-*ntH6cbRAS*	This study
pUCBB-pBAD-cbRAS	pUCBB:: P_{araBAD}*ntH6-cbRAS*	This study
pUCBB-pBAD-laRAS	pUCBB:: P_{araBAD}*ntH6-laRAS*	This study
pUCBB-pBAD-moRAS	pUCBB:: P_{araBAD}*ntH6-moRAS*	This study
Acceptor Substrate Plasmids		
pUCBB-ABCR	pUCBB:: P_{lacP}-*hdhA*-P_{lacP}-*hpaB*- P_{lacP}-*hpaC*-P_{lacP}-ntH6-cbRAS	This study
pUCBB-ABC	pUCBB:: P_{lacP}-*hdhA*-P_{lacP}-*hpaB*- P_{lacP}-*hpaC*	This study
Donor Substrate Plasmids		
pCDFBB-TBC4	pCDFBB:: P_{lacP}-*rsTAL*- P_{lacP}-*hdhA*-P_{lacP}-*hpaB*- P_{lacP}-*hpaC*-P_{araBAD}*ntH6-at4CL2*	This study
pACBB-TBC4	pACBB:: P_{lacP}-*rsTAL*- P_{lacP}-*hdhA*-P_{lacP}-*hpaB*-P_{lacP}-*hpaC*-P_{araBAD}*ntH6-at4CL2*	This study
pCDFBB-TBC	pCDFBB:: P_{lacP}-*rsTAL*- P_{lacP}-*hdhA*-P_{lacP}-*hpaB*- P_{lacP}-*hpaC*	This study

Hydroxycinnamoyl Transfer Plasmids		
pACBB-4cbR	pACBB::P_{araBAD}*ntH6-at4CL2*-P_{araBAD}*ntH6-cbRAS*	This study
pACBB-4laR	pACBB::P_{araBAD}*ntH6-at4CL2*-P_{araBAD}*ntH6-laRAS*	This study
pACBB-4moR	pACBB::P_{araBAD}*ntH6-at4CL2*-P_{araBAD}*ntH6-moRAS*	This study

Figure 3.S1. Simultaneous conversion of p-coumaric acid and 4-HPL to caffeic acid and 3,4-DHPL, respectively, by *E. coli* BW27784 transformants of pUCBB-hpaBC. As in Figure 3.3, transformants were cultured in minimal media at 30 °C, but were supplemented with both 1 mM *p*-coumaric acid **5** and 1 mM 4-HPL **2.** Media samples were taken over time and processed for HPLC analysis. Samples were compared to standard compounds in order to identify compound peaks. Solid trace, pUCBB-hpaBC transformants; dotted trace, empty pUCBB transformants (control).

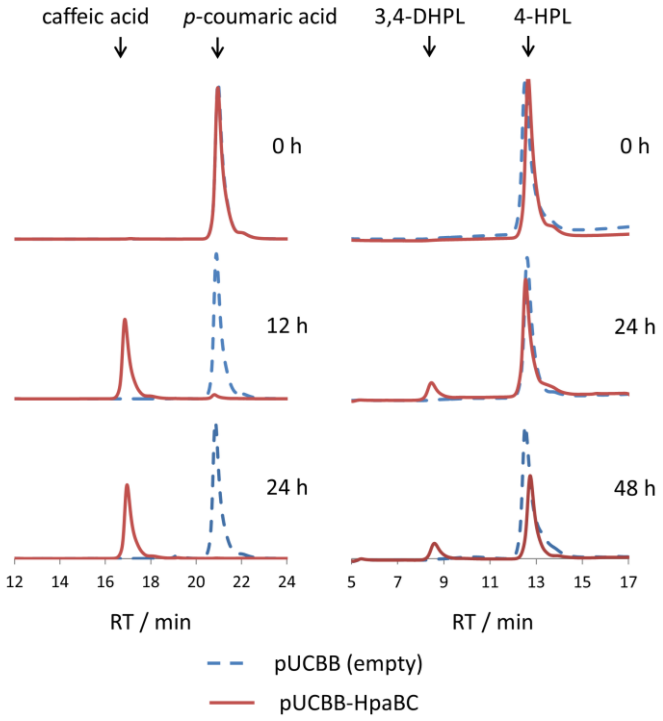


Figure 3.S2. UV spectra of rosmarinic acid and isorinic acid. The chromatogram and spectrum for the RA **7** standard are shown in blue; the chromatogram and spectra for RA **7** and IA **8** produced by *E. coli* BW27784 cells harboring pUCBB-ABC, pCDFBB-TBC, and pACBB-4moR 48 h after induction are shown in blue. Experimental details are as described in the main text.

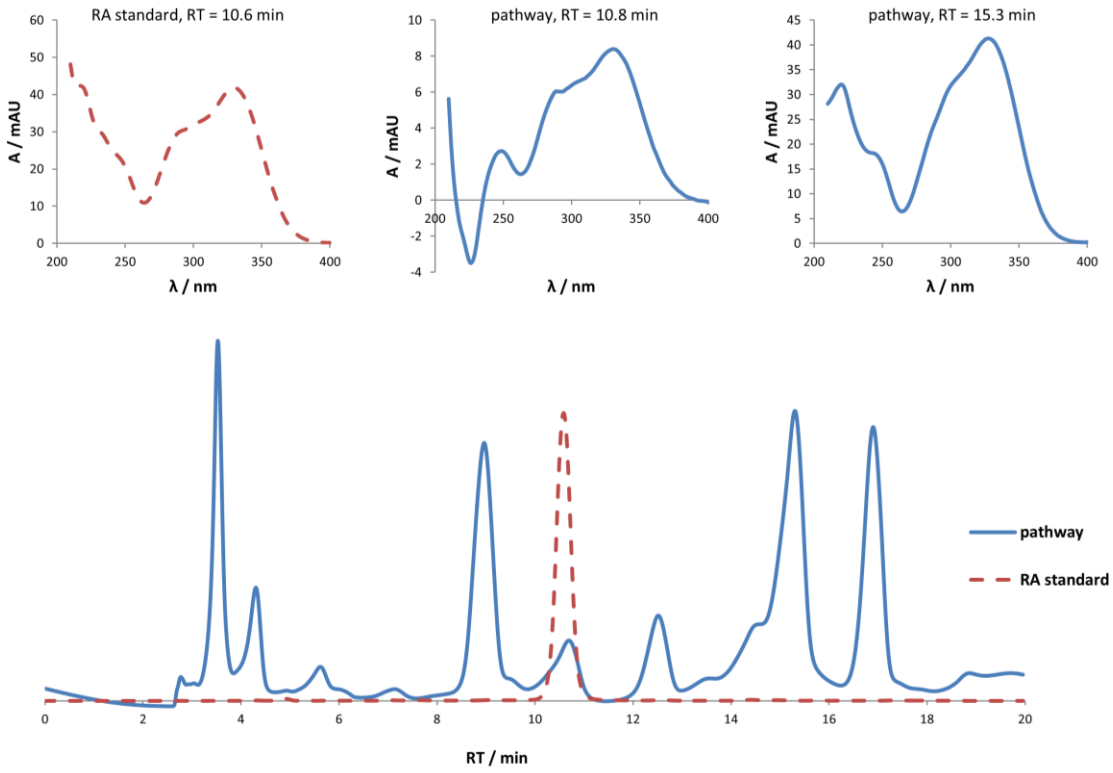


Figure 3.S3. Accumulation of a dark pigment assumed to be melanin in *E. coli* cells expressing *hpaBC*. A) *E. coli* BW27784 cells harboring either pUCBB (empty) or pUCBB-hpaBC streaked on LB-agar plates and cultured overnight at 37 °C. B) *E. coli* BW27784 cells harboring either pUCBB (empty) or pUCBB-hpaBC grown from a single colony in liquid LB media at 37 °C.

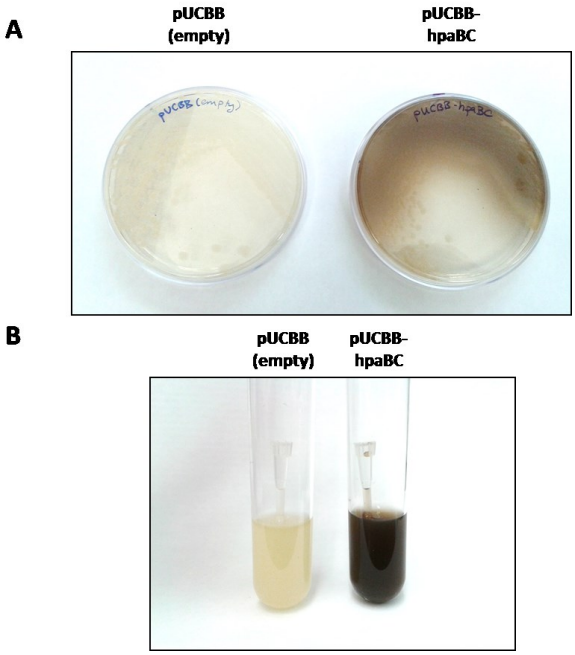


Figure 3.S4. Expression of three RAS orthlogs. *E. coli* BW27784 were co-transformed with pUCBB-ABC, pCDFBB-TBC and pACBB-4cbR, pACBB-4laR or pACBB-4moR and cultured in 100 mL modified M9 media. After 24 h of induction, cell lysates were subjected to anti-His-tag Western blotting to check the relative expression of the three RAS enzymes. T, total lysate; S, soluble fraction.

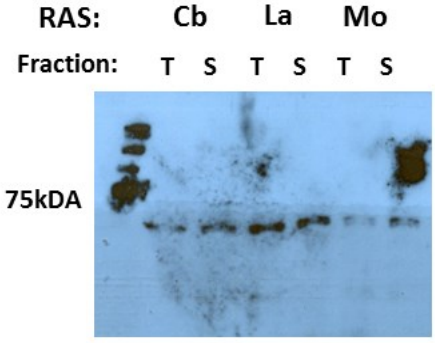
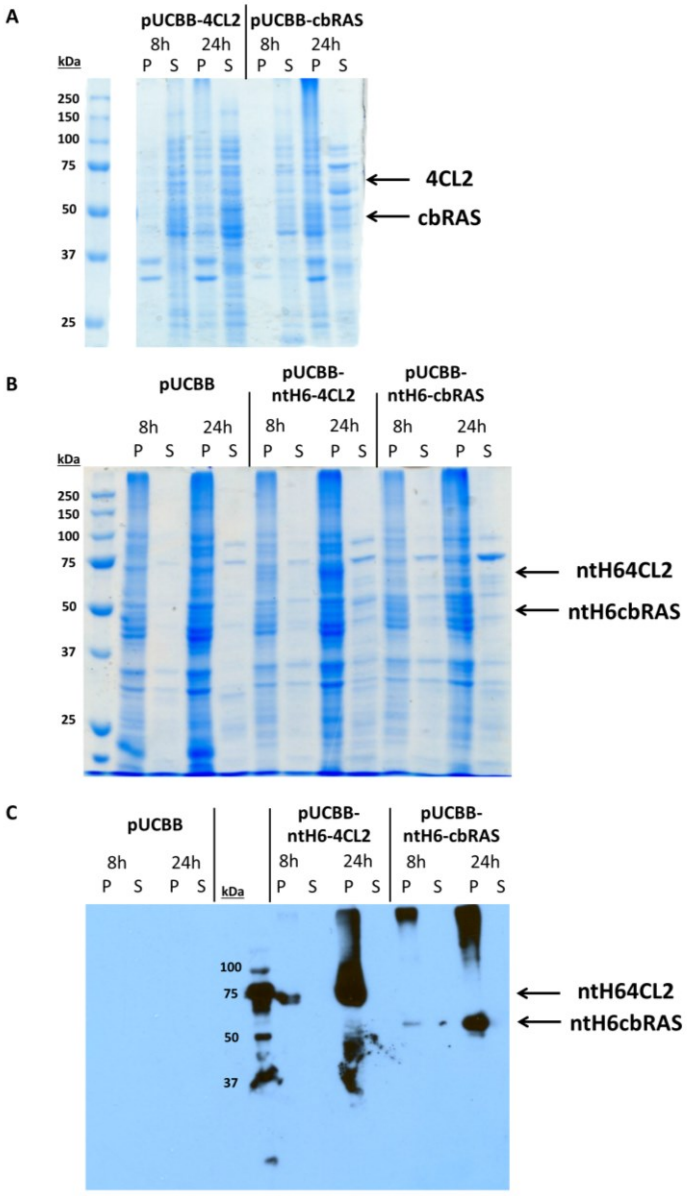


Figure 3.S5. Addition of an N-terminal His-tag to At4CL2 and CbRAS for increased expression levels. P, pellet; S, soluble fraction. A) No visible expression of *at4CL2* and *cbRAS* before addition of an N-terminal His-tag *via* SDS-PAGE. B) SDS-PAGE of At4CL2 and CbRAS after the addition of an N-terminal His-tag shows a visible band for *at4CL2* expression in the pellet fraction after 24h. C) An anti-His-tag Western blot shows translation of both *at4CL2* and *cbRAS* with N-terminal His-tags.



Chapter 4. Exploring Engineered Metabolons: Investigation of Cargo Encapsulation within Eut Bacterial Microcompartments

Summary

Bacterial microcompartments (BMCs) are proteinaceous organelles found in many bacterial taxa for the sequestration of metabolic enzymes and volatile or toxic intermediates. Engineered ethanolamine utilization (Eut) BMCs have shown promise as a means to encapsulate heterologous biosynthetic pathways in *E. coli*. Our group recently showed that the co-expression of Eut BMC shell proteins and a protein equipped with a signal peptide ($EutC^{1-19}$) led to protein encapsulation within Eut BMC shells. In this work, I have explored the possibility of localizing heterologous biosynthetic pathway enzymes to Eut BMCs in order to alter the pathway's product profile. $EutC^{1-19}$-tagged eGFP fusions of the enzyme HpaB localized to recombinant Eut shells, while tagged fusions of HdhA did not. In order to probe the protein-protein interactions governing enzyme encapsulation within Eut BMCs, we created mutants of EutS, a shell protein that is necessary and sufficient for recombinant shell formation and $EutC^{1-19}$-mediated encapsulation. One mutant of the lumenal helix of EutS, $EutS_{L29A}$, disrupted $EutC^{1-19}$-eGFP localization to Eut shells, suggesting that a specific interaction exists between $EutC^{1-19}$ and the lumenal helix of EutS. Ongoing work is underway to identify other Eut encapsulation signal peptides and investigate their interactions with EutS and other Eut shell proteins.

1. Introduction

Bacterial microcompartments (BMCs) are proteinaceous organelles found in a broad range of bacterial taxa.[71, 168] The compartments consist of polyhedral selectively-permeable protein shells which are made up of lattices of proteins with a conserved BMC domain (PF00936).[90] Metabolic enzymes are encapsulated within the shell, often encoded within a single operon in the bacterial genome.[169] In nature, BMCs limit the diffusion of metabolic intermediates, which serves to increase the efficiency of metabolic pathways and/or protect the cell from toxic intermediates. The first BMCs discovered were the carboxysomes in cyanobacteria, which sequester the enzymes carbonic anhydrase and RuBisCo, increasing the local concentration of CO_2 relative to RuBisCo as well as excluding O_2 and preventing the oxygenase reaction of RuBisCo.[170-171] More recently, BMCs involved in the metabolism of non-standard carbon sources – referred to as metabolons – have been characterized. The two best-studied examples are the 1,2-propanediol utilization (Pdu) and ethanolaimine utilization (Eut) BMCs, which serve to sequester propionaldehyde and acetaldehyde, respectively, which would otherwise be toxic to the cell.[90, 172-174] The components of both the Eut and Pdu BMCs are encoded by operons containing open reading frames for multiple BMC-domain proteins and metabolic enzymes.[175-176]

The ability to selectively sequester enzymes and volatile or toxic intermediates in a sub-cellular compartment would have a number of valuable applications in biotechnology. Multiple enzymes in a biocatalytic pathway could be localized to a BMC

and the BMC purified, creating a one-step means of supplying a “package” of catalytic enzymes for *in vitro* biocatalysis. Alternatively, key enzymes in a heterologous *in vivo* biosynthetic pathway could be localized to engineered BMCs to limit the diffusion of certain intermediates. As a first step towards developing a system for the engineering of BMCs, Choudhary et al. showed that heterologously expressing the five shell proteins of the Eut operon from *Salmonella enterica* LT2 (EutSMNLK) in *Escherichia coli* led to the formation of heterologous BMC shells.[87] Interestingly, EutS expressed alone also led to the formation of BMC shells in *E. coli*, while the rest of the shell proteins expressed alone did not. Additionally, Choudhary et al. discovered that an N-terminal signal peptide from EutC, a subunit of the ethanolamine ammonia lyase encapsulated in native Eut BMCs, led to the localization of fluorescent markers and other enzymes into the interior of both EutS-only recombinant shells and EutSMNLK recombinant shells.[87] While other groups have achieved the formation of empty BMC shells in *E. coli* using the Pdu operon from *Citrobacter freundii*[85] and the carboxysome CsoS operon from *Halothiabacillus neapolitanus*[86], these systems required the expression of at least five shell proteins for shell formation. Furthermore, while N-terminal targeting peptides have been shown to allow cargo protein localization within native Pdu BMCs,[88-89] to our knowledge these signal peptides have not been utilized in conjunction with recombinant Pdu shells.

In order to fully utilize the recombinant Eut BMC system for biocatalytic applications, we need to understand the mechanism and dynamics of encapsulation of

cargo proteins within the recombinant shells. While Choudhary et al. showed that the signal peptide EutC$^{1-19}$ is sufficient to encapsulate cargo proteins within the shell, no one has yet demonstrated the encapsulation of non-native metabolic enzymes to recombinant BMCs.[87] Additionally, native Eut compartments contain at least six enzyme components for the metabolism of ethanolamine, suggesting the presence of functional signal peptides on the N-termini of other Eut cargo proteins.[177] The identification of additional functional signal peptides could allow more precise control over the stoichiometry of enzymes localized to heterologous BMCs. Furthermore, an understanding of the protein-protein interactions governing cargo protein encapsulation is critical to the tuning of engineered BMCs. In this work, we explore these aspects of recombinant Eut BMCs in order to build their utility for *in vitro* biocatalysis and metabolic engineering.

2. Results and Discussion

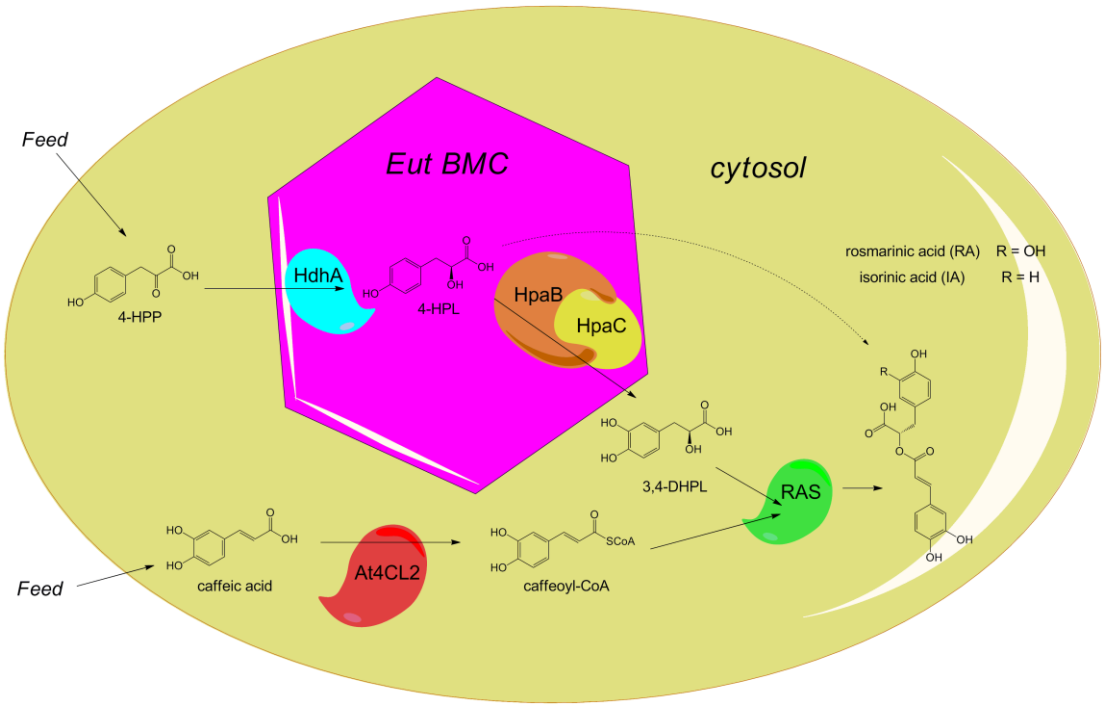
2.1 Exploring the encapsulation of key enzymes to improve a rosmarinic acid biosynthetic pathway

One potential use of engineered Eut BMCs is the encapsulation of enzymes in a recombinant biosynthetic pathway to limit the diffusion of certain metabolic intermediates. This could be of particular use in the novel rosmarinic acid (RA) biosynthetic pathway constructed in Chapter 3. The final enzyme in the pathway, rosmarinic acid synthase (RAS) accepts both 3,4-dihydroxyphenyllactate (3,4-DHPL)

and 4-hydroxyphenyllactate (4-HPL) as substrates, leading to the biosynthesis of both RA and the side product isorinic acid (IA), respectively. While RA has value as an antioxidant and potentially a pharmaceutical (reviewed in [23, 46]), the value of IA as a bioactive compound is unknown. However, in the engineered RA biosynthetic pathway, IA is the favored product, resulting in a ~4:1 ratio of IA to RA produced (Figure 3.6). One approach to decreasing the production of IA and altering the product profile to favor RA biosynthesis would be to decrease the concentration of 4-HPL available to RAS.

We propose that localization of part of the RA biosynthetic pathway – specifically, the enzymes responsible for the conversion of 4-hydroxyphenylpyruvate (4-HPP) to 3,4-DHPL (HdhA and the HpaBC complex) – to the interior of a BMC would limit the diffusion of 4-HPL in the cytosol and thus decrease the availability of 4-HPL to cytosolic 3,4-DHPL (Scheme 4.1). Because little is understood about the selective permeability of Eut BMCs,[177] we cannot assume that it will be permeable to 3,4-DHPL but impermeable to 4-HPL; rather, we hypothesize that the rate of conversion of 4-HPL to 3,4-DHPL will be greater than the rate of 4-HPL diffusion out of the membrane, thus decreasing the cytosolic concentration of 4-HPL. Stochastic simulation of encapsulation of biosynthetic pathways has predicted an increase in pathway flux based on the assumption of general diffusion limitation.[178] To test this hypothesis, we are working to assemble a simplified version of the RA biosynthetic pathway, where 4-HPP and caffeic acid will be fed to cells and At4CL2 will be expressed in the cytoplasm, while HdhA and HpaB will be expressed with the N-terminal $EutC^{1-19}$ signal sequence for localization to the BMC interior (Scheme

4.1). HpaC may be localized to the BMC interior through protein-protein interaction with HpaB; alternatively, an HpaBC protein fusion could be created for the localization of the HpaBC complex to the BMC.



Scheme 4.1. Proposed model of a partially-sequestered RA biosynthetic pathway. The EutC$^{1-19}$ signal sequence will be used to target HdhA and HpaBC to a recombinant Eut BMC; meanwhile, un-tagged RAS and At4CL2 will be expressed in the cytosol. If the rate of conversion of 4-HPL to 3,4-DHPL is greater than the rate of 4-HPL diffusion out of the BMC shell, then the pathway with BMC-localized HdhA and HpaBC will lead to a decreased cytosolic 4-HPL concentration compared with a completely cytosolic pathway. This, in turn, would result in an increase in the ratio of RA to IA in the pathway product profile. 4-HPP, 4-hydroxyphenylpyruvate; 4-HPL,4- hydroxyphenyllactate; 3,4-DHPL, 3,4-dihydroxyphenyllactate. For an explanation of enzyme names and functions, see Chapter 3.

In order to determine the feasibility of this model, we constructed EutC$^{1-19}$-tagged eGFP fusions of HdhA and HpaB and assessed the encapsulation of these proteins in

recombinant Eut BMC shells. These fusions were designed to contain the $EutC^{1-19}$ signal sequence, the enzyme, and a C-terminal fusion of eGFP, with flexible $(Gly\text{-}Ser)_6$-Gly linkers separating the open reading frame (ORF) of the enzyme from both the signal peptide and the eGFP reporter (Figure 4.1). These fusions were cloned into pACBB under the control of a constitutive *lac* promoter to give the plasmids pACBB-$EutC^{1-19}$-HdhA-eGFP and pACBB-$EutC^{1-19}$-HpaB-eGFP. *E. coli* BW27784 cells were cotransformed with one of these fusion constructs or pACBB-$EutC^{1-19}$-eGFP and a Eut shell protein plasmid, either empty pUCBB, pUCBB-EutS, or pUCBB-EutSMNLK; transformants were cultured in minimal media overnight at 30 °C and observed *via* fluorescence microscopy. When expressed in the absence of recombinant Eut shells, $EutC^{1-19}$-HpaB-eGFP is diffuse throughout the cells, while $EutC^{1-19}$-HdhA-eGFP shows polar aggregation in some cells (Figure 4.2). When coexpressed with EutS alone, $EutC^{1-19}$-HdhA-eGFP continues to show fluorescent foci that resemble polar aggregation and are distinct from the slightly-off-center localization pattern of $EutC^{1-19}$-eGFP; $EutC^{1-19}$-HpaB-eGFP is still diffuse throughout the cells (Figure 4.3). Finally, when coexpressed with EutSMNLK, $EutC^{1-19}$-HdhA-eGFP continues to show signs of polar aggregation, while $EutC^{1-19}$-HpaB-eGFP forms loci resembling the localization pattern of $EutC^{1-19}$-eGFP (Figure 4.4). These results suggest that it may be feasible to localize $EutC^{1-19}$-tagged HpaB to recombinant

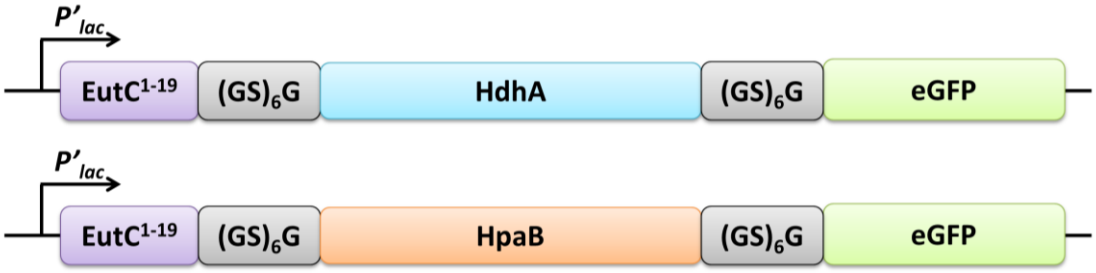


Figure 4.1. Enzyme-eGFP fusion constructs for the visualization of encapsulation into recombinant Eut BMCs. The ORF of each enzyme is separated by a flexible $(Gly\text{-}Ser)_6$-Gly linker (grey) from the signal peptide (purple) and the eGFP ORF. The fusions are expressed constitutively from a modified *lac* promoter.

EutSMNLK shells; however, the expression of the HdhA fusion would need to be tuned to better understand whether it can be localized to a recombinant BMC or not.

If encapsulation of both $EutC^{1\text{-}19}$-HdhA and $EutC^{1\text{-}19}$-HpaB can be achieved individually, the next step will be to verify the colocalization of the enzymes to recombinant BMC shells. This can be done by creating enzyme-fluorophore fusions using two distinct fluorophores that can localize to the recombinant BMC interior. After several failed attempts to achieve localization of red fluorophore to recombinant BMCs, our group has recently achieved the simultaneous colocalization of eGFP and BFP (blue fluorescent protein) to recombinant BMCs (data not shown), which would allow the detection of colocalization of two enzyme-fluorophore fusions. Encapsulation of the two enzymes would also be verified by BMC purification and analysis *via* SDS-PAGE and N-terminal peptide sequencing. This method could also be used to determine whether the protein-protein interaction between HpaB and HpaC is strong enough that $EutC^{1\text{-}19}$-HpaB

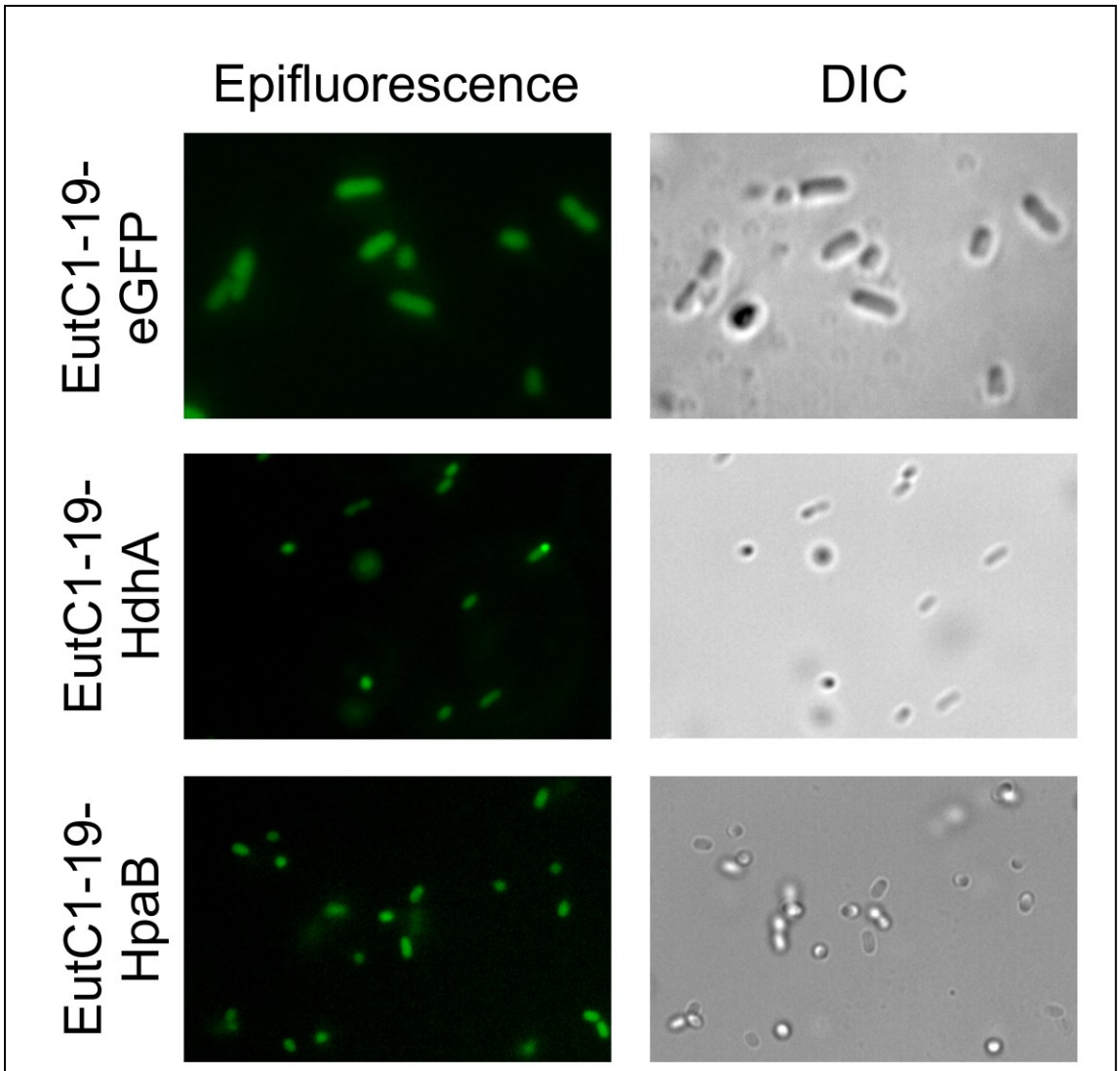


Figure 4.2. Expression of EutC1-19-tagged enzyme-eGFP fusions in the absence of shell proteins. *E. coli* BW27784 cells were transformed with pACBB-EutC$^{1-19}$-eGFP, pACBB-EutC$^{1-19}$-HdhA-eGFP or pACBB-EutC$^{1-19}$-HpaB-eGFP, cultured overnight in minimal media, and subjected to fluorescence microscopy. Expression of EutC$^{1-19}$-eGFP and EutC$^{1-19}$-HpaB-eGFP appears to be diffuse (top and bottom); some cells expressing EutC$^{1-19}$-HdhA-eGFP shows some foci near the poles which appear consistent with aggregation (middle).

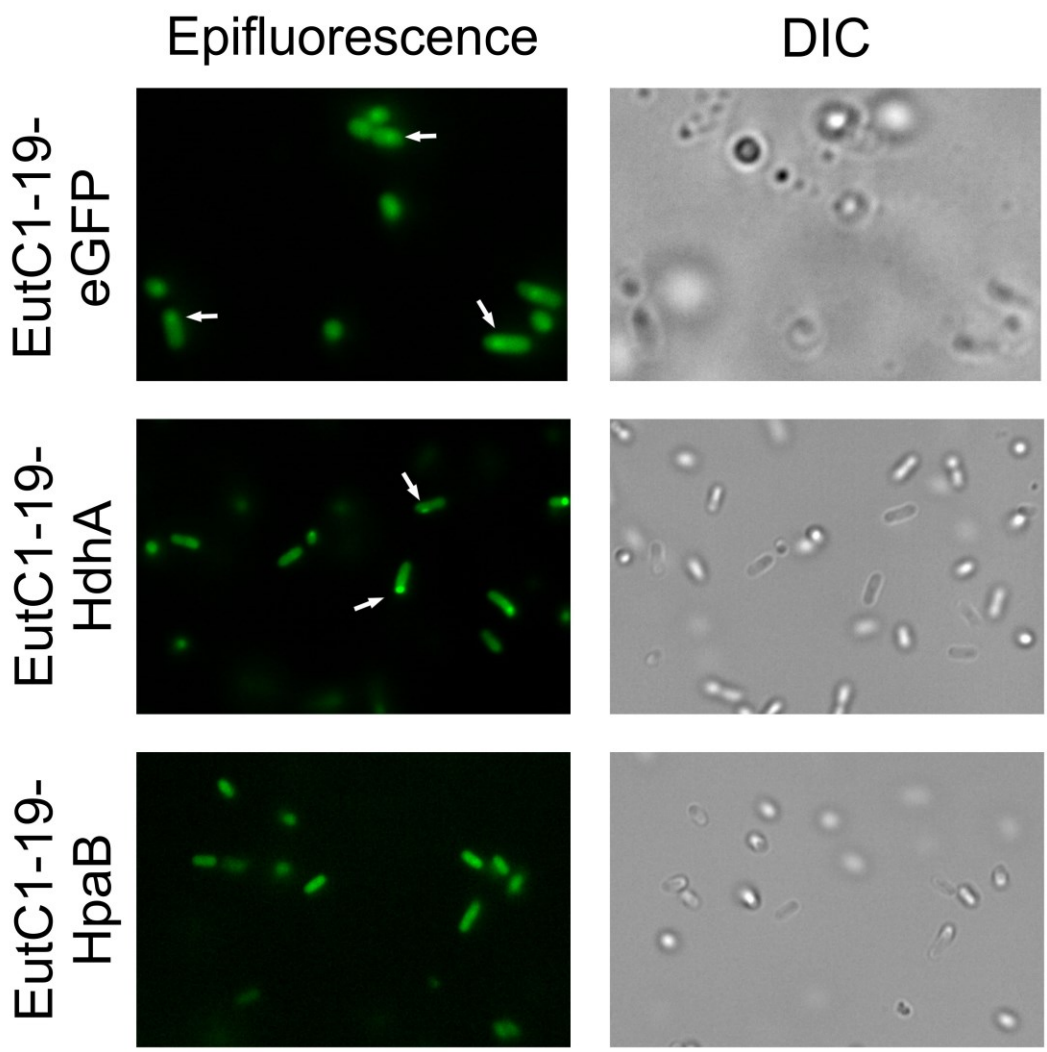


Figure 4.3. Coexpression of EutC1-19-tagged enzyme-eGFP fusions with EutS. *E. coli* BW27784 cells were cotransformed with pACBB-EutC$^{1-19}$-eGFP, pACBB-EutC$^{1-19}$-HdhA-eGFP or pACBB-EutC$^{1-19}$-HpaB-eGFP, cultured overnight in minimal media, and subjected to fluorescence microscopy. Expression of EutC$^{1-19}$-eGFP appears in off-center foci consistent with BMC compartmentalization seen in previous work (top); some cells expressing EutC$^{1-19}$-HdhA-eGFP shows some foci near the poles which appear consistent with aggregation (middle); expression of EutC$^{1-19}$-HpaB-eGFP appears to be diffuse (bottom).

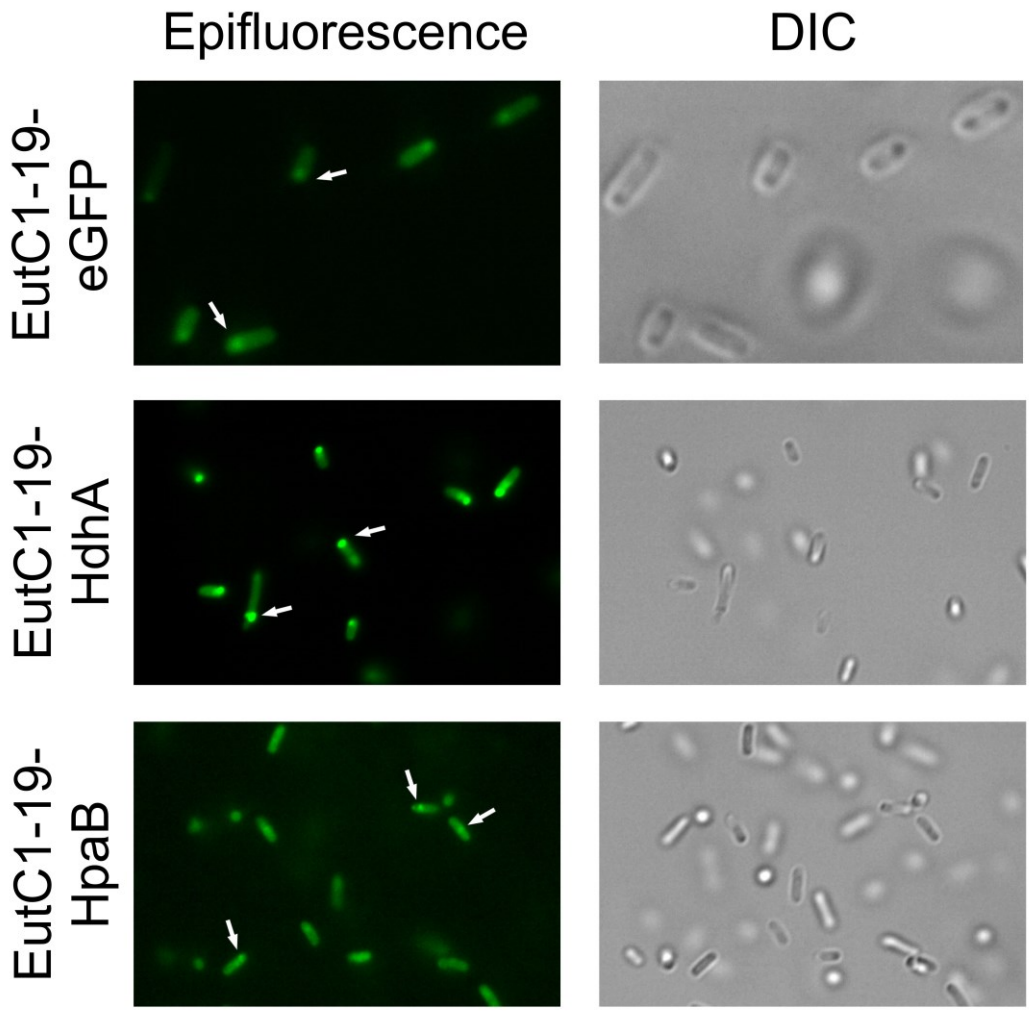


Figure 4.4. Coexpression of EutC1-19-tagged enzyme-eGFP fusions with EutSMNLK. *E. coli* BW27784 cells were cotransformed with pACBB-EutC$^{1-19}$-eGFP, pACBB-EutC$^{1-19}$-HdhA-eGFP or pACBB-EutC$^{1-19}$-HpaB-eGFP, cultured overnight in minimal media, and subjected to fluorescence microscopy. Expression of EutC$^{1-19}$-eGFP and EutC$^{1-19}$-HpaB-eGFP appears in off-center foci consistent with BMC compartmentalization (top and bottom); cells expressing EutC$^{1-19}$-HdhA-eGFP shows foci near the poles consistent with aggregation (middle).

could “pull” HpaC into the BMCs during encapsulation; if not, a fusion of the HpaB and HpaC enzymes could be created to allow the cofactor-recycling complex to be encapsulated. Finally, once HdhA and HpaB are successfully encapsulated *in vivo*, *E. coli* cells could be transformed with plasmids for the expression $EutC^{1-19}$-HdhA, $EutC^{1-19}$-HpaB, At4CL2 and RAS and coexpressed with or without recombinant shell proteins, and the levels of pathway intermediates and products could be measured (Scheme 4.1). If HdhA and HpaB localization within a recombinant BMC does limit the diffusion of 4-HPL, then when the recombinant shell proteins are expressed, we should see a decrease in the detectable levels of cytosolic 4-HPL and an increase in the ratio of RA to IA production from the pathway.

While this work constitutes an important first step towards the encapsulation of metabolic pathways within engineered BMCs, this approach currently has several pitfalls. First, it is possible that $EutC^{1-19}$-tagged HdhA may not localize to the BMCs at all, potentially due to poor protein folding with the $EutC^{1-19}$-tag or interference of HdhA’s quaternary structure with encapsulation. Our repeated failure to localize $EutC^{1-19}$-tagged red fluorescent proteins to recombinant BMCs suggests that some proteins are less efficiently encapsulated than others for reasons we do not yet understand, which limits the application potential of this system. Second, it is unclear why $EutC^{1-19}$-tagged HpaB would be encapsulated by EutSMNLK shells, but not EutS-only shells (Figures 4.3 and 4.4). This again suggests that encapsulation efficiency is dependent not only on the presence of the $EutC^{1-19}$ signal peptide, but also on protein-protein interactions which

currently remain uncharacterized. Finally, because only one signal peptide has been identified that can localize proteins to recombinant Eut BMCs, there is no way to fine-tune the stoichiometry of enzymes encapsulated within a single shell. Ultimately, in order to effectively use recombinant Eut BMCs in biocatalytic applications, we need a more thorough understanding of the mechanism of protein encapsulation within Eut BMCs.

2.2 *In vivo* investigation of the encapsulation mechanism of the EutC$^{1-19}$ signal peptide in recombinant Eut BMCs

Recently, some light has been shed on the mechanism of cargo protein encapsulation with native Pdu-type BMCs. Fan et al. showed that encapsulation of the cargo protein PduP relies on protein-protein interactions of the PduP N-terminal peptide with the lumenal helix of the PduA shell protein, and identified key shell and cargo protein residues mediating this interaction through site-directed mutagenesis.[179] They found that mutating any of three key residues of PduA's lumenal helix – H81, V84 and I88 – led to the disruption of PduP encapsulation within Pdu compartments and abolished the interaction between PduP and PduA, as observed through *in vitro* His-tag affinity bait-and-prey pulldown experiments (Figure 4.5A). They also identified the three residues of the helical PduP signal peptide involved in this interaction – E7, I10, and L14 – through similar experimentation.[179] The authors suggest that this interaction study provides evidence for a "cognate pair" model of cargo protein encapsulation within BMCs, in which certain cargo-shell pairs specifically interact to mediate encapsulation.

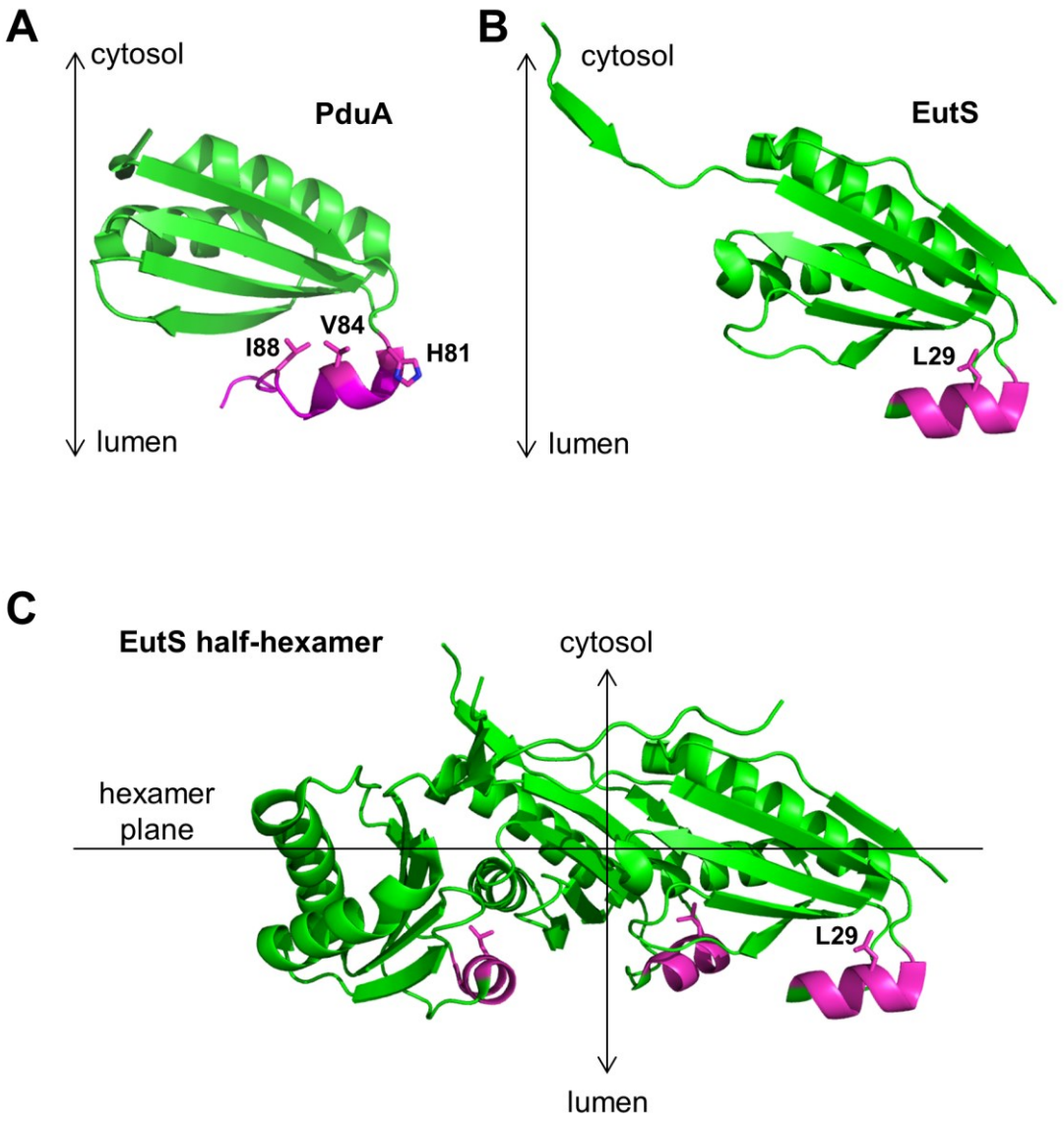


Figure 4.5. Structure of the lumenal helix of BMC shell proteins involved in cargo protein encapsulation. The shell protein lumenal helices are shown in magenta, and residues shown to be involved in cargo protein encapsulation are shown as stick structures and labeled. A, structure of PduA, with encapsulation residues implicated by Fan et al.;[177] B, structure of EutS, showing L29 residue involved in $EutC^{1-19}$-tagged protein encapsulation; C, half of a EutS hexamer, showing little space for interaction of the $EutC^{1-19}$ signal peptide.

While the cognate pair model of encapsulation within BMCs is compelling, direct evidence of the model – i.e. demonstrated specific interaction of another cognate shell-cargo protein pair – has not been published, and bioinformatic analysis offers an alternative model for encapsulation. We observed that the N-termini of multiple Pdu and Eut cargo proteins, including EutC, exhibit the same pattern as the PduP signal peptide, with a ExxOxx(x)O motif, where O is a hydrophobic residue (Figure 4.6). These conserved residues are consistent with a negative-hydrophobic-hydrophobic pattern on one side of an α-helix, which could interact with a positive-hydrophobic-hydrophobic pattern on one side of a shell protein lumenal helix, as is the case with PduA (Figure 4.5). This suggests the possibility of a more degenerate mechanism of cargo protein encapsulation, in which a number of cargo signal peptides could interact with a number of shell lumenal helices for the encapsulation of multiple proteins simultaneously.

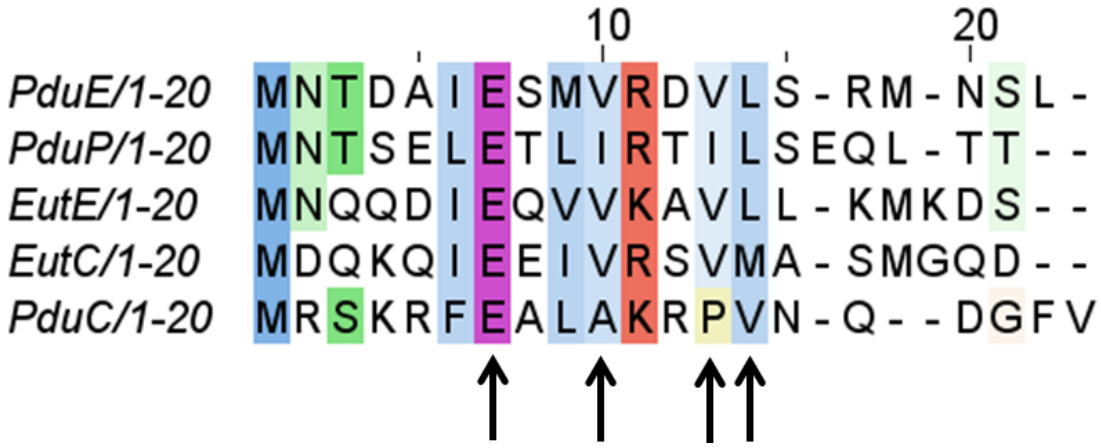


Figure 4.6. Multiple sequence alignment of the N-termini of several Eut and Pdu cargo proteins. Alignment was performed using ClustalW2. Residues of the putative conserved ExxOxx(x)O motif are indicated with arrows.

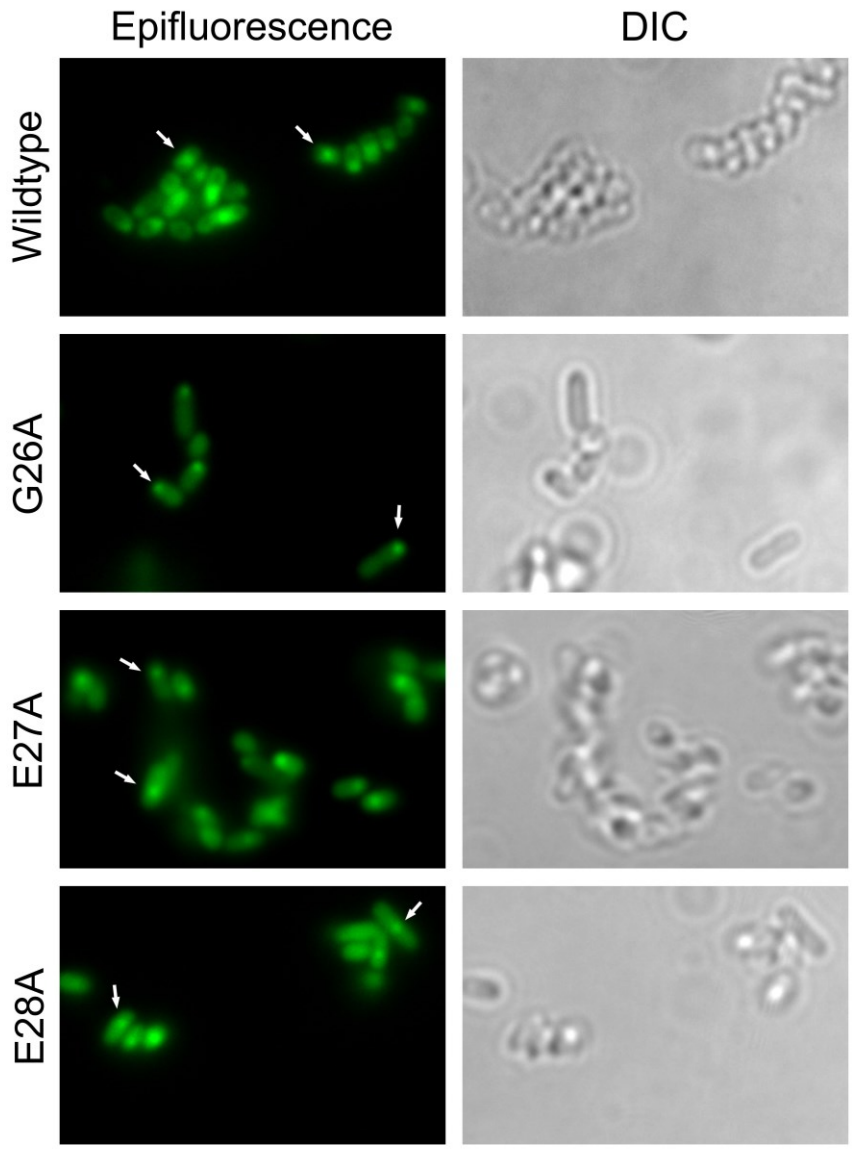


Figure 4.7A. Alanine scanning of EutS lumenal helix to determine key residues in EutC1-19-tagged protein encapsulation. *E. coli* C2566 cells were cotransformed with pACBB-EutC$^{1-19}$-eGFP and a plasmid expressing the indicated EutS mutants from a pUCBB backbone. Cells were cultured overnight at 30 °C and subjected to fluorescence microscopy. All mutants in this panel exhibit foci consistent with encapsulation inside EutS shells.

In order to better understand the mechanism of cargo protein encapsulation in Eut BMCs, we decided to investigate the protein-protein interactions mediating protein encapsulation within recombinant Eut BMCs. So far, we had identified only one viable encapsulation signal peptide, $EutC^{1-19}$, and we presume that it interacts with the shell protein EutS, since it can target proteins to shells composed entirely of EutS (Section 2.1, Figure 4.2). First, to determine whether the lumenal helix of Eut S mediates localization of $EutC^{1-19}$-tagged proteins in a manner analogous to the PduA-PduP interaction, we performed alanine scanning of all of the residues on the lumenal helix of EutS and tested the effect of each mutation on encapsulation of $EutC^{1-19}$-eGFP into EutS-only shells *via* fluorescence microscopy. Unlike the PduA-PduP interaction, in which three residues were found to be necessary for interaction (Figure 4.5A), only two of the EutS residues screened led to abolishment of $EutC^{1-19}$-eGFP encapsulation (Figure 4.7). While one of these non-encapsulating mutants, $EutS_{L29A}$, is consistent with the key encapsulation residues identified on the PduA lumenal helix, the mutant $EutS_{K32A}$ does not fit the positive-hydrophobic-hydrophobic pattern exhibited by PduA (Figure 4.5A and B). These data suggest that, while the lumenal helix of EutS is important for $EutC^{1-19}$-mediated encapsulation within EutS shells, the encapsulation mechanism is defined by a set of protein-protein interactions distinct from those described for PduA and PduP.

In addition to understanding the EutS-$EutC^{1-19}$ interaction, we wanted to probe the validity of the cognate pair and degenerate models of protein encapsulation within

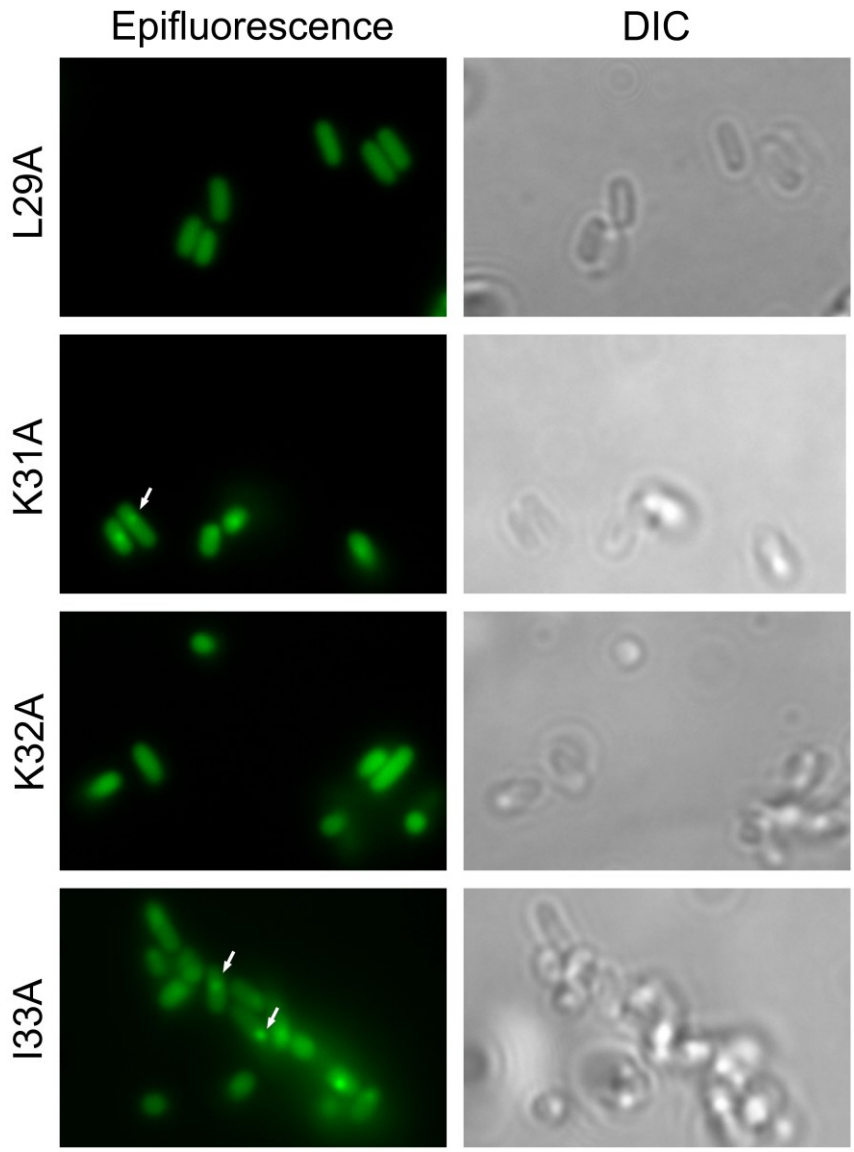


Figure 4.7B. Alanine scanning of EutS lumenal helix to determine key residues in EutC1-19-tagged protein encapsulation. *E. coli* C2566 cells were cotransformed with pACBB-EutC$^{1-19}$-eGFP and a plasmid expressing the indicated EutS mutants from a pUCBB backbone. Cells were cultured overnight at 30 °C and subjected to fluorescence microscopy. Mutants EutS$_{L29A}$ and EutS$_{K32A}$ exhibit diffuse fluorescence, indicating a disruption to EutC$^{1-19}$-eGFP encapsulation.

Eut shells. As a first step, we investigated whether the $EutC^{1-19}$ signal peptide could target proteins for encapsulation *via* interactions with other shell proteins besides EutS, as predicted by the degenerate model, or whether it interacts exclusively with EutS, as predicted by the cognate pair model. Because EutS is the only shell protein which forms BMCs when expressed alone, we are unable to test the localization of $EutC^{1-19}$-tagged proteins to each shell protein individually. However, by coexpressing one of the non-encapsulating EutS mutants with other shell proteins, we can look for $EutC^{1-19}$-tagged protein encapsulation mediated by any of the other four shell proteins. First, to confirm that $EutS_{L29A}$ and $EutS_{K32A}$ are still able to form shells with other shell proteins, we coexpressed them with an mCherry-tagged version of EutM, which exhibit red fluorescent foci indicating incorporation into BMC shells when coexpressed with wild type EutS (Figure 4.8). While mCherry-EutM coexpressed with $EutS_{L29A}$ forms fluorescent foci similar to wild type, coexpression with $EutS_{K32A}$ led to diffuse fluorescence, indicating that $EutS_{K32A}$ is unable to form BMC shells with other shell proteins and suggesting that the K32A mutation affects encapsulation in EutS only shell indirectly by disrupting EutS hexamer formation. Next, we coexpressed $EutS_{L29A}$ with EutM alone, and saw that these compartments were unable to encapsulate $EutC^{1-19}$-eGFP, indicating that EutM is not involved in $EutC^{1-19}$-mediated cargo protein encapsulation (Figure 4.9). Additionally, we coexpressed $EutS_{L29A}$ with the full suite of Eut shell proteins – EutMNLK – to determine if any other shell proteins appear to be involved in $EutC^{1-19}$-mediated encapsulation. We observed no encapsulation of $EutC^{1-19}$-

eGFP with $EutS_{L29A}$MNLK shells, suggesting that $EutC^{1-19}$ only interacts with EutS (Figure 4.10). Taken together, these data support the cognate pair model of cargo protein encapsulation.

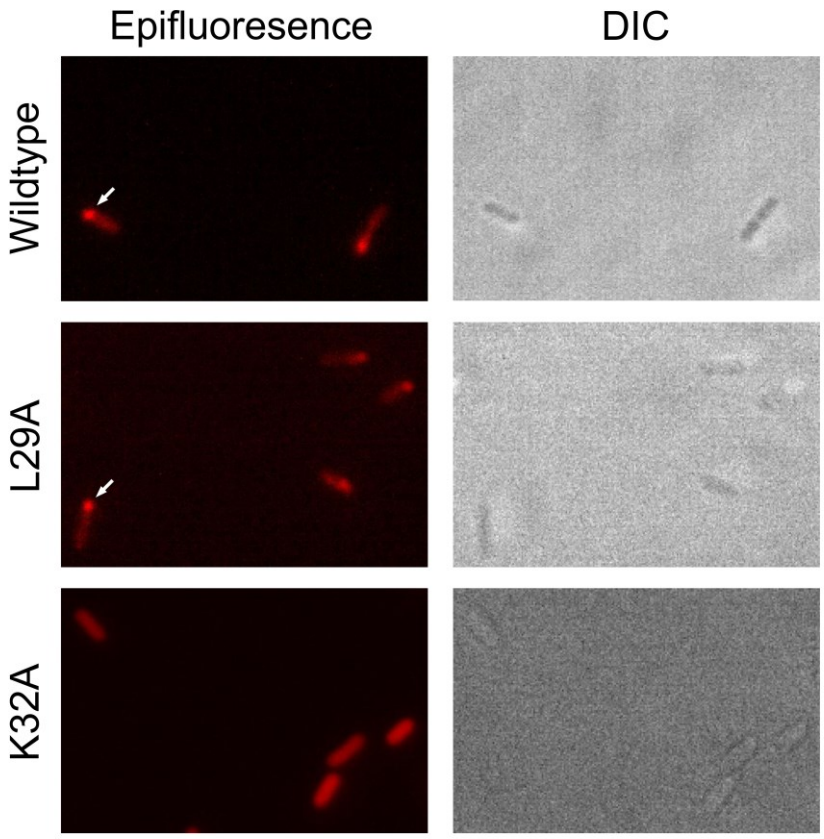


Figure 4.8. Coexpression of EutS mutants with mCherry-tagged EutM to visualize BMC shell formation. *E. coli* C2566 cells were cotransformed with a plasmid containing the indicated EutS mutants and mCherry-EutM on a pACBB backbone. Cells were cultured overnight at 30 °C and subjected to fluorescence microscopy. mCherry-EutM forms foci when expressed with the $EutS_{L29A}$, indicating shell formation; fluorescence is diffuse with S_{K32A} indicating a discruption of shell formation.

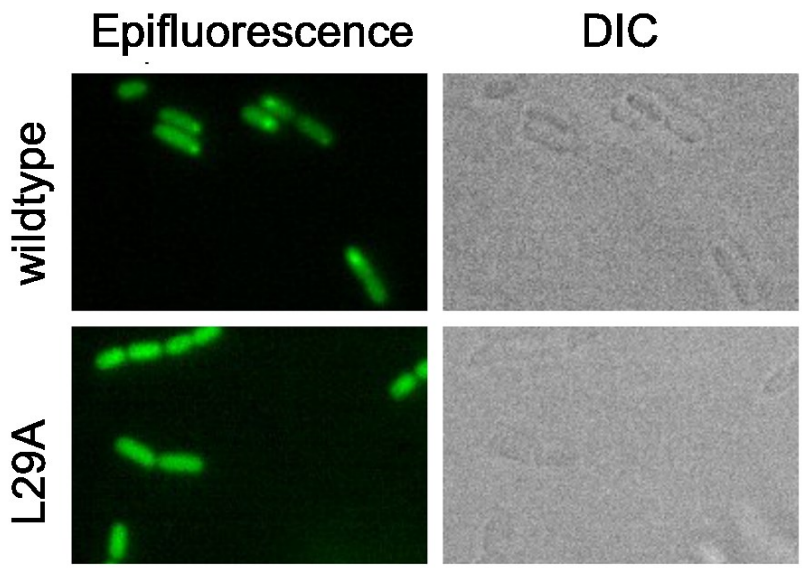


Figure 4.9. Coexpression of EutSL29A and EutM shows no evidence of EutM-mediated encapsulation of EutC1-19-eGFP. *E. coli* C2566 cells were cotransformed with pBBRBB-EutC$^{1-19}$-eGFP and a containing either wild type EutS or EutS$_{L29A}$ and EutM on a pACBB backbone. Cells were cultured overnight at 30 °C and subjected to fluorescence microscopy. Cells expressing EutS$_{L29A}$ exhibit diffuse eGFP fluorescence, indicating no

While these preliminary results support the cognate pair model, more data are needed to establish it as the primary model of cargo protein encapsulation within BMCs. First, while we were unable to show binding of the same signal peptide to multiple shell proteins, it may be true that multiple signal peptides are able to interact with the same shell protein. In order to investigate this, other signal sequences able to mediate encapsulation of cargo to the Eut shells must be identified. Previously, the ethanol

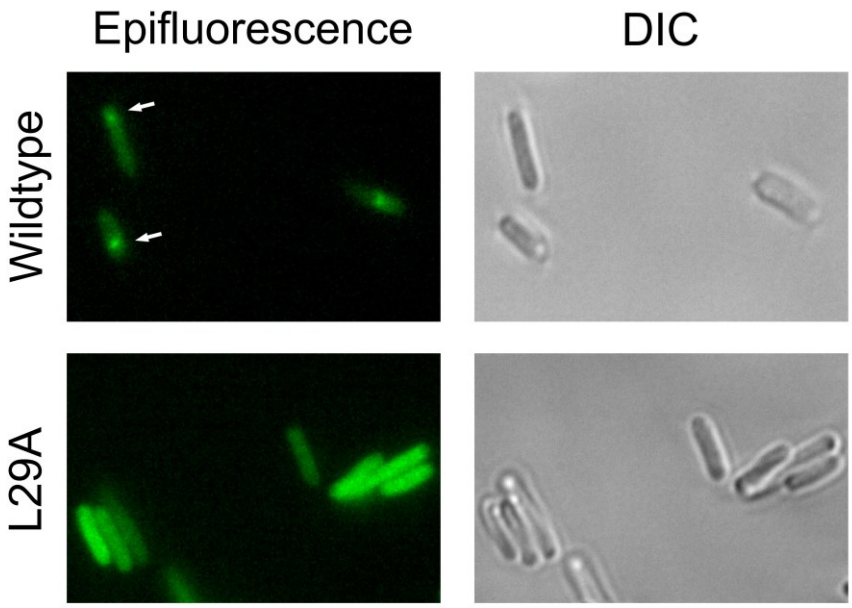


Figure 4.10. Coexpression of EutSL29A and EutMNLK shows no evidence of encapsulation of EutC1-19-eGFP mediated by other shell proteins. *E. coli* C2566 cells were cotransformed with pBBRBB-EutC$^{1-19}$-eGFP and a containing either wild type EutS or EutS$_{L29A}$ and EutMNLK on a pACBB backbone. Cells were cultured overnight at 30 °C and subjected to fluorescence microscopy. Cells expressing EutS$_{L29A}$ exhibit diffuse eGFP fluorescence, indicating no encapsulation.

dehydrogenase EutG was predicted to have an N-terminal signal peptide for encapsulation, based on the presence of an N-terminal extension that is absent in non-BMC-associated ethanol dehydrogenases.[89] Our lab has attempted to show encapsulation of EutG$^{1-19}$-tagged proteins to recombinant Eut BMCs, but no localization was observed.[87] It is possible that EutG does have an N-terminal signal peptide, but that more than the first 19 residues are required to mediate encapsulation; attempts to localize proteins tagged with longer N-terminal extensions of EutG are underway. However, based on the presence of a ExxOxx(x)O motif similar to that found in the

signal peptide of PduP, the N-terminus of EutE could also potentially be an encapsulation signaling peptide for Eut BMCs (Figure 4.6). If an N-terminal portion of EutE is found to mediate the encapsulation of cargo proteins within EutSMNLK shells, then further investigation could be done to determine whether the EutE signal peptide interacts with EutS or one of the other shell proteins. If it is found to encapsulate enzymes within EutS-only shells, then the interaction of multiple signal sequences with the EutS protein would lend credence to the degenerate model of encapsulation. If it does not bind to EutS only shells, but binds to $EutS_{L29A}MNLK$ shells, this would provide the first solid basis for the cognate pair model by suggesting that the EutE signal sequence forms a cognate pair with a second shell protein.

Another way to examine the validity of these models would be *via in vitro* studies of protein-protein interactions between Eut shell and cargo proteins. The development of a method for visualizing shell/cargo protein interactions would allow for the exhaustive assessment of all potential shell-cargo cognate pairs. To this end, we have attempted to show the *in vitro* interaction of shell and cargo proteins through His-tag affinity pull-downs, using EutS and $EutC^{1-19}$-eGFP as a proof of concept, since we have visualized their interaction *in vivo via* microscopy. In this method, the His-tagged "bait" protein is and bound to Talon affinity resin, after which a cell lysate containing the potential binding partner, the "prey," is incubated with the resin. An interaction between the bait and prey would result in an elution showing proteins *via* SDS-PAGE. Initial pull-downs with EutS and $EutC^{1-19}$-eGFP, using C-terminally His-tagged version of each protein as bait, have

showed no interaction. We note that, while we have so far expressed bait and prey in separate cell lines, Fan et al. were able to show interactions between PduA and PduP *via* similar pull-down experiments by expressing both bait and prey in the same cell.[179] It is also notable that, while the lumenal helix is clearly important in shell-cargo interactions that mediate encapsulation (Figure 4.7B), there appears to be little space within the hexameric crystal structure of EutS to accommodate the presence of a helical signal peptide (Figure 4.5C). Based on these observations, we hypothesize that the protein-protein interactions governing encapsulation are relatively transient, with interaction occurring before shell protein hexamer formation, and release of the cargo protein into the lumen of the BMC as shell proteins form the lattices of hexamers that make up the shell. If the EutS in cell lysates used in preliminary pull-down experiments had already formed shells, this would explain the failure to show interaction between EutS and $EutC^{1-19}$-eGFP. To circumvent this problem, we are currently developing alternative approaches to these pull-downs, including expression of bait and prey in the same cell, as well as constitutive expression of $EutC^{1-19}$-eGFP followed by late induction of EutS expression to try and catch the interaction before BMC shell formation can happen. If we can thus establish an *in vitro* method for looking at shell/cargo protein interactions, we will proceed to test the interaction of all possible shell/cargo pairs to determine whether cognate pairs exist, or whether a degenerate pattern of interaction is seen.

3. Conclusions

While engineered BMCs show promise for future applications in *in vitro* biocatalysis and metabolic engineering, there is still much to learn about the mechanism of cargo protein encapsulation in order to utilize their full potential. We have shown the localization of a fluorescent reporter fused to one key enzyme in an engineered RA biosynthetic pathway to the interior of a Eut BMC, yet the localization of a second enzyme was unsuccessful and requires further optimization. In an effort to understand the protein-protein interactions governing encapsulation of $EutC^{1-19}$-tagged proteins, we identified a key residue of the lumenal helix of EutS which, when mutated, disrupts the localization of $EutC^{1-19}$-eGFP to the interior of EutS shells. This allowed us to coexpress other Eut shell proteins with the non-encapsulating $EutS_{L29A}$ mutant determine their involvement in $EutC^{1-19}$-mediated encapsulation. The inability of $EutS_{L29A}$MNLK shells to encapsulate $EutC^{1-19}$-eGFP suggests that EutS alone interacts with the $EutC^{1-19}$ signal peptide, providing circumstantial evidence in support of a shell/cargo cognate pair model of encapsulation. To provide further evidence supporting or dismissing this model, we are working to identify alternative Eut cargo protein signal sequences, as well as to develop a means to detect *in vitro* protein-protein interactions between all possible shell/cargo cognate pairs in the suite of Eut BMC proteins. This work will provide not only a better understanding of cargo protein encapsulation within Eut BMCs but will contribute to the set of tools available for recombinant Eut BMC engineering.

4. Materials and Methods

4.1 Materials, Strains and Growth Conditions

Restriction enzymes, T4 DNA ligase, and Phusion® High Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA) were used for the creation of the EutC$^{1-19}$-tagged enzyme-eGFP fusions, as well as building co-expression plasmids of the EutS mutants with other shell proteins. A QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA, USA) was used for site-directed mutagenesis of EutS. Glass coverslips (22x22mm, 1.5 thickness, Corning Inc., Tewksbury, MA, USA) and BacLight^{TM} mounting oil (Life Technologies, Grand Island, NY, USA) were used in light microscopy for encapsulation studies. Protein structure images were prepared using PyMol (Schrodinger, www.pymol.org), and multiple sequence alignment was carried out using ClustalW2.1 (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

E. coli C2566 (New England Biolabs, Ipswich, MA, USA) was used for all DNA propagation and manipulation, as well as all encapsulation studies of EutS lumenal helix mutants (section 2.2). *E. coli* BW27784 (*E. coli* Genetic Stock Center, New Haven, CT) [119] was used for localization studies with the EutC$^{1-19}$-tagged enzyme-eGFP fusions, since this was the strain used for the construction of the RA biosynthetic pathway (Chaper 3). Cultures for DNA manipulation were grown at 37 °C in Luria-Burtani (LB) medium with the appropriate antibiotics: ampicillin (100 µg/mL), chloramphenicol (50 µg/mL), and/or kanamycin (30 µg/mL). For encapsulation experiments, single colonies

of transformants containing plasmids for Eut shell expression and a EutC$^{1-19}$-tagged fluorescent reporter were cultured in LB medium at 37 °C for six to eight hours; then, these seed colonies were used to inoculate fresh LB cultures to a 1:1000 dilution factor. These cultures were incubated with shaking at 30 °C for 15 h before mounting and microscopy.

4.2 Construction of Plasmids

The EutC$^{1-19}$-tagged enzyme-eGFP fusions were created using the plasmid pACBB-EutC$^{1-19}$-GS-MCS-GS-eGFP, which contains a modified, constitutive lac promoter ($P_{lacP'}$) and ribosome binding site followed by the EutC$^{1-19}$ signal peptide, a (Gly-Ser)$_6$Gly linker, a multiple cloning site, a second (Gly-Ser)$_6$Gly linker, and the ORF for eGFP. The ORFs of *hdhA* and *hpaB* were amplified from the plasmids pUCBB-hdhA and pUCBB-hpaB (Table 3.S5) using forward primers with a Hin*d*III site at the 5' end and a reverse primer with a KpnI site at the 5' end. The resulting PCR products were digested with Hin*d*III and KpnI and ligated into pACBB-EutC$^{1-19}$-GS-MCS-GS-eGFP which had been digested with the same enzymes, resulting in the plasmids into pACBB-EutC$^{1-19}$-HdhA-eGFP and pACBB-EutC$^{1-19}$-HpaB-eGFP.

Alanine scanning of the lumenal helix of EutS was carried out *via* site-directed mutagenesis by site directed mutagenesis, using the plasmid pUCBB-EutS as a template[87]. All codons for residues from Gly_{26} to Ile_{33} were altered to code for Ala, except residue Ala_{30}, which was not mutated. The resulting pUCBB-EutS* mutant

plasmids were used for initial studies of EutC$^{1-19}$ encapsulation (Figure 4.7). For coexpression of the the EutS mutants with mCherry-EutM, wildtype EutM or EutMNLK on the same plasmid, the EutS mutants were inserted into pUCBB-mCherry-EutM, pUCBB-EutM or pUCBB-EutMNLK through BioBrick™ prefix stacking, as described for our group's customized set of BioBrick™ plasmids.[87, 157] Briefly, the pUCBB-EutS* plasmids were digested with EcoRI and SpeI. The resulting DNA fragement, containing the $P_{lacP'}$ promoter, ribosome binding site and EutS* ORF was then ligated into the pACBB destination plasmids that had been digested with EcoRI and XbaI. Because XbaI and SpeI are isoschizomers, the ligation results in neither an XbaI or EcoRI site being restored at ligation point. This resulted in the plasmids pUCBB-EutS*-mCherry-EutM, pUCBB-EutS*-EutM or pUCBB-EutS*MNLK. Finally, these plasmids were all digested with EcoRI and SpeI, and the resulting BioBricks containing multiple gene-encoding sequences were ligated into a pACBB backbone that had been digested with the same enzymes, giving pACBB-EutS*-mCherry-EutM, pACBB-EutS*-EutM or pACBB-EutS*MNLK. All plasmids used in this study are listed in Table 4.1.

4.3 Light Microscopy

To visualize cargo protein encapsulation within recombinant Eut BMCs, *E. coli* cells were cultured as described in section 4.1, diluted 10:1 in phosphate-buffered saline and then mounted onto glass coverslips. Cells were observed using a Nikon TiE microscope with a Lumencor SpectraX light source at 470nm through a Plan Apo 100x, 1.45n.a objective, and images were collected with an Andor Zyla camera controlled *via*

the Nikon Elements 4.6 software. All instruments are housed and maintained by the University of Minnesota – University Imaging Centers, http://uic.umn.edu. Post-capture alignment, false-coloring and cropping was conducted in ImageJ, Fiji and Nikon NIS Elements 4.6.

Table 4.1. Plasmids used in this study.

Plasmid	Description	Source
Standard Plasmid Backbones		
pUCBB	pMB1 (colE1); Amp^r $P_{lacP'}$	[157]
pACBB	p15A; Cm^r $P_{lacP'}$	[157]
pBBRBB	RK2; Kan^r $P_{lacP'}$	[157]
Enzyme-eGFP Fusion Plasmids		
pACBB-EutC$^{1-19}$-GS-MCS-GS-eGFP	pACBB:: $P_{lacP'}$ *eutC$^{1-19}$ tag-(GS)$_6$G-EcoRV-HindIII-Acc65I-KpnI-SalI-(GS)$_6$-eGFP*	E. Schmidt laboratory stock
pACBB-EutC$^{1-19}$-HdhA-eGFP	pACBB:: $P_{lacP'}$*eutC$^{1-19}$-(GS)$_6$G-hdhA-(GS)$_6$G-eGFP*	This study
pACBB-EutC$^{1-19}$-HpaB-eGFP	pACBB:: $P_{lacP'}$*eutC$^{1-19}$-(GS)$_6$G-hpaB-(GS)$_6$G-eGFP*	This study
EutS Plasmids		
pUCBB-EutS	pUCBB:: $P_{lacP'}$*eutS*	[87]
pUCBB-EutS$_{G26A}$	pUCBB:: $P_{lacP'}$*eutS$_{G26A}$*	This study
pUCBB-EutS$_{E27A}$	pUCBB:: $P_{lacP'}$*eutS$_{E27A}$*	This study
pUCBB-EutS$_{E28A}$	pUCBB:: $P_{lacP'}$*eutS$_{E28A}$*	This study

pUCBB-EutS$_{L29A}$	pUCBB:: P$_{lacP}$-*eutS$_{L29A}$*	This study
pUCBB-EutS$_{K31A}$	pUCBB:: P$_{lacP}$-*eutS$_{K31A}$*	This study
pUCBB-EutS$_{K32A}$	pUCBB:: P$_{lacP}$-*eutS$_{K32A}$*	This study
Encapsulation Study Plasmids		
pACBB-EutC$^{1-19}$-eGFP	pACBB:: P$_{lacP}$-*eutC$^{1-19}$-eGFP*	[87]
pBBRBB-EutC$^{1-19}$-eGFP	pBBRBB:: P$_{lacP}$-*eutC$^{1-19}$-eGFP*	[87]
pACBB- mCherry-EutM	pACBB:: P$_{lacP}$-*mCherry-eutM*	M.A. Held laboratory stock
pACBB- EutS-mCherry-EutM	pACBB:: P$_{lacP}$-*eutS*-P$_{lacP}$-*mCherry-eutM*	This study
pACBB- EutS$_{L29A}$-mCherry-EutM	pACBB:: P$_{lacP}$-*eutS$_{L29A}$*-P$_{lacP}$-*mCherry-eutM*	This study
pACBB- EutS$_{K32A}$-mCherry-EutM	pACBB:: P$_{lacP}$-*eutS$_{K32A}$*-P$_{lacP}$-*mCherry-eutM*	This study
pUCBB-EutM	pUCBB:: P$_{lacP}$-*eutM*	[87]
pACBB- EutSM	pACBB:: P$_{lacP}$-*eutS*-P$_{lacP}$-*eutM*	This study
pACBB- EutS$_{L29A}$M	pACBB:: P$_{lacP}$-*eutS$_{L29A}$*-P$_{lacP}$-*eutM*	This study
pUCBB-EutMNLK	pUCBB:: P$_{lacP}$-*eutMN*-P$_{lacP}$-*eutLK*	[87]
pACBB- EutSMNLK	pACBB:: P$_{lacP}$-*eutS*-P$_{lacP}$-*eutMN*-P$_{lacP}$-*eutLK*	M.A. Held laboratory stock
pACBB- EutS$_{L29A}$MNLK	pACBB:: P$_{lacP}$-*eutS*-P$_{lacP}$-*eutMN*-P$_{lacP}$-*eutLK*	This study

Chapter 5. Concluding Remarks

As the importance of plant natural products in pharmaceutical discovery has become more apparent, the need for microbial production platforms becomes more dire. The last two decades have seen the development of a number of valuable tools for the biosynthesis of medicinally-relevant plant compounds, including the flavonoid class of phenylpropanoid compounds. However, production of much of the range of plant phenylpropanoid classes, including many important pharmaceuticals, has yet to be reached with this technology. My thesis work aimed to expand the capabilities of microbial biosynthesis of medicinal plant phenylpropanoids by exploring and developing new biosynthetic pathways and metabolic engineering tools in *E. coli*. I worked toward engineering the biosynthesis of ligans and hydroxycinnamic acid esters, whose *de novo* biosynthesis in *E. coli* had not yet been achieved. I have also begun an investigation of the utility of engineered bacterial microcompartments (BMCs) as a means of controlling spatial organization in metabolic pathway engineering. This work has contributed significantly to our understanding of the capabilities and limitations of existing technology for microbial biosynthesis and has expanded the accessible range of phenylpropanoid products that can be produced in *E. coli*.

When we began exploring the possibility of engineering lignan biosynthesis in *E. coli*, very little was understood about the mechanism of regio- and stereo-controlled oxidative radical coupling. I was able to achieve random oxidative radical coupling of

coniferyl alcohol in *E. coli* through overexpression of a bacterial laccase, showing that this could be a feasible candidate for an oxidase in bacterial lignan production. Additionally, I was able to achieve the soluble expression of the characterized dirigent protein FiDRP1 in *E. coli*, which turned out to be a very difficult protein based on my experience and later work from other groups.[117] However, my difficulty in expressing FiDRP1 and my inability to show its function in *E. coli* proved to be a significant obstacle. Later publications by other groups showed the necessity of post-translational modifications in not only dirigent protein function but also folding and stability, illustrating the infeasibility of using *E. coli* as a host for lignan biosynthesis.[102] Despite these setbacks and the ultimate decision to retire this project, this work has helped provide important information for future attempts at lignan biosynthesis in microbes by suggesting that a eukaryotic host may be necessary.

In addition to lignan biosynthesis, I chose to pursue a platform for hydroxycinnamic acid ester (HCE) biosynthesis in *E. coli* with a focus on rosmarinic acid (RA), a potent antioxidant with potential for anti-cancer and neuroprotective applications.[23, 46] The beginning of this work coincided with a seemingly new interest in the *de novo* biosynthesis of hydroxycinnamic acids in *E. coli*.[47, 50] While previous groups had achieved production of hydroxycinnamoyl amides from fed precursors in *E. coli*,[52-53] at the beginning of this work, there were no reports of microbial biosynthesis of HCEs. Recognizing the potential difficulty of reconstituting the plant RA biosynthetic pathway in *E. coli*, I was able to use recent studies on bacterial caffeic acid biosynthesis

to restructure the RA pathway and create a novel biosynthetic route using bacterial and plant enzymes. This approach proved successful, and I was able to show titers of RA similar to those shown by Eudes et al. before pathway optimization, who had meanwhile published a study of the *de novo* biosynthesis of hydroxycinnamoyl anthranilates in *E. coli*.[54] Through probing the engineered pathway for bottlenecks, I was able to propose several approaches for future optimization of through boosting precursor availability and enzyme engineering. Additionally, due to my modular design of this biosynthetic pathway, this could prove to be a promising production platform not only for RA, but for a range of other HCEs and hydroxycinnamoyl amides.

Recent work in the Schmidt-Dannert laboratory has shown the potential utility of engineered ethanolamine utilization (Eut) BMCs in metabolic engineering applications.[87] I saw the potential to use this system to eliminate cross-talk within the redesigned RA biosynthetic pathway by sequestering a pathway intermediate, leading me to test the ability to localize two heterologous enzymes to Eut BMCs. Mixed results with the ability to localize enzymes to the BMCs using the N-terminal signal peptide $EutC^{1-19}$ led me to investigate the mechanism of cargo protein encapsulation within Eut shells. Previous work from other groups and our own observation suggested two potential models for enzyme encapsulation within BMCs: specific interactions between cognate pairs of shell and cargo proteins, or general interactions between several signal peptides and shell proteins. Through site-directed mutagenesis of the shell protein EutS, I was able to provide evidence that a specific interaction between EutS and $EutC^{1-19}$ mediates cargo

protein encapsulation within recombinant Eut BMCs. In addition, I have suggested ongoing work to expand our understanding of protein localization through the identification of other signal peptides and a thorough *in vitro* investigation of interactions between Eut shell and cargo proteins. This work has laid the foundation to deepen our understanding of Eut BMCs and increase their utility in metabolic engineering applications.

Taken together, the work presented in this thesis represents an exploration of the potential of microbial biosynthesis of plant phenylpropanoid natural products. Throughout the body of this thesis, I have presented multiple avenues for future directions for this research. For example, there is great potential for a microbial biosynthetic platform for lignans in a eukaryotic host such as *Pichia pastoris*. Optimization of RA biosynthesis in my system through increasing L-tyrosine biosynthesis, enzyme engineering and/or partial pathway colocalization within a BMC could lead to an economically viable microbial production platform. Finally, a study to better understand the protein-protein interactions that govern cargo protein encapsulation within BMCs will allow a much greater degree of control in metabolon engineering. In conclusion, my exploration of plant phenylpropanoid biosynthesis in *E. coli* has expanded our capabilities in microbial biosynthetic platforms and laid the foundation for further development of metabolic engineering tools.

Bibliography

[1] **J. A. Chemler, M. A. Koffas** (2008). "Metabolic engineering for plant natural product biosynthesis in microbes." Curr Opin Biotechnol, **19**, 597-605.

[2] **M. Mora-Pale, S. P. Sanchez-Rodriguez, R. J. Linhardt, J. S. Dordick, M. A. G. Koffas** (2013). "Metabolic engineering and in vitro biosynthesis of phytochemicals and non-natural analogues." Plant Sci, **210**, 10-24.

[3] **M. Feher, J. M. Schmidt** (2003). "Property distributions: Differences between drugs, natural products, and molecules from combinatorial chemistry." J Chem Inf Comp Sci, **43**, 218-227.

[4] **R. H. Baltz** (2006). "Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration?" J Ind Microbiol Biotechnol, **33**, 507-513.

[5] **D. D. Baker, M. Chu, U. Oza, V. Rajgarhia** (2007). "The value of natural products to future pharmaceutical discovery." Nat Prod Rep, **24**, 1225-1244.

[6] **D. J. Newman, G. M. Cragg** (2007). "Natural products as sources of new drugs over the last 25 years." J Nat Prod, **70**, 461-477.

[7] **M. A. Hicks, K. L. Prather** (2014). "Bioprospecting in the genomic age." Adv Appl Microbiol, **87**, 111-146.

[8] **C. J. Paddon, P. J. Westfall, D. J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M. D. Leavell, A. Tai, A. Main, D. Eng, D. R. Polichuk, K. H. Teoh, D. W. Reed, T. Treynor, J. Lenihan, M. Fleck, S. Bajad, G. Dang, D. Dengrove, D. Diola, G. Dorin, K. W. Ellens, S. Fickes, J. Galazzo, S. P. Gaucher, T. Geistlinger, R. Henry, M. Hepp, T. Horning, T. Iqbal, H. Jiang, L. Kizer, B. Lieu, D. Melis, N. Moss, R. Regentin, S. Secrest, H. Tsuruta, R. Vazquez, L. F. Westblade, L. Xu, M. Yu, Y. Zhang, L. Zhao, J. Lievense, P. S. Covello, J. D. Keasling, K. K. Reiling, N. S. Renninger, J. D. Newman** (2013). "High-level semi-synthetic production of the potent antimalarial artemisinin." Nature, **496**, 528-+.

[9] **P. K. Ajikumar, W. H. Xiao, K. E. J. Tyo, Y. Wang, F. Simeon, E. Leonard, O. Mucha, T. H. Phon, B. Pfeifer, G. Stephanopoulos** (2010). "Isoprenoid

Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli." Science, **330**, 70-74.

[10] **M. C. Song, E. J. Kim, E. Kim, K. Rathwell, S. J. Nam, Y. J. Yoon** (2014). "Microbial biosynthesis of medicinally important plant secondary metabolites." Nat Prod Rep.

[11] **K. T. Watts, P. C. Lee, C. Schmidt-Dannert** (2004). "Exploring recombinant flavonoid biosynthesis in metabolically engineered Escherichia coli." Chembiochem, **5**, 500-507.

[12] **K. T. Watts, P. C. Lee, C. Schmidt-Dannert** (2006). "Biosynthesis of plant-specific stilbene polyketides in metabolically engineered Escherichia coli." BMC Biotechnol, **6**.

[13] **T. Vogt** (2010). "Phenylpropanoid Biosynthesis." Mol Plant, **3**, 2-20.

[14] **K. M. Herrmann, L. M. Weaver** (1999). "The shikimate pathway." Annu Rev Plant Physiol Plant Mol Biol, **50**, 473-503.

[15] **H. Maeda, N. Dudareva** (2012). "The Shikimate Pathway and Aromatic Amino Acid Biosynthesis in Plants." Annu Rev Plant Biol, **63**, 73-105.

[16] **G. Gosset** (2009). "Production of aromatic compounds in bacteria." Curr Opin Biotechnol, **20**, 651-658.

[17] **M. I. Chavez-Bejar, A. R. Lara, H. Lopez, G. Hernandez-Chavez, A. Martinez, O. T. Ramirez, F. Bolivar, G. Gosset** (2008). "Metabolic engineering of Escherichia coli for L-tyrosine production by expression of genes coding for the chorismate mutase domain of the native chorismate mutase-prephenate dehydratase and a cyclohexadienyl dehydrogenase from Zymomonas mobilis." Appl Environ Microbiol, **74**, 3284-3290.

[18] **M. Berner, D. Krug, C. Bihlmaier, A. Vente, R. Muller, A. Bechthold** (2006). "Genes and enzymes involved in caffeic acid biosynthesis in the actinomycete Saccharothrix espanaensis." J Bacteriol, **188**, 2666-2673.

[19] **J. S. Williams, M. Thomas, D. J. Clarke** (2005). "The gene stlA encodes a phenylalanine ammonia-lyase that is involved in the production of a stilbene antibiotic in Photorhabdus luminescens TT01." Microbiol-Sgm, **151**, 2543-2550.

[20] **L. K. Xiang, B. S. Moore** (2002). "Inactivation, complementation, and heterologous expression of encP, a novel bacterial phenylalanine ammonia-lyase gene." J Biol Chem, **277**, 32505-32509.

[21] **M. Meydani** (2009). "Potential health benefits of avenanthramides of oats." Nutr Rev, **67**, 731-735.

[22] **R. Upadhyay, L. J. M. Rao** (2013). "An Outlook on Chlorogenic Acids-Occurrence, Chemistry, Technology, and Biological Activities." Crit Rev Food Sci, **53**, 968-984.

[23] **V. P. Bulgakov, Y. V. Inyushkina, S. A. Fedoreyev** (2012). "Rosmarinic acid and its derivatives: biotechnology and applications." Crit Rev Biotechnol, **32**, 203-217.

[24] **B. Winkel-Shirley** (2001). "Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology." Plant Physiol, **126**, 485-493.

[25] **R. A. Dixon, F. Chen, D. J. Guo, K. Parvathi** (2001). "The biosynthesis of monolignols: a "metabolic grid", or independent pathways to guaiacyl and syringyl units?" Phytochemistry, **57**, 1069-1084.

[26] **R. Vanholme, B. Demedts, K. Morreel, J. Ralph, W. Boerjan** (2010). "Lignin Biosynthesis and Structure." Plant Physiol, **153**, 895-905.

[27] **L. B. Davin, N. G. Lewis** (2005). "Dirigent phenoxy radical coupling: advances and challenges." Curr Opin Biotechnol, **16**, 398-406.

[28] **R. Dexter, A. Qualley, C. M. Kish, C. J. Ma, T. Koeduka, D. A. Nagegowda, N. Dudareva, E. Pichersky, D. Clark** (2007). "Characterization of a petunia acetyltransferase involved in the biosynthesis of the floral volatile isoeugenol." Plant J, **49**, 265-275.

[29] **T. Koeduka, E. Fridman, D. R. Gang, D. G. Vassao, B. L. Jackson, C. M. Kish, I. Orlova, S. M. Spassova, N. G. Lewis, J. P. Noel, T. J. Baiga, N. Dudareva, E. Pichersky** (2006). "Eugenol and isoeugenol, characteristic aromatic constituents of spices, are biosynthesized via reduction of a coniferyl alcohol ester." Proc Natl Acad Sci U S A, **103**, 10128-10133.

[30] **M. B. Quin, C. Schmidt-Dannert** (2014). "Designer microbes for biosynthesis." Curr Opin Biotechnol, **29C**, 55-61.

[31] **K. Buchholz, J. Collins** (2013). "The roots--a short history of industrial microbiology and biotechnology." Appl Microbiol Biotechnol, **97**, 3747-3762.

[32] **Y. Wang, S. Chen, O. Yu** (2011). "Metabolic engineering of flavonoids in plants and microorganisms." Appl Microbiol Biotechnol, **91**, 949-956.

[33] **V. J. Martin, D. J. Pitera, S. T. Withers, J. D. Newman, J. D. Keasling** (2003). "Engineering a mevalonate pathway in Escherichia coli for production of terpenoids." Nat Biotechnol, **21**, 796-802.

[34] **H. Renault, J. E. Bassard, B. Hamberger, D. Werck-Reichhart** (2014). "Cytochrome P450-mediated metabolic engineering: current progress and future challenges." Curr Opin Plant Biol, **19C**, 27-34.

[35] **H. C. Tseng, K. L. J. Prather** (2012). "Controlled biosynthesis of odd-chain fuels and chemicals via engineered modular metabolic pathways." Proc Natl Acad Sci U S A, **109**, 17925-17930.

[36] **A. Bar-Even, D. Salah Tawfik** (2013). "Engineering specialized metabolic pathways--is there a room for enzyme improvements?" Curr Opin Biotechnol, **24**, 310-319.

[37] **R. E. Cobb, J. C. Ning, H. Zhao** (2014). "DNA assembly techniques for next-generation combinatorial biosynthesis of natural products." J Ind Microbiol Biotechnol, **41**, 469-477.

[38] **B. W. Biggs, B. De Paepe, C. N. Santos, M. De Mey, P. Kumaran Ajikumar** (2014). "Multivariate modular metabolic engineering for pathway and strain optimization." Curr Opin Biotechnol, **29C**, 156-162.

[39] **R. J. Conrado, J. D. Varner, M. P. DeLisa** (2008). "Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy." Curr Opin Biotechnol, **19**, 492-499.

[40] **E. I. Hwang, M. Kaneko, Y. Ohnishi, S. Horinouchi** (2003). "Production of plant-specific flavanones by Escherichia coli containing an artificial gene cluster." Appl Environ Microbiol, **69**, 2699-2706.

[41] **J. V. Becker, G. O. Armstrong, M. J. van der Merwe, M. G. Lambrechts, M. A. Vivier, I. S. Pretorius** (2003). "Metabolic engineering of Saccharomyces cerevisiae for the synthesis of the wine-related antioxidant resveratrol." FEMS Yeast Res, **4**, 79-85.

[42] **E. Leonard, K. H. Lim, P. N. Saw, M. A. Koffas** (2007). "Engineering central metabolic pathways for high-level flavonoid production in Escherichia coli." Appl Environ Microbiol, **73**, 3877-3886.

[43] **W. Zha, S. B. Rubin-Pitel, Z. Shao, H. Zhao** (2009). "Improving cellular malonyl-CoA level in Escherichia coli via metabolic engineering." Metab Eng, **11**, 192-198.

[44] **F. Koopman, J. Beekwilder, B. Crimi, A. van Houwelingen, R. D. Hall, D. Bosch, A. J. van Maris, J. T. Pronk, J. M. Daran** (2012). "De novo production of the flavonoid naringenin in engineered Saccharomyces cerevisiae." Microb Cell Fact, **11**, 155.

[45] **C. N. Santos, M. Koffas, G. Stephanopoulos** (2011). "Optimization of a heterologous pathway for the production of flavonoids from glucose." Metab Eng, **13**, 392-400.

[46] **M. Petersen, M. S. Simmonds** (2003). "Rosmarinic acid." Phytochemistry, **62**, 121-125.

[47] **O. Choi, C. Z. Wu, S. Y. Kang, J. S. Ahn, T. B. Uhm, Y. S. Hong** (2011). "Biosynthesis of plant-specific phenylpropanoids by construction of an artificial biosynthetic pathway in Escherichia coli." J Ind Microbiol Biotechnol, **38**, 1657-1665.

[48] **H. Zhang, G. Stephanopoulos** (2013). "Engineering E. coli for caffeic acid biosynthesis from renewable sugars." Appl Microbiol Biotechnol, **97**, 3333-3341.

[49] **S. Y. Kang, O. Choi, J. K. Lee, B. Y. Hwang, T. B. Uhm, Y. S. Hong** (2012). "Artificial biosynthesis of phenylpropanoic acids in a tyrosine overproducing Escherichia coli strain." Microb Cell Fact, **11**, 153.

[50] **Y. Lin, Y. Yan** (2012). "Biosynthesis of caffeic acid in Escherichia coli using its endogenous hydroxylase complex." Microb Cell Fact, **11**, 42.

[51] **B. G. Kim, W. D. Jung, H. Mok, J. H. Ahn** (2013). "Production of hydroxycinnamoyl-shikimates and chlorogenic acid in Escherichia coli: production of hydroxycinnamic acid conjugates." Microb Cell Fact, **12**, 15.

[52] **K. Kang, K. Back** (2009). "Production of phenylpropanoid amides in recombinant Escherichia coli." Metab Eng, **11**, 64-68.

[53] **K. Kang, M. Park, S. Park, Y. S. Kim, S. Lee, S. G. Lee, K. Back** (2009). "Production of plant-specific tyramine derivatives by dual expression of tyramine N-hydroxycinnamoyltransferase and 4-coumarate:coenzyme A ligase in Escherichia coli." Biotechnol Lett, **31**, 1469-1475.

[54] **A. Eudes, D. Juminaga, E. E. Baidoo, F. W. Collins, J. D. Keasling, D. Loque** (2013). "Production of hydroxycinnamoyl anthranilates from glucose in Escherichia coli." Microb Cell Fact, **12**, 62.

[55] **M. N. Cha, H. J. Kim, B. G. Kim, J. H. Ahn** (2014). "Synthesis of Chlorogenic Acid and p-Coumaroyl Shikimates from Glucose Using Engineered Escherichia coli." J Microbiol Biotechnol, **24**, 1109-1117.

[56] **Y. Lin, X. Shen, Q. Yuan, Y. Yan** (2013). "Microbial biosynthesis of the anticoagulant precursor 4-hydroxycoumarin." Nat Commun, **4**, 2603.

[57] **Y. Lin, X. Sun, Q. Yuan, Y. Yan** (2013). "Combinatorial biosynthesis of plant-specific coumarins in bacteria." Metab Eng, **18**, 69-77.

[58] **S. J. Kim, D. G. Vassao, S. G. Moinuddin, D. L. Bedgar, L. B. Davin, N. G. Lewis** (2014). "Allyl/propenyl phenol synthases from the creosote bush and

engineering production of specialty/commodity chemicals, eugenol/isoeugenol, in Escherichia coli." Arch Biochem Biophys, **541**, 37-46.

[59] **H. J. Kuo, Z. Y. Wei, P. C. Lu, P. L. Huang, K. T. Lee** (2014). "Bioconversion of pinoresinol into matairesinol by use of recombinant Escherichia coli." Appl Environ Microbiol, **80**, 2687-2692.

[60] **T. Lutke-Eversloh, G. Stephanopoulos** (2007). "L-tyrosine production by deregulated strains of Escherichia coli." Appl Microbiol Biotechnol, **75**, 103-110.

[61] **A. J. Munoz, G. Hernandez-Chavez, R. de Anda, A. Martinez, F. Bolivar, G. Gosset** (2011). "Metabolic engineering of Escherichia coli for improving L-3,4-dihydroxyphenylalanine (L-DOPA) synthesis from glucose." J Ind Microbiol Biotechnol, **38**, 1845-1852.

[62] **N. Yakandawala, T. Romeo, A. D. Friesen, S. Madhyastha** (2008). "Metabolic engineering of Escherichia coli to enhance phenylalanine production." Appl Microbiol Biotechnol, **78**, 283-291.

[63] **A. Escalante, R. Calderon, A. Valdivia, R. de Anda, G. Hernandez, O. T. Ramirez, G. Gosset, F. Bolivar** (2010). "Metabolic engineering for the production of shikimic acid in an evolved Escherichia coli strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system." Microb Cell Fact, **9**, 21.

[64] **N. Ran, J. W. Frost** (2007). "Directed evolution of 2-keto-3-deoxy-6-phosphogalactonate aldolase to replace 3-deoxy-D-arabino-heptulosonic acid 7-phosphate synthase." J Am Chem Soc, **129**, 6130-6139.

[65] **T. Lutke-Eversloh, G. Stephanopoulos** (2008). "Combinatorial pathway analysis for improved L-tyrosine production in Escherichia coli: identification of enzymatic bottlenecks by systematic gene overexpression." Metab Eng, **10**, 69-77.

[66] **T. Lutke-Eversloh, G. Stephanopoulos** (2007). "A semi-quantitative high-throughput screening method for microbial L-tyrosine production in microtiter plates." J Ind Microbiol Biotechnol, **34**, 807-811.

[67] **C. N. Santos, W. Xiao, G. Stephanopoulos** (2012). "Rational, combinatorial, and genomic approaches for engineering L-tyrosine production in Escherichia coli." Proc Natl Acad Sci U S A, **109**, 13538-13543.

[68] **R. J. Ellis** (2001). "Macromolecular crowding: obvious but underappreciated." Trends Biochem Sci, **26**, 597-604.

[69] **M. Arrio-Dupont, G. Foucault, M. Vacher, P. F. Devaux, S. Cribier** (2000). "Translational diffusion of globular proteins in the cytoplasm of cultured muscle cells." Biophys J, **78**, 901-907.

[70] **C. C. Hyde, S. A. Ahmed, E. A. Padlan, E. W. Miles, D. R. Davies** (1988). "Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from Salmonella typhimurium." J Biol Chem, **263**, 17857-17871.

[71] **S. Cheng, Y. Liu, C. S. Crowley, T. O. Yeates, T. A. Bobik** (2008). "Bacterial microcompartments: their properties and paradoxes." Bioessays, **30**, 1084-1095.

[72] **S. An, R. Kumar, E. D. Sheets, S. J. Benkovic** (2008). "Reversible compartmentalization of de novo purine biosynthetic complexes in living cells." Science, **320**, 103-106.

[73] **B. Winkel-Shirley** (1999). "Evidence for enzyme complexes in the phenylpropanoid and flavonoid pathways." Physiol Plantarum, **107**, 142-149.

[74] **I. Meynial Salles, N. Forchhammer, C. Croux, L. Girbal, P. Soucaille** (2007). "Evolution of a Saccharomyces cerevisiae metabolic pathway in Escherichia coli." Metab Eng, **9**, 152-159.

[75] **H. C. Chang, C. M. Kaiser, F. U. Hartl, J. M. Barral** (2005). "De novo folding of GFP fusion proteins: high efficiency in eukaryotes but not in bacteria." J Mol Biol, **353**, 397-409.

[76] **H. P. Fierobe, A. Mechaly, C. Tardif, A. Belaich, R. Lamed, Y. Shoham, J. P. Belaich, E. A. Bayer** (2001). "Design and production of active cellulosome chimeras. Selective incorporation of dockerin-containing enzymes into defined functional complexes." J Biol Chem, **276**, 21257-21261.

[77] **J. E. Dueber, G. C. Wu, G. R. Malmirchegini, T. S. Moon, C. J. Petzold, A. V. Ullal, K. L. Prather, J. D. Keasling** (2009). "Synthetic protein scaffolds provide modular control over metabolic flux." Nat Biotechnol, **27**, 753-759.

[78] **D. O. Moon, M. O. Kim, J. D. Lee, Y. H. Choi, G. Y. Kim** (2010). "Rosmarinic acid sensitizes cell death through suppression of TNF-alpha-induced NF-kappaB activation and ROS generation in human leukemia U937 cells." Cancer Lett, **288**, 183-191.

[79] **Y. Wang, O. Yu** (2012). "Synthetic scaffolds increased resveratrol biosynthesis in engineered yeast cells." J Biotechnol, **157**, 258-260.

[80] **R. J. Conrado, G. C. Wu, J. T. Boock, H. Xu, S. Y. Chen, T. Lebar, J. Turnsek, N. Tomsic, M. Avbelj, R. Gaber, T. Koprivnjak, J. Mori, V. Glavnik, I. Vovk, M. Bencina, V. Hodnik, G. Anderluh, J. E. Dueber, R. Jerala, M. P. DeLisa** (2012). "DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency." Nucleic Acids Res, **40**, 1879-1889.

[81] **C. J. Delebecque, A. B. Lindner, P. A. Silver, F. A. Aldaye** (2011). "Organization of intracellular reactions with rationally designed RNA assemblies." Science, **333**, 470-474.

[82] **M. Sutter, D. Boehringer, S. Gutmann, S. Gunther, D. Prangishvili, M. J. Loessner, K. O. Stetter, E. Weber-Ban, N. Ban** (2008). "Structural basis of enzyme encapsulation into a bacterial nanocompartment." Nat Struct Mol Biol, **15**, 939-947.

[83] **F. P. Seebeck, K. J. Woycechowsky, W. Zhuang, J. P. Rabe, D. Hilvert** (2006). "A simple tagging system for protein encapsulation." J Am Chem Soc, **128**, 4516-4517.

[84] **B. Worsdorfer, K. J. Woycechowsky, D. Hilvert** (2011). "Directed evolution of a protein container." Science, **331**, 589-592.

[85] **J. B. Parsons, S. Frank, D. Bhella, M. Liang, M. B. Prentice, D. P. Mulvihill, M. J. Warren** (2010). "Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement." Mol Cell, **38**, 305-315.

[86] **W. Bonacci, P. K. Teng, B. Afonso, H. Niederholtmeyer, P. Grob, P. A. Silver, D. F. Savage** (2012). "Modularity of a carbon-fixing protein organelle." Proc Natl Acad Sci U S A, **109**, 478-483.

[87] **S. Choudhary, M. B. Quin, M. A. Sanders, E. T. Johnson, C. Schmidt-Dannert** (2012). "Engineered protein nano-compartments for targeted enzyme localization." PLoS One, **7**, e33342.

[88] **C. Fan, T. A. Bobik** (2011). "The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment." J Bacteriol, **193**, 5623-5628.

[89] **C. Fan, S. Cheng, Y. Liu, C. M. Escobar, C. S. Crowley, R. E. Jefferson, T. O. Yeates, T. A. Bobik** (2010). "Short N-terminal sequences package proteins into bacterial microcompartments." Proc Natl Acad Sci U S A, **107**, 7509-7514.

[90] **S. J. Tsai, T. O. Yeates** (2011). "Bacterial microcompartments insights into the structure, mechanism, and engineering applications." Prog Mol Biol Transl Sci, **103**, 1-20.

[91] **F. Cai, M. Sutter, S. L. Bernstein, J. N. Kinney, C. A. Kerfeld** (2014). "Engineering Bacterial Microcompartment Shells: Chimeric Shell Proteins and Chimeric Carboxysome Shells." ACS Synth Biol.

[92] **C. Canel, R. M. Moraes, F. E. Dayan, D. Ferreira** (2000). "Podophyllotoxin." Phytochemistry, **54**, 115-120.

[93] **M. Gordaliza, P. A. Garcia, J. M. del Corral, M. A. Castro, M. A. Gomez-Zurita** (2004). "Podophyllotoxin: distribution, sources, applications and new cytotoxic derivatives." Toxicon, **44**, 441-459.

[94] **N. Braidy, A. Matin, F. Rossi, M. Chinain, D. Laurent, G. J. Guillemin** (2014). "Neuroprotective effects of rosmarinic acid on ciguatoxin in primary human neurons." Neurotox Res, **25**, 226-234.

[95] **V. D. Kancheva, O. T. Kasaikina** (2013). "Bio-antioxidants - a chemical base of their antioxidant activity and beneficial effect on human health." Curr Med Chem, **20**, 4784-4805.

[96] **C. C. Tangney, H. E. Rasmussen** (2013). "Polyphenols, inflammation, and cardiovascular disease." Curr Atheroscler Rep, **15**, 324.

[97] **V. B. Gencel, M. M. Benjamin, S. N. Bahou, R. A. Khalil** (2012). "Vascular effects of phytoestrogens and alternative menopausal hormone therapy in cardiovascular disease." Mini Rev Med Chem, **12**, 149-174.

[98] **M. Khaled, Z. Z. Jiang, L. Y. Zhang** (2013). "Deoxypodophyllotoxin: a promising therapeutic agent from herbal medicine." J Ethnopharmacol, **149**, 24-34.

[99] **S. Farkya, V. S. Bisaria, A. K. Srivastava** (2004). "Biotechnological aspects of the production of the anticancer drug podophyllotoxin." Appl Microbiol Biotechnol, **65**, 504-519.

[100] **L. B. Davin, H. B. Wang, A. L. Crowell, D. L. Bedgar, D. M. Martin, S. Sarkanen, N. G. Lewis** (1997). "Stereoselective bimolecular phenoxy radical coupling by an auxiliary (dirigent) protein without an active center." Science, **275**, 362-366.

[101] **S. C. Halls, L. B. Davin, D. M. Kramer, N. G. Lewis** (2004). "Kinetic study of coniferyl alcohol radical binding to the (+)-pinoresinol forming dirigent protein." Biochemistry, **43**, 2587-2595.

[102] **B. Pickel, A. Schaller** (2013). "Dirigent proteins: molecular characteristics and potential biotechnological applications." Appl Microbiol Biotechnol, **97**, 8427-8438.

[103] **A. Piscitelli, C. Pezzella, P. Giardina, V. Faraco, S. Giovanni** (2010). "Heterologous laccase production and its role in industrial applications." Bioeng Bugs, **1**, 252-262.

[104] **B. Viswanath, B. Rajesh, A. Janardhan, A. P. Kumar, G. Narasimha** (2014). "Fungal laccases and their applications in bioremediation." Enzyme Res, **2014**, 163242.

[105] **S. Ba, A. Arsenault, T. Hassani, J. P. Jones, H. Cabana** (2013). "Laccase immobilization and insolubilization: from fundamentals to applications for the elimination of emerging contaminants in wastewater treatment." Crit Rev Biotechnol, **33**, 404-418.

[106] **J. R. Jeon, P. Baldrian, K. Murugesan, Y. S. Chang** (2012). "Laccase-catalysed oxidations of naturally occurring phenols: from in vivo biosynthetic pathways to green synthetic applications." Microb Biotechnol, **5**, 318-332.

[107] **P. Giardina, V. Faraco, C. Pezzella, A. Piscitelli, S. Vanhulle, G. Sannia** (2010). "Laccases: a never-ending story." Cell Mol Life Sci, **67**, 369-385.

[108] **M. C. Machczynski, E. Vijgenboom, B. Samyn, G. W. Canters** (2004). "Characterization of SLAC: a small laccase from Streptomyces coelicolor with unprecedented activity." Protein Sci, **13**, 2388-2397.

[109] **S. Ralph, J. Y. Park, J. Bohlmann, S. D. Mansfield** (2006). "Dirigent proteins in conifer defense: gene discovery, phylogeny, and differential wound- and insect-induced expression of a family of DIR and DIR-like genes in spruce (Picea spp.)." Plant Mol Biol, **60**, 21-40.

[110] **S. G. Ralph, S. Jancsik, J. Bohlmann** (2007). "Dirigent proteins in conifer defense II: Extended gene discovery, phylogeny, and constitutive and stress-induced gene expression in spruce (Picea spp.)." Phytochemistry, **68**, 1975-1991.

[111] **K. W. Kim, S. G. Moinuddin, K. M. Atwell, M. A. Costa, L. B. Davin, N. G. Lewis** (2012). "Opposite stereoselectivities of dirigent proteins in Arabidopsis and schizandra species." J Biol Chem, **287**, 33957-33972.

[112] **J. Liu, R. D. Stipanovic, A. A. Bell, L. S. Puckhaber, C. W. Magill** (2008). "Stereoselective coupling of hemigossypol to form (+)-gossypol in moco cotton is mediated by a dirigent protein." Phytochemistry, **69**, 3038-3042.

[113] **B. Pickel, M. A. Constantin, J. Pfannstiel, J. Conrad, U. Beifuss, A. Schaller** (2010). "An enantiocomplementary dirigent protein for the enantioselective laccase-catalyzed oxidative coupling of phenols." Angew Chem Int Ed Engl, **49**, 202-204.

[114] **A. Schaller, A. Stintzi** (2009). "Enzymes in jasmonate biosynthesis - structure, function, regulation." Phytochemistry, **70**, 1532-1538.

[115] **E. Hofmann, P. Zerbe, F. Schaller** (2006). "The crystal structure of Arabidopsis thaliana allene oxide cyclase: insights into the oxylipin cyclization reaction." Plant Cell, **18**, 3201-3217.

[116] **S. C. Halls, N. G. Lewis** (2002). "Secondary and quaternary structures of the (+)-pinoresinol-forming dirigent protein." Biochemistry, **41**, 9455-9461.

[117] **C. Kazenwadel, J. Klebensberger, S. Richter, J. Pfannstiel, U. Gerken, B. Pickel, A. Schaller, B. Hauer** (2013). "Optimized expression of the dirigent protein AtDIR6 in Pichia pastoris and impact of glycosylation on protein structure and function." Appl Microbiol Biotechnol, **97**, 7215-7227.

[118] **B. Pickel, J. Pfannstiel, A. Steudle, A. Lehmann, U. Gerken, J. Pleiss, A. Schaller** (2012). "A model of dirigent proteins derived from structural and functional similarities with allene oxide cyclase and lipocalins." FEBS J, **279**, 1980-1993.

[119] **A. Khlebnikov, K. A. Datsenko, T. Skaug, B. L. Wanner, J. D. Keasling** (2001). "Homogeneous expression of the P(BAD) promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter." Microbiology, **147**, 3241-3247.

[120] **C. Schmidt-Dannert, D. Umeno, F. H. Arnold** (2000). "Molecular breeding of carotenoid biosynthetic pathways." Nature Biotechnology, **18**, 750-753.

[121] **S. E. Bloch, C. Schmidt-Dannert** (2014). "Construction of a Chimeric Biosynthetic Pathway for the De Novo Biosynthesis of Rosmarinic Acid in *Escherichia coli*." Chembiochem, **[Epub ahead of print.]**.

[122] **J. Paluszczak, V. Krajka-Kuzniak, W. Baer-Dubowska** (2010). "The effect of dietary polyphenols on the epigenetic regulation of gene expression in MCF7 breast cancer cells." Toxicol Lett, **192**, 119-125.

[123] **Y. Xu, Z. Jiang, G. Ji, J. Liu** (2010). "Inhibition of bone metastasis from breast carcinoma by rosmarinic acid." Planta Med, **76**, 956-962.

[124] **C. Anusuya, S. Manoharan** (2011). "Antitumor initiating potential of rosmarinic acid in 7,12-dimethylbenz(a)anthracene-induced hamster buccal pouch carcinogenesis." J Environ Pathol Toxicol Oncol, **30**, 199-211.

[125] **K. Venkatachalam, S. Gunasekaran, V. A. Jesudoss, N. Namasivayam** (2013). "The effect of rosmarinic acid on 1,2-dimethylhydrazine induced colon carcinogenesis." Exp Toxicol Pathol, **65**, 409-418.

[126] **S. Fallarini, G. Miglio, T. Paoletti, A. Minassi, A. Amoruso, C. Bardelli, S. Brunelleschi, G. Lombardi** (2009). "Clovamide and rosmarinic acid induce neuroprotective effects in in vitro models of neuronal death." Br J Pharmacol, **157**, 1072-1084.

[127] **N. A. Kelsey, H. M. Wilkins, D. A. Linseman** (2010). "Nutraceutical antioxidants as novel neuroprotective agents." Molecules, **15**, 7792-7814.

[128] **Y. Shimojo, K. Kosaka, Y. Noda, T. Shimizu, T. Shirasawa** (2010). "Effect of rosmarinic acid in motor dysfunction and life span in a mouse model of familial amyotrophic lateral sclerosis." J Neurosci Res, **88**, 896-904.

[129] **D. H. Park, S. J. Park, J. M. Kim, W. Y. Jung, J. H. Ryu** (2010). "Subchronic administration of rosmarinic acid, a natural prolyl oligopeptidase inhibitor, enhances cognitive performances." Fitoterapia, **81**, 644-648.

[130] **T. Hamaguchi, K. Ono, A. Murase, M. Yamada** (2009). "Phenolic compounds prevent Alzheimer's pathology through different effects on the amyloid-beta aggregation pathway." Am J Pathol, **175**, 2557-2565.

[131] **J. Wang, H. Xu, H. Jiang, X. Du, P. Sun, J. Xie** (2012). "Neurorescue effect of rosmarinic acid on 6-hydroxydopamine-lesioned nigral dopamine neurons in rat model of Parkinson's disease." J Mol Neurosci, **47**, 113-119.

[132] **X. Jin, P. Liu, F. Yang, Y. H. Zhang, D. Miao** (2013). "Rosmarinic acid ameliorates depressive-like behaviors in a rat model of CUS and Up-regulates BDNF levels in the hippocampus and hippocampal-derived astrocytes." Neurochem Res, **38**, 1828-1837.

[133] **H. Nie, Z. Peng, N. Lao, H. Wang, Y. Chen, Z. Fang, W. Hou, F. Gao, X. Li, L. Xiong, Q. Tan** (2014). "Rosmarinic acid ameliorates PTSD-like symptoms in a rat model and promotes cell proliferation in the hippocampus." Prog Neuropsychopharmacol Biol Psychiatry, **51**, 16-22.

[134] **H. D. Zinsmeister, H. Becker, T. Eicher** (1991). "Bryophytes, a Source of Biologically-Active, Naturally-Occurring Material." Angew Chem Int Ed Engl, **30**, 130-147.

[135] **T. Eicher, M. Ott, A. Speicher** (1996). "Bryophyte constituents .7. New synthesis of (+)-rosmarinic acid and related compounds." Synthesis-Stuttgart, 755-&.

[136] **K. Pabsch, M. Petersen, N. N. Rao, A. W. Alfermann, C. Wandrey** (1991). "Chemoenzymatic Synthesis of Rosmarinic Acid." Recl Trav Chim Pay B, **110**, 199-205.

[137] **E. Reimann, H. J. Maas, T. Pflug** (1997). "Synthesis of (+/-)-rosmarinic acid methylester." Monatsh Chem, **128**, 995-1008.

[138] **L. J. Huang, C. H. Li, Z. M. Lu, Z. B. Ma, D. Q. Yu** (2006). "Total synthesis and biological evaluation of (+)- and (-)-butyl ester of rosmarinic acid." J Asian Nat Prod Res, **8**, 561-566.

[139] **M. S. Petersen** (1991). "Characterization of Rosmarinic Acid Synthase from Cell-Cultures of Coleus-Blumei." Phytochemistry, **30**, 2877-2881.

[140] **W. W. Su, F. Lei, L. Y. Su** (1993). "Perfusion Strategy for Rosmarinic Acid Production by Anchusa-Officinalis." Biotechnol Bioeng, **42**, 884-890.

[141] **M. Georgiev, A. Pavlov, M. Ilieva** (2006). "Selection of high rosmarinic acid producing Lavandula vera MM cell lines." Process Biochem, **41**, 2068-2071.

[142] **B. Francoise, S. Hossein, H. Halimeh, N. F. Zahra** (2007). "Growth optimization of Zataria multiflora Boiss. tissue cultures and rosmarinic acid production improvement." Pak J Biol Sci, **10**, 3395-3399.

[143] **S. Y. Lee, H. Xu, Y. K. Kim, S. U. Park** (2008). "Rosmarinic acid production in hairy root cultures of Agastache rugosa Kuntze." World J Microb Biot, **24**, 969-972.

[144] **M. Georgiev, R. Abrashev, E. Krumova, K. Demirevska, M. Ilieva, M. Angelova** (2009). "Rosmarinic acid and antioxidant enzyme activities in

Lavandula vera MM cell suspension culture: a comparative study." Appl Biochem Biotechnol, **159**, 415-425.

[145] **D. Eberle, P. Ullmann, D. Werck-Reichhart, M. Petersen** (2009). "cDNA cloning and functional characterisation of CYP98A14 and NADPH:cytochrome P450 reductase from Coleus blumei involved in rosmarinic acid biosynthesis." Plant Mol Biol, **69**, 239-253.

[146] **J. Negrel, F. Javelle** (1997). "Purification, characterization and partial amino acid sequencing of hydroxycinnamoyl-CoA:tyramine N-(hydroxycinnamoyl)transferase from tobacco cell-suspension cultures." Eur J Biochem, **247**, 1127-1135.

[147] **Q. Yang, K. Reinhard, E. Schiltz, U. Matern** (1997). "Characterization and heterologous expression of hydroxycinnamoyl/benzoyl-CoA : anthranilate N-hydroxycinnamoyl/benzoyltransferase from elicited cell cultures of carnation, Dianthus caryophyllus L." Plant Mol Biol, **35**, 777-789.

[148] **M. Yu, P. J. Facchini** (1999). "Purification, characterization, and immunolocalization of hydroxycinnamoyl-CoA: tyramine N-(hydroxycinnamoyl)transferase from opium poppy." Planta, **209**, 33-44.

[149] **K. Yonekura-Sakakibara, Y. Tanaka, M. Fukuchi-Mizutani, H. Fujiwara, Y. Fukui, T. Ashikari, Y. Murakami, M. Yamaguchi, T. Kusumi** (2000). "Molecular and biochemical characterization of a novel hydroxycinnamoyl-CoA: Anthocyanin 3-O-glucoside-6 ''-O-acyltransferase from Perilla frutescens." Plant Cell Physiol, **41**, 495-502.

[150] **K. Back, S. M. Jang, B. C. Lee, A. Schmidt, D. Strack, K. M. Kim** (2001). "Cloning and characterization of a hydroxycinnamoyl-CoA:tyramine N-(hydroxycinnamoyl)transferase induced in response to UV-C and wounding from Capsicum annuum." Plant Cell Physiol, **42**, 475-481.

[151] **C. Comino, S. Lanteri, E. Portis, A. Acquadro, A. Romani, A. Hehn, R. Larbat, F. Bourgaud** (2007). "Isolation and functional characterization of a cDNA coding a hydroxycinnamoyltransferase involved in phenylpropanoid biosynthesis in Cynara cardunculus L." BMC Plant Biol, **7**.

[152] **C. Landmann, S. Hucherig, B. Fink, T. Hoffmann, D. Dittlein, H. A. Coiner, W. Schwab** (2011). "Substrate promiscuity of a rosmarinic acid synthase from lavender (Lavandula angustifolia L.)." Planta, **234**, 305-320.

[153] **M. Sander, M. Petersen** (2011). "Distinct substrate specificities and unusual substrate flexibilities of two hydroxycinnamoyltransferases, rosmarinic acid synthase and hydroxycinnamoyl-CoA:shikimate hydroxycinnamoyl-transferase, from Coleus blumei Benth." Planta, **233**, 1157-1171.

[154] **C. Weitzel, M. Petersen** (2011). "Cloning and characterisation of rosmarinic acid synthase from Melissa officinalis L." Phytochemistry, **72**, 572-578.

[155] **I. A. Kim, B. G. Kim, M. Kim, J. H. Ahn** (2012). "Characterization of hydroxycinnamoyltransferase from rice and its application for biological synthesis of hydroxycinnamoyl glycerols." Phytochemistry, **76**, 25-31.

[156] **Y. Lin, X. Sun, Q. Yuan, Y. Yan** (2014). "Extending shikimate pathway for the production of muconic acid and its precursor salicylic acid in Escherichia coli." Metab Eng, **23**, 62-69.

[157] **J. E. Vick, E. T. Johnson, S. Choudhary, S. E. Bloch, F. Lopez-Gallego, P. Srivastava, I. B. Tikh, G. T. Wawrzyn, C. Schmidt-Dannert** (2011). "Optimized compatible set of BioBrick vectors for metabolic pathway engineering." Appl Microbiol Biotechnol, **92**, 1275-1286.

[158] **T. Knight** (2003). "Idempotent vector design for standard assembly of Biobricks." MIT Synthetic Biology Working Group Technical Reports.

[159] **N. Bernard, K. Johnsen, T. Ferain, D. Garmyn, P. Hols, J. J. Holbrook, J. Delcour** (1994). "NAD(+)-dependent D-2-hydroxyisocaproate dehydrogenase of Lactobacillus delbrueckii subsp. bulgaricus. Gene cloning and enzyme characterization." Eur J Biochem, **224**, 439-446.

[160] **J. Ehlting, D. Buttner, Q. Wang, C. J. Douglas, I. E. Somssich, E. Kombrink** (1999). "Three 4-coumarate:coenzyme A ligases in Arabidopsis thaliana represent two evolutionarily divergent classes in angiosperms." Plant J, **19**, 9-20.

[161] **A. Berger, J. Meinhard, M. Petersen** (2006). "Rosmarinic acid synthase is a new member of the superfamily of BAHD acyltransferases." Planta, **224**, 1503-1510.

[162] **G. della-Cioppa, S. J. Garger, G. G. Sverlow, T. H. Turpen, L. K. Grill** (1990). "Melanin production in Escherichia coli from a cloned tyrosinase gene." Biotechnology (N Y), **8**, 634-638.

[163] **V. H. Lagunas-Munoz, N. Cabrera-Valladares, F. Bolivar, G. Gosset, A. Martinez** (2006). "Optimum melanin production using recombinant Escherichia coli." J Appl Microbiol, **101**, 1002-1008.

[164] **M. I. Chavez-Bejar, V. E. Balderas-Hernandez, A. Gutierrez-Alejandre, A. Martinez, F. Bolivar, G. Gosset** (2013). "Metabolic engineering of Escherichia coli to optimize melanin synthesis from glucose." Microb Cell Fact, **12**, 108.

[165] **Y. F. Yao, C. S. Wang, J. Qiao, G. R. Zhao** (2013). "Metabolic engineering of Escherichia coli for production of salvianic acid A via an artificial biosynthetic pathway." Metab Eng, **19**, 79-87.

[166] **C. Tokuda, Y. Ishikura, M. Shigematsu, H. Mutoh, S. Tsuzuki, Y. Nakahira, Y. Tamura, T. Shinoda, K. Arai, O. Takahashi, H. Taguchi** (2003). "Conversion of Lactobacillus pentosus D-lactate dehydrogenase to a D-hydroxyisocaproate dehydrogenase through a single amino acid replacement." J Bacteriol, **185**, 5023-5026.

[167] **Y. Ishikura, S. Tsuzuki, O. Takahashi, C. Tokuda, R. Nakanishi, T. Shinoda, H. Taguchi** (2005). "Recognition site for the side chain of 2-ketoacid substrate in d-lactate dehydrogenase." J Biochem, **138**, 741-749.

[168] **M. Beeby, T. A. Bobik, T. O. Yeates** (2009). "Exploiting genomic patterns to discover new supramolecular protein assemblies." Protein Sci, **18**, 69-79.

[169] **C. A. Kerfeld, S. Heinhorst, G. C. Cannon** (2010). "Bacterial microcompartments." Annu Rev Microbiol, **64**, 391-408.

[170] **B. D. Rae, B. M. Long, M. R. Badger, G. D. Price** (2013). "Functions, compositions, and evolution of the two types of carboxysomes: polyhedral

microcompartments that facilitate CO2 fixation in cyanobacteria and some proteobacteria." Microbiol Mol Biol Rev, **77**, 357-379.

[171] **M. R. Badger, G. D. Price** (2003). "CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution." J Exp Bot, **54**, 609-622.

[172] **G. D. Havemann, T. A. Bobik** (2003). "Protein content of polyhedral organelles involved in coenzyme B-12-dependent degradation of 1,2-propanediol in Salmonella enterica serovar typhimurium LT2." J Bacteriol, **185**, 5086-5095.

[173] **M. R. Rondon, R. Kazmierczak, J. C. Escalantesemerena** (1995). "Glutathione Is Required for Maximal Transcription of the Cobalamin Biosynthetic and 1,2-Propanediol Utilization (Cob/Pdu) Regulon and for the Catabolism of Ethanolamine, 1,2-Propanediol, and Propionate in Salmonella-Typhimurium Lt2." J Bacteriol, **177**, 5434-5439.

[174] **E. M. Sampson, T. A. Bobik** (2008). "Microcompartments for B-12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate." J Bacteriol, **190**, 2966-2971.

[175] **T. A. Bobik, G. D. Havemann, R. J. Busch, D. S. Williams, H. C. Aldrich** (1999). "The propanediol utilization (pdu) operon of Salmonella enterica serovar typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B-12-dependent 1,2-propanediol degradation." J Bacteriol, **181**, 5967-5975.

[176] **E. Kofoid, C. Rappleye, I. Stojiljkovic, J. Roth** (1999). "The 17-gene ethanolamine (eut) operon of Salmonella typhimurium encodes five homologues of carboxysome shell proteins." J Bacteriol, **181**, 5317-5329.

[177] **M. Held, M. B. Quin, C. Schmidt-Dannert** (2013). "Eut bacterial microcompartments: insights into their function, structure, and bioengineering applications." J Mol Microbiol Biotechnol, **23**, 308-320.

[178] **R. J. Conrado, T. J. Mansell, J. D. Varner, M. P. DeLisa** (2007). "Stochastic reaction-diffusion simulation of enzyme compartmentalization reveals improved catalytic efficiency for a synthetic metabolic pathway." Metab Eng, **9**, 355-363.

[179] **C. Fan, S. Cheng, S. Sinha, T. A. Bobik** (2012). "Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments." Proc Natl Acad Sci U S A, **109**, 14995-15000.