Expressing and Engineering Natural Product Enzymes in Bacterial Hosts

By

Taylor B. Cook

A dissertation submitted in partial fulfillment of the requirements for the degree of:

Doctor of Philosophy

(Chemical and Biological Engineering)

UNIVERSITY OF WISCONSIN-MADISON

2021

Date of Final Oral Examination: 02 June 2021

The dissertation is approved by the following members of the Final Oral Committee:

Brian F. Pfleger, Professor, Chemical and Biological Engineering
Eric V. Shusta, Professor, Chemical and Biological Engineering
Philip Romero, Assistant Professor, Chemical and Biological Engineering
Michael G. Thomas, Professor, Bacteriology

Abstract

Drug discovery has depended primarily on the isolation of natural products from living organisms because these compounds often have inherent properties applicable to modern medicine. Microorganisms from the environment proved to be rich sources of natural products with antibiotic, anticancer, and immunosuppressant activities, but as this resource was exhausted over the last century, researchers have had to resort to alternative strategies for developing novel drugs. One such strategy is to engineer microorganisms to produce non-native natural products of interest. This is a multi-disciplinary effort that applies synthetic biology and metabolic engineering towards the manipulation of natural product chemistry. The biosynthetic pathways for two natural product classes, polyketides and nonribosomal peptides, are particularly challenging to manipulate due to their reliance on large, modular enzymes. In this dissertation, we identified *Pseudomonas putida* as an ideal host for expressing and engineering these enzymes. To facilitate strain engineering of this bacterial host, we developed a genome editing toolkit that enabled chromosomal deletions, integrations, and point mutations. We then applied these methods towards expressing heterologous pathways in *P. putida* to produce polyketides and nonribosomal peptides. Chromosomal integration of a heterologous gene cluster responsible for prodigiosin biosynthesis resulted in a strain capable of producing 1.1 g/L of product. *P. putida* has a relatively high-GC chromosome, but we found that expression of a low-GC gene cluster for glidobactin A biosynthesis resulted in higher heterologous titers compared to a related high-GC gene cluster. Further strain and pathway engineering, including improvements to pathway expression and deletion of a native carbon sink, resulted in a strain capable of producing 470 mg/L glidobactin A. *P. putida* natively

produces a siderophore called pyoverdine through a nonribosomal peptide synthetase (NRPS). We utilized this pathway to explore subdomain substitutions for altering substrate specificity in NRPSs. Subdomain boundaries for creating functional chimeras were identified, but the most active variants still performed poorly compared to the wild-type enzyme. The findings discussed here will enable the production of novel compounds through the heterologous expression of uncharacterized pathways and the manipulation of key catalytic domains in modular megaenzymes.

To Nana and Pappy

Acknowledgments

They say God laughs at man's plans, and if the past year doesn't make you believe that, then I don't know what will. Deadly pandemic aside, I am impressed at how often I've managed to change my plans over the last six years. I've always been a little unsure of myself. I still am, but I've learned to be more comfortable with it (and to fake it when necessary). Whether it's canceled projects or canceled flights, I've had to learn how to accept and move on from failed plans many times over the last five years. This was only possible thanks to the countless people that have helped me over the years.

I'd like to start by thanking everyone in the Pfleger and Thomas Labs for helping me get to this point. Thank you to Brian for the opportunity to work in this crazy group. I appreciate the freedom you gave me to explore my research ideas and the constructive environment that you promoted within the lab, and thank you to Michael Thomas for essentially being a second advisor and the ultimate example of how to perform rigorous science. To all the Pfleger and Thomas Lab members that were around when I first started research, thank you for the constant support and making me feel comfortable to express myself. Most of my accomplishments would not have been possible without the mentorship of Jackie Rand, who set me on the right path in the *P. putida* world. Hopefully I can return the favor when you start looking for a job in Boston! Thank you to Mark Politz and Chris Mehrer for offering great advice and showing an interest in my projects when I needed it. You two were great at showing me a different perspective (in

politics as well as science!), and I'm all the better for it. Thank you to Vlad Vinnik and Kurt Throckmorton for welcoming an awkward engineer into the Thomas Lab.

To Ryan Clark, Néstor Hernandez-Lozada, and Travis Korosh, thank you for all of the fascinating conversations, whether it was about music, politics, or science. To Dylan Courtney and Will Cordell, thank you for teaching me so much about metabolism or bioreactors and for playing Dark Souls with me. To Mike Jindra, thank you for being a lively bench mate and talking to me about fashion. To Francesca Gambacorta, thank you for picking up the slack in lab when I didn't do my job. It was obvious from the beginning that you would be a much better Lab Daddy than me. To Maya Venkataraman, thank you for being a great office-mate and tolerating my random thoughts about politics. To my favorite post-docs, Will Bothfeld and Qiang Yan, thank you for the helpful ideas and encouragement, especially during my last year of grad school. To my mentees, Sophia Liu and Aditya Ailiani, thank you for all the interest and hard work you put into research. One final thank you to all of you for being such great friends. It was never hard going into work knowing that I enjoyed the company of everyone around me.

A round of thanks to everyone in the Library trivia group (Jackie, Chris, Matt Long, Tim Smith, Hector), especially for knowing so much more than me about history and sports that aren't basketball. Thank you to the Mavrikakis members of the Waffle Wednesday crew (Sean Tacey, Ellen Murray, Michael Rebarchik) for giving me a reason to wake up early in the middle of the week. A second shout-out to Sean Tacey for obsessing over basketball and memes as much as me and teaching me everything I know about ultimate.

To all my other non-bio friends: Keishla Rivera-Dones, Coogan Thompson, Hector Fuster, thank you for always looking out for me and letting me be my awkward self around you. A special thank you to Hector again for being my best friend and the reason I had a social life outside of the Pfleger Lab for most of grad school. To Curran and Laura Gahan, thank you for including me in all your social events at The Sheelin. To Frank Nguyen and Parth Mangrolia, thank you for including me in your brunch and happy hour outings and for hosting the best watch parties. To Jonathan and Erin Sheavly, thank

you for introducing me to the best cat in the world, Theo.

Words can't describe how much I owe my parents, David and Nancy Cook, for always supporting and loving me, even when I wanted to move further away from home. You are a constant that I need in a fast-paced and busy world. To my sisters, Emily and Kaitlyn, thank you for becoming my best friends over the last few years. I'm so proud of you two for all of your accomplishments, and I'm looking forward to all of us being done with school and entering adulthood. To Nana, thank you for your unconditional love and always asking me what food I wanted when I planned to fly home. It hurts that you aren't here to celebrate this moment with us. You were the first one to say that I was going to cure cancer. While that isn't *technically* correct, it is fitting that I ended up researching an anticancer compound (and started dating a pharmacist that specializes in oncology!).

Last but not least, thank you Lauren for your kindness and support during an especially difficult year. You continue to inspire me with how hard you work and dedicate yourself to your family and friends. Whether we're acting like kids by watching Disney movies or growing up by learning how to cook together, so many of my highlights for the last 2.5 years have been with you.

"Your flowery description's no better than his! We sent for the Great Light and you bring us this? We didn't ask what it seems like, we asked what it IS!"

- Aaron Weiss, inspired by the works of M. R. Bawa Muhaiyaddeen

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Surveying Earth's natural diversity to address societal problems

Humankind has a long history of making substantial improvements to its way of life through the discovery and experimentation of existing biological systems. The domestication of plants and animals enhanced our access to food and enabled the expansion of human civilizations thousands of years ago. People often observed and tinkered with the world around them without knowledge of the underlying phenomena. For example, the discovery of fermented grains and fruits was a boon to human recreation, long before Louis Pasteur determined the role that yeast played in the process [1]. Even experimentation with harmful organisms was beneficial, as the first vaccines were developed in the 1800s by discovering the infectious agent and producing a weakened version [2]. Around the same time, medicinal chemistry was maturing as chemists discovered the bioactive compounds, also known as natural products, responsible for the therapeutic effects of plant extracts. This trend continued in the twentieth century, and medicine depended primarily on natural products from plants and microorganisms as candidates for novel drugs. These compounds were commonly used as antibiotics, but natural product-derived drugs have also been developed for treating various cancers [3].

Natural products are often involved in biomolecular interactions beneficial to the organisms that synthesize them. As organisms evolve, so do the genes, pathways, and metabolites within them, so it follows that existing natural products have been subject to natural selection over time [4]. Sometimes, the evolutionary path towards the function of a natural product is obvious. For example, compounds with antibiotic activity benefit host organisms by preventing the growth of other species that can compete for nutrients. These chemical species have a natural fit in modern medicine where they are administered to patients to treat bacterial and fungal infections.

For other types of medical treatments, there is not always a clear connection between a compound's native function and its clinical purpose. Anti-cancer drugs isolated from microorganisms are unlikely to have encountered mammalian cancer cells before their discovery, and the exact mechanism for how they benefit the host's survival is often unknown. However, regardless of the natural product's activity, evolution primarily selects for specific interactions with biomolecules, such as proteins. This selection results in small molecules with three-dimensional structures poised to interact with other proteins and enzymes, including those involved in the generation of cancer cells in the human body [5]. Synthetic ligands are relatively more two-dimensional compared to natural product-derived molecules, leading to a much higher fraction of screened compounds with no detectable bioactivity. Drug development in the 1970s and 1980s focused primarily on synthetic ligands, but this period was short-lived compared to natural product-derived development throughout the twentieth century, likely due to these structural differences. The evolutionary and structural advantage of natural products is why they are considered "privileged" compounds, and the majority of clinical drugs today are derived from natural products or mimics [6].

As medicinal chemistry progressed into the twentieth century, chemists discovered the potential of microorganisms for drug discovery. The first antibiotics isolated from a microorganism were the penicillins, discovered by Alexander Fleming in 1929 [7]. Penicillins are naturally produced by filamentous fungi in the genus *Penicillium*, and they were one of the early drugs used to treat common bacterial infections, including those known to

cause scarlet fever, pneumonia, and meningitis [8]. Early pharmaceutical companies, such as Pfizer, developed the commercial production of the antibiotic by isolating new *Penicillium* strains and optimizing growth conditions in large fermentation tanks, a necessary advancement for scaling up penicillin production and distribution. The mass production of penicillin was instrumental in enabling the treatment of bacterial infections for Allied soldiers during World War II, and this successful story prompted an exhaustive search for more bioactive compounds from microorganisms throughout the twentieth century.

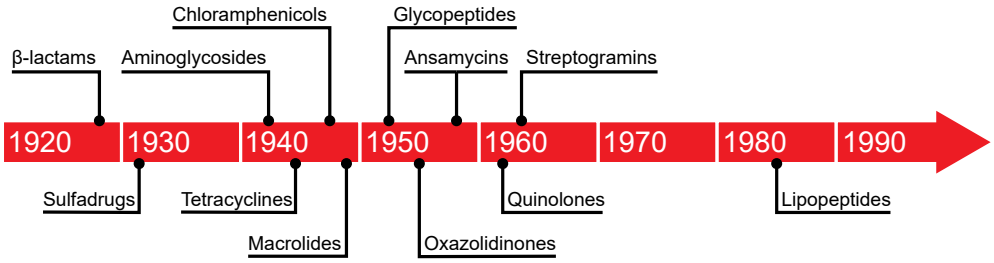


Figure 1.1: **Timeline of antibiotic discovery in the 1900s.** Antibiotic discovery from natural products peaked from the 1930s to the 1960s, until synthetic molecules were prioritized in the search of new drugs. Aminoglycocides, chloramphenicols, and tetracyclines were all discovered from Actinomycetes. The most recently discovered class of antibiotics are the lipopeptides, also found primarily in Actinomycetes. Figure was adapted from Lewis 2012.

Drug discovery from microorganisms continued successfully through the 1960s, and the major bacterial source of novel compounds was a group of soil bacteria, the Actinomycetes. These bacteria provided several classes of antibiotics, including chloramphenicols and tetracyclines (Figure 1.1) [9]. However, the broad use of antibiotics in medicine increased the prevalence of bacterial and fungal infections with antibiotic resistance, reducing the effectiveness of these treatments. As early antibiotics were being phased out, scientists were exhausting soil bacteria as a resource for novel compounds, creating the current medical crisis where antibiotic discovery is racing against the spread of multidrug-resistant pathogens [10]. Penicillin mass production is an exemplary story about identifying a natural biological system, i.e. a *Penicillium* fungus, and taking advantage of its native abilities to advance medical treatments, but this strategy for drug discovery is less viable today. It is increasingly difficult to discover more microorganisms that produce novel drug candidates. Even if a new compound is discovered, its native host may not be amenable to large-scale cultivations, or researchers may not be able to

determine the culture conditions that induce production of the bioactive compound. Alternative strategies for discovering novel drug candidates and scaling up their production are necessary if drug discovery is to keep up with the current needs in modern medicine. A relatively new strategy for addressing this problem is the modification of metabolic pathways within microorganisms. This growing field has taken advantage of modern methods in synthetic biology and metabolic engineering to alter the genetic material in microorganisms such that they produce a compound of interest. A second advantage to this strategy is that the host organisms for these projects are often amenable to large-scale cultivations, providing an avenue for scaling up production for clinical trials and hopefully mass production.

1.2 Designing biological systems via rational engineering and directed evolution

Engineering metabolic pathways begins by manipulating DNA sequences within an organism, and the principles governing DNA design are based on our knowledge of molecular biology: the study of processes involving DNA, RNA, and protein synthesis in living cells. These processes are often described as the central dogma of molecular biology, where genes are encoded and preserved in DNA, RNA is transcribed based on these gene sequences, and protein is translated from the "message" provided by RNA [11]. Each type of molecule in this process has a "language" based on molecular building blocks. DNA and RNA use nucleotides to encode a message that determines a sequence of amino acids, which in turn determines the function of a protein. In the field of synthetic biology, researchers design DNA sequences with the goal of controlling the amounts of DNA, RNA, and protein in the host organism as well as tuning protein function. Many proteins function as enzymes, which catalyze specific reactions that make up cellular metabolism. The major challenge of synthetic biology is reliably controlling the processes in the central dogma. Metabolic engineering is a common application of synthetic biology where genetic modifications in a host organism changes cellular metabolism, often for the goal

of overproducing a metabolite of interest.

Molecular biology and metabolism vary considerably between organisms, so strategies in metabolic engineering depend on *a priori* knowledge of the organism of interest. When a species is well-studied, rational approaches to engineering metabolism are appropriate because researchers can reliably control gene expression and predict how changes in enzyme production affect metabolism. Engineering poorly studied organisms, often called non-model organisms, requires random approaches where a large number of strain and pathway variants are analyzed for performance in high-throughput screens or selections in a manner similar to natural evolutionary processes.

1.2.1 Rational approaches to genetic and metabolic engineering

The metabolic networks in living organisms are complex, and in the case of bacteria, contain thousands of reactions and metabolites [12]. Despite this complexity, metabolic changes can be categorized into three basic groups: downregulation, upregulation, and mutation. Downregulation involves the reduction or deletion of activity of a reaction or pathway. This modification is often used when a pathway is purported to compete for metabolites necessary for the synthesis of a target compound. Upregulation results in an increase in reaction or pathway activity and is useful when a set of reactions is known to synthesize precursor metabolites necessary for production. Lastly, mutation focuses on changes to enzyme activity rather than abundance. Enzymes can be tuned by mutations in their sequence, or they can be replaced altogether by similar enzymes from other organisms. These three strategies for metabolic engineering are often combined in an iterative fashion while monitoring their effect on production, a process commonly referred to as a design-build-test cycle [13].

There are many variables that researchers must consider when shaping metabolism in engineered organisms, and we must understand the host's metabolism well enough in order to effectively design these strategies. An essential tool for this research is genome-scale metabolic models, which are mathematical representations of the metabolic path-

ways known in an organism of interest [14]. These models simulate the flow of metabolites in an organism under standard laboratory conditions. They can also be predictive and be used to identify targets for downregulation and upregulation that would increase flux towards a desired metabolite [15]. This is a powerful tool because it allows researchers to quickly identify metabolic engineering strategies without performing experiments.

Once alterations to an organism's metabolism are identified, methods for changing its genetic material are required. These methods introduce DNA designed by researchers, also known as recombinant DNA, into the cell. The primary goal of recombinant DNA is to produce a gene product, either RNA or protein, encoded by a gene sequence, a process called gene expression. The product has a desired function in the cell, such as enzymatic activity or gene regulation. Most DNA constructs are designed to synthesize a protein, and in this case, the intermediate RNA is called messenger RNA (mRNA). Although less common, it is possible to design DNA to express non-coding RNA (ncRNA), which does not have the sequences necessary to produce protein, but still confers a function in the cell as an RNA molecule.

Specific sequences encoded in DNA are required for gene expression (Figure 1.2a). First, a promoter sequence activates transcription, the synthesis of RNA by RNA polymerase. When a protein is the desired gene product, downstream of the promoter is the ribosome binding site (RBS) sequence. The RNA encoded by this sequence recruits the ribosome to mRNA and initiates translation, the synthesis of protein from RNA. Next, a coding sequence (CDS) is downstream the RBS. The CDS contains a sequence of codons, groups of three DNA/RNA bases, that determine the amino acid sequence of the encoded protein, and therefore its structure and function. In the case of bacteria, multiple CDSs can exist next to each other while under control of the same promoter, creating an operon, although functional RBS sequences must be located upstream of each CDS. Lastly, a terminator is usually included at the end of the operon, which signals the RNA polymerase to stop transcribing RNA.

Each component described above provides an opportunity for controlling the final amount of RNA or protein produced. Promoter strength is a common variable for con-

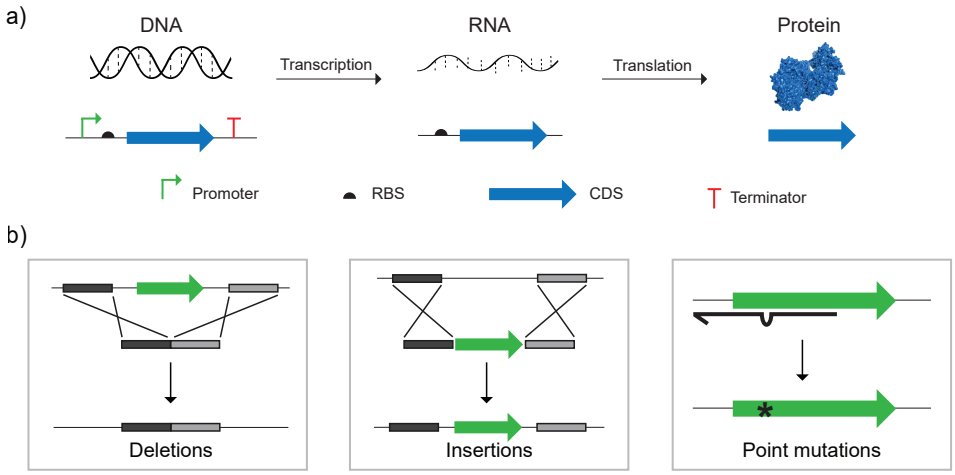


Figure 1.2: **Genetic parts required for gene expression and various types of genome editing. a)** Gene expression begins with a DNA sequence that contains a promoter, RBS, CDS, and a terminator. During transcription, the DNA containing the RBS and CDS are transcribed into RNA. Lastly, the CDS is translated into the encoded protein. **b)** The three types of genome editing are deletions, insertions, and point mutations. A DNA construct interacts with the chromosome so that the desired edit is achieved.

trolling gene expression, where "strong" promoters directly lead to more RNA production, which then results in more protein production. Promoters that have a constant expression level are called constitutive promoters. Alternatively, inducible promoters exhibit varying levels of transcription based on the addition of a chemical inducer or an environmental condition (e.g. pH or temperature). Protein production can also be controlled at the translational level. RBS sequences vary in strength and affect the translation initiation rate by the ribosome. Codon usage in a CDS can also affect expression levels by modulating the efficiency of translation elongation. Organisms have preferred codons for gene expression, and exogenous genes typically have higher expression levels if they do not contain codons that the host organism rarely uses in its native genes [16]. Lastly, codons can also be mutated to change the protein sequence, which can affect protein stability or activity.

The variables described above allow for great control of protein production and activity, especially in well-studied organisms, but they only cover transcriptional and translational control. The first step in the central dogma, DNA replication, can also affect how a DNA construct performs. Once recombinant DNA has entered the cell, it is usually maintained as a replicative plasmid, a circular DNA molecule that replicates indepen-

dently of the chromosome. It is also possible to insert DNA directly into the chromosome in a process called genome editing. Introducing exogenous DNA is faster with replicative plasmids and there are more copies of plasmid DNA compared to the chromosome, which increases gene copy number. However, DNA that has recombined with the chromosome is more stable and requires fewer cellular resources to maintain. In addition, strategies for altering native metabolic pathways in an organism requires genome editing. For these reasons, methods for editing the genome are a necessity in metabolic engineering.

Genome editing enables several genetic changes that affect metabolism. The most straight-forward genetic change is the deletion of native genes from the chromosome, which is often used to remove competing pathways if they are not essential for cell growth (Figure 1.2b). In cases where essential pathways are deleterious to metabolite production, there are some methods that enable repression of the pathway, which reduces expression without completely removing it from the cell [17]. It is also possible to integrate new DNA directly into the chromosome, which enables the overexpression of pathways required for metabolite production. If strong promoters are known for the host, then one option is to insert them in front of native genes to increase their expression. If a pathway needed to produce a metabolite of interest is not encoded in the host's chromosome, then genes from a different organism can be integrated with a strong promoter, resulting in heterologous expression. Lastly, a versatile genome editing strategy is to introduce point mutations into the chromosome. This technique enables more subtle changes to native gene expression through changes to promoter and RBS sequences, but as discussed above, it is also possible to modify protein activity by mutating codons in a gene of interest. [18].

The introduction so far highlights the versatility of common metabolic engineering strategies. Researchers have the tools necessary to design and create DNA constructs to effect a desired metabolic shift. However, there are many variables governing transcription and translation to consider. It is difficult to predict the exact strength of a promoter or RBS needed to achieve the desired gene expression. In addition, the exact amount of enzyme required for optimal metabolic flux is usually unknown. Metabolic engineering

often involves multiple gene targets, which considerably increases the number of genetic variables. These challenges can sometimes be overcome by creating multiple DNA constructs with different promoters, RBS sequences, or genes. These efforts are laborious and time consuming, but often yield favorable results in model organisms [19]. This strategy is limited for non-model organisms, as the exact role of native enzymes may not be clear, and there may be unknown regulatory processes that interfere with the desired metabolic changes. In these cases, rational approaches are less reliable, and researchers must turn to random and high-throughput methods for metabolic engineering.

1.2.2 Laboratory evolution of metabolic pathways and enzymes

Biotechnology research has greatly benefited from methods inspired by natural selection. Scientists have developed strategies for diversifying biological systems, including whole organisms, pathways, or proteins, followed by isolating variants that have the desired performance. These systems are usually poorly understood and it is difficult to hypothesize specific changes that improve function. Researchers address this issue by generating a large library of variants, in the range of thousands up to millions, through a mostly random process. Due to the magnitude of these libraries, it is necessary to devise a high-throughput screen or selection based on the activity of the engineering target. This process is referred to as laboratory evolution or directed evolution, and it is often cyclical, where individuals isolated from early rounds of evolution are the starting material for later rounds (Figure 1.3). From a molecular biology perspective, each cycle begins by mutating DNA sequences that encode the engineering target. This DNA can be a single gene, multiple genes, or even an entire chromosome. This results in a library of DNA constructs that are assessed in a host organism by a selection or screen, and afterwards researchers determine the DNA sequences of individuals with the desired activity. Mutants with improved performance are mutated further, and the cycle continues until the design goal is achieved.

A selection or screen must be carefully chosen when performing laboratory evolution.

This step is entirely dependent on the activity of the engineering target and the size of the library to be analyzed. When a library is subjected to a selection, individuals that do not reach a desired activity threshold are intentionally lost. This allows researchers to quickly reduce the number of variants to characterize, and selections are preferred when analyzing larger libraries, especially when a large percentage of the library is expected to be non-functional. A screen is defined as a rapid assay that generates a simple output for every individual in a library. Collecting information on every individual makes screens more cumbersome for large libraries and requires more work if the fraction of non-functional variants is too high. Nevertheless, an obvious link between a design goal and a simple output for a screen or selection is not always available, and therefore the act of analyzing large libraries is usually the rate-limiting step in laboratory evolution. Advancements in automation technology have started to enable the laboratory evolution of biological systems where a screen or selection is not available [20]. Liquid-handling robots can be programmed to perform many experimental steps at a much faster rate than individual researchers are capable of, increasing the speed at which large libraries are analyzed.

Methods for diversifying biological systems depend on the engineering target, whether it is a microorganism, a pathway, or a single protein. To evolve microorganisms in the lab, researchers must generate mutations in the chromosome that affect how native genes are expressed. These mutations may improve cell growth on a specific nutrient source or shift cellular metabolism so that production of a metabolite of interest increases. A passive approach to generating mutants is adaptive laboratory evolution (ALE), where a strain is passaged continuously in selective media while the population of cells in the cultures

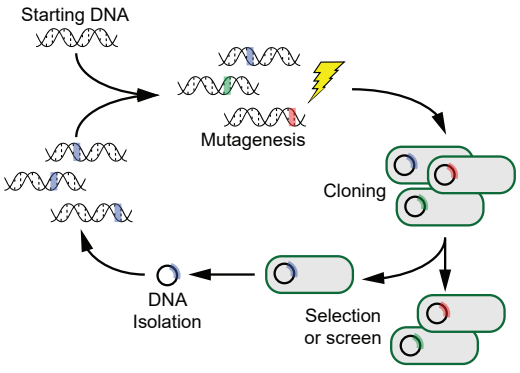


Figure 1.3: **Directed evolution is a cyclical process with multiple steps.** Mutant libraries are generated from a single DNA sequence (chromosome, group of genes, or single gene). A selection or screen is used to isolate variants with the desired activity. Variants of interest are then replicated and used as the starting point in another round of evolution.

obtain beneficial mutations that improve their growth characteristics (Figure 1.4a) [21]. An alternative method for mutating chromosomes is transposon mutagenesis, where a DNA construct is randomly inserted into the chromosome [22]. This process results in a library of strains that have an insertion in a different location on the chromosome, which is likely to disrupt a gene and prevent its expression.

Evolving pathways and proteins involves more active methods for creating diversity. The DNA encoding these targets is much smaller than a chromosome, so researchers can take advantage of many methods for generating synthetic DNA. When optimizing a metabolic pathway, the expression levels and the activity of each enzyme are altered by changing the promoter and RBS sequences for every gene, as well as the gene products. A common strategy is to generate combinatorial libraries of DNA constructs encoding the pathway of interest, where researchers choose from different promoters and RBS sequences of varying strength and from multiple homologues for each enzyme (Figure 1.4b).

In metabolic engineering, it may be possible to significantly improve production by focusing on a single enzyme. Researchers have dedicated decades of research towards understanding the relationship between sequence and function in proteins, particularly for enzymes. There have been some successes in protein engineering via rational design, but this approach often requires structural information on the protein of interest, which is not always easy to obtain. Multiple methods have been developed for creating protein libraries, including random mutagenesis, saturation mutagenesis, and gene shuffling (Figure 1.4c) [23]. Directed evolution is the ideal strategy for protein engineering due to the ability for these techniques to generate massive protein libraries and the availability of screens and selections that researchers have coupled to various protein activities. Evolving enzymes has another benefit for metabolic engineering in that enzyme variants with altered specificity can be created to build novel metabolic pathways in addition to optimizing existing pathways [24].

There is clearly no shortage of cutting edge techniques to engineer biological systems, either through rational or evolutionary approaches. However, the application of these

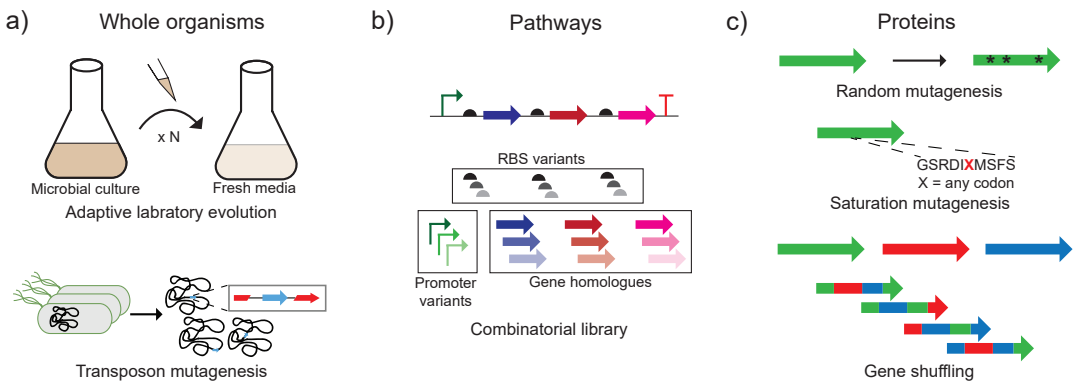


Figure 1.4: **Methods for generating diversity in biological systems. a)** Generating mutants of whole organisms. In ALE, cells are passaged from a culture to fresh media N times until the design goal is met. Transposon mutagenesis is used to create a library of strain mutants with random knockouts in the chromosome. **b)** Pathway variants are generated by cloning combinatorial libraries, where all combinations of promoter, RBS, and gene variants are tested for performance. **c)** Protein libraries for directed evolution are generated through three main techniques. Random mutagenesis introduces point mutations throughout the entire gene or a continuous section of the gene. Saturation mutagenesis focuses on one or a few codons that are mutagenized so that every amino acid is sampled at each one. Gene shuffling randomly combines blocks from parent genes to create a library of chimeras.

tools towards natural product biosynthesis has been limited. Secondary metabolism is generally not understood as well as primary metabolism, and much work has to be devoted towards characterizing the enzymes involved in natural product biosynthesis. Pathway size and complexity also increase the difficulty in creating and testing combinatorial libraries. Evolutionary approaches to investigating natural products are also limited because it is rarely possible to couple secondary metabolism to cell growth. Pathway or enzyme variants have to be analyzed with screens, limiting library size and the rate at which natural product biosynthesis can be engineered.

1.3 Overview of Thesis

Throughout this thesis, we have tackled multiple challenges associated with improving and modifying natural product biosynthesis in bacteria. Our strategies included genome editing, protein directed evolution, and metabolic engineering. These accomplishments were largely made possible by the wealth of synthetic biology tools available for two bacteria, *Escherichia coli* and *Pseudomonas putida*.

1.3.1 Chapter 2

In Chapter 2, we review efforts to develop bacterial hosts for expressing heterologous genes and pathways, with a focus on *P. putida*. There is considerable interest in identifying ideal strains for producing heterologous natural products, particularly polyketides and non-ribosomal peptides, and recent progress in bacterial hosts is reviewed. Polyketides and non-ribosomal peptides are synthesized by large, multi-modular enzymes, which increases the difficulty in modifying catalyic domains. We compare strategies to engineer these mega-enzymes in order to synthesize natural product derivatives.

1.3.2 Chapter 3

In Chapter 3, a set of synthetic biology tools developed for *P. putida* were characterized. We compared gene expression from several inducible promoters and determined the plasmid copy number for broad host range origins. Multiple scarless deletions in the *P. putida* chromosome were created using Cas9-asssisted homologous recombination.

1.3.3 Chapter 4

In Chapter 4, we expanded our genome editing tools to engineer *P. putida* for the heterologous production of polyketides and non-ribosomal peptides. We successfully constructed strains capable of producing two bioactive natural products, prodigiosin and glidobactin A. We compared heterologous production titers on various carbon sources and devised a workflow for increasing production titers from heterologous pathways containing polyketide synthases and non-ribosomal peptide synthetases. Lastly, we identified a carbon sink in native *P. putida* metabolism and improve production of heterologous products after its removal.

1.3.4 Chapter 5

In Chapter 5, we describe our efforts to characterize enzymes involved in the synthesis of bacterial siderophores in collaboration with Prof. Michael Thomas. A thorough analysis of a specificity-determining enzymatic domain in non-ribosomal peptide biosynthesis was completed through directed evolution of the enterobactin-producing enzyme, EntF, from *E. coli*. We explored strategies for swapping domains and modules in non-ribosomal peptide synthetases responsible for pyoverdine synthesis in *P. putida*. Lastly, we demonstrated that a heme peroxidase was responsible for the conversion of L-tyrosine to L-DOPA in the biosynthesis of fabrubactin, a siderophore from *Agrobacterium tumefaciens*.

1.3.5 Chapter 6

In Chapter 6, we explored the utility of Pseudomonads as heterologous hosts for extracellular enzymes. After identifying optimal promoters and plasmid origins for heterologous protein production in *P. putida*, an enzyme that is natively secreted by a bacterial host was overexpressed along with its associated chaperone in multiple *Pseudomonas* strains.

1.3.6 Chapter 7

In Chapter 7, we expressed mammalian enzymes that complemented some steps in *P. putida* metabolism. These enzymes are classified as acyl-CoA dehydrogenases (ACADs) and have homology to two enzymes that catalyze the first two steps in levulinic acid catabolism. One ACAD was able to recover growth on levulinic acid in a strain missing the first enzyme in the levulinic acid pathway.

1.3.7 Chapter 8

Finally, in Chapter 8, we identify steps to further the work done here to engineer the production of natural products. Our workflow for expressing heterologous pathways in *P.*

putida could be applied towards the discovery and production of novel natural products by introducing uncharacterized gene clusters from slow-growing cyanobacteria. Evaluating the level of heterologous expression throughout the *P. putida* chromosome would enhance future strain development. A third project would provide more insight on the activation of amino acids in non-ribosomal peptide biosynthesis by applying directed evolution on EntF chimeras. We end this discussion by taking a broader look at how synthetic biology will play a role in natural products research.

Chapter 2

Leveraging synthetic biology for the production and engineering of natural product biosynthesis

Authors and Contributors:

- **Taylor B. Cook** Wrote the manuscript.
- **Dr. Brian F. Pfleger** Provided guidance, edited, and reviewed the manuscript.

A section of this chapter was originally written as an article in *MedChemComm* [25].

2.1 Abstract

Bacteria have historically been a rich source of natural products (e.g. polyketides and nonribosomal peptides) that possess medically-relevant activities. Despite extensive discovery programs in both industry and academia, a plethora of biosynthetic pathways remain uncharacterized and the corresponding molecular products untested for potential bioactivities. This knowledge gap comes in part from the fact that many putative natural product producers have not been cultured in conventional laboratory settings in which the corresponding products are produced at detectable levels. Next-generation sequencing technologies are further increasing the knowledge gap by obtaining metagenomic sequence information from complex communities where production of the desired compound cannot be isolated in the laboratory. For these reasons, many groups are turning to synthetic biology to produce putative natural products in heterologous hosts. This strategy depends on the ability to heterologously express putative biosynthetic gene clusters and produce relevant quantities of the corresponding products. Actinobacteria remain the most abundant source of natural products and the most promising heterologous hosts for natural product discovery and production. However, researchers are discovering more natural products from other groups of bacteria, such as myxobacteria and cyanobacteria. Therefore, phylogenetically similar heterologous hosts have become promising candidates for synthesizing these novel molecules. The downside of working with these microbes is the lack of well-characterized genetic tools for optimizing expression of gene clusters and product titers. Synthetic biology tools are advancing rapidly for some heterologous hosts, making them easier to manipulate and enabling more advanced strategies for engineering natural product synthesis, such as protein engineering for altering enzyme specificity in order to synthesize novel derivatives. This chapter examines heterologous expression of natural product gene clusters in terms of the motivations for this research, the traits desired in an ideal host, tools available to the field, and a survey of recent progress. An emphasis is placed on the synthetic biology of the soil bacterium, *Pseudomonas putida*, and recent successes in developing this microbe as a host for engineering polyketide and nonribosomal peptide production.

2.2 Introduction

Bacteria are valuable sources of natural products with medically relevant activities [26]. For most of the twentieth century, more than 80% of medical compounds were derived from or inspired by natural products [27].2 Almost half of the natural products synthesized by bacteria possess some bioactivity, including antibiotic, anticancer, and immunosuppressant activities [28]. Demand for novel drugs with improved activities is increasing in part to the rise in multi-drug resistant infections and the ever-present need for a diverse set of cancer treatments [29, 30]. Two of the most intriguing classes of natural products are polyketides and non-ribosomal peptides, which contribute considerably to the number of known bioactive natural products. These compound families are made by megaenzymes with multiple catalytic domains in an assembly-line fashion. Since their discovery, the modular nature of these enzymes has promised the ability to use combinatorial biosynthesis to produce diverse compounds that could be screened for novel bioactivity [31]. However, after the initial explosion of new antibiotics in the 1950's and 1960's, the role of natural products in drug development decreased considerably as synthetic chemistry techniques accelerated the pace of discovery beyond the speed with which new bioactive compounds could be isolated from novel microbes and/or their underlying biosynthetic machinery engineered to produce diverse compound libraries [9]. With the advent of next-generation DNA sequencing technologies, the number of putative biosynthetic gene clusters (BGCs) encoding PKSs and NRPSs has since increased exponentially to over 70 000 clusters, but the number of clusters associated with specific compounds remains under 1000 [32]. This gap is caused in part by challenges in culturing native producers in conditions that maximize biosynthesis of the desired compound.

It has been estimated that 99% of bacteria have not yet been cultivated in conventional laboratory media [33], and many of the remaining bacteria have slow growth rates, do not produce natural products in tested cultivation conditions, and/or are not genetically tractable [34, 35]. Even in cases where the native host has been cultivated, substantial engineering may be required to produce relevant levels of putative secondary metabolites [36, 37]. For these reasons, heterologous expression has become an essential

tool in the genomic era of natural product discovery and development. While heterologous expression of natural product BGCs can provide substantial advantages, pitfalls are often encountered when developing and deploying heterologous hosts. Unfortunately, the common heterologous expression workhorse, *Escherichia coli*, has not proven to be a widely useful host for producing complex natural products including polyketides and non-ribosomal peptides. Therefore many groups have turned to non-model bacteria for heterologous production.

Optimizing a BGC for expression in a heterologous host is commonly described as "refactoring". Refactoring a pathway includes several strategies for distributing coding sequences into synthetic genetic circuits. These synthetic circuits contain a combination of promoters, RBSs, and terminators with known transcriptional and translational activities [38]. Coding sequences may also be codon optimized using de novo DNA synthesis so that their codon usage matches that of the heterologous host [39]. Characterized synthetic biology tools, which include genetic tools and methods for genetically modifying BGCs and hosts of interest, are necessary for researchers to refactor BGCs for optimal expression in heterologous hosts.

In this review, we discuss the role of heterologous expression in the discovery and engineered production of bioactive polyketides and non-ribosomal peptides from bacteria. We contextualize these recent advancements by identifying the various groups of bacteria that produce these compounds and by comparing the heterologous hosts that researchers are using to express BGCs of interest. Our comparisons will focus primarily on what synthetic biology tools are available for individual hosts and to what extent are researchers taking advantage of available tools to modify heterologous hosts and/or BGCs for improved production.

2.3 Motivations for the heterologous expression of BGCs

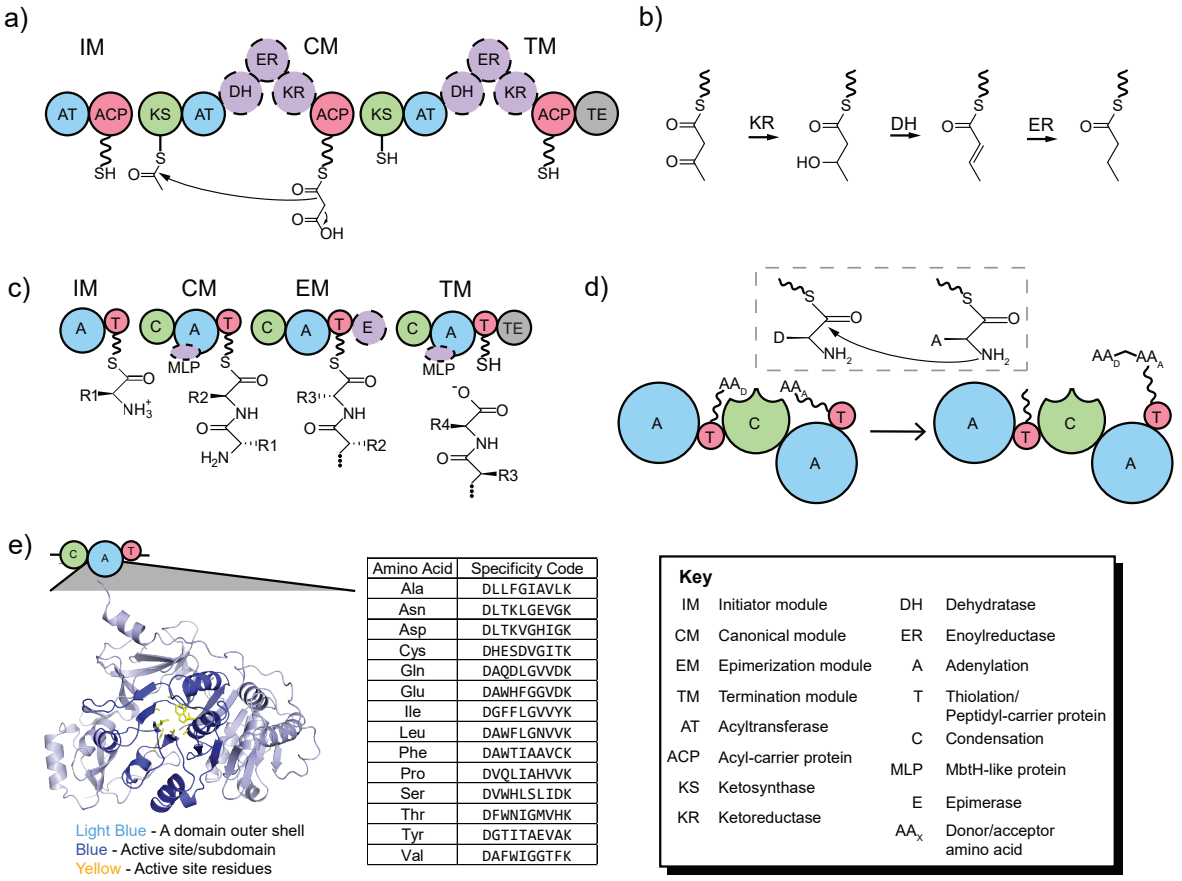
Polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) can generate extraordinary chemical diversity by incorporating a wide variety of substrates in the initiation and elongation biosynthetic steps [40, 41, 42]. Additional tailoring catalytic domains, deviations from the canonical assembly-line enzymology, and hybrid PKS–NRPS enzymes further increase the complexity of these enzymes [43, 44]. For more information on PKS and NRPS enzymology, Fischbach and Walsh provide an in-depth review on the relevant enzymatic domains and mechanisms [45]. We provide our own description of these components below, incorporating more recent findings.

2.3.1 Polyketide synthase and nonribosomal peptide synthetase enzymology

PKSs and NRPSs are similar in that they are usually modular enzymes with multiple catalytic domains encoded in a single peptide. Each enzyme class has highly conserved domains, and their sequential organization and specificity determine the metabolite's structure. Evolutionary processes have achieved outstanding structural diversity through the recombination and mutation of individual domains [46, 47]. Based on previous biochemical studies and bioinformatics, it is possible to predict the structure of polyketides and nonribosomal peptides using genetic information [48, 49].

Polyketide synthesis commonly begins with the condensation of acetyl-CoA with malonyl-CoA, and successive steps incorporate more malonyl-CoA building blocks. PKSs often deviate from these substrates, with propionyl-CoA commonly replacing acetyl-CoA as the starting unit [50]. Other starter units include benzoyl-CoA and fatty acyl-ACP [51, 52]. Methylmalonyl, ethylmalonyl, and hydroxymalonyl-CoA are common alternative extender units in polyketide biosynthesis [41].

Type I PKSs have a canonical assembly line structure, where each module is responsible for the activation and condensation of a single extender unit, and each module contains up to five catalytic domains (Figure 2.1a). The acyltransferase (AT) domain



Amino Acid	Specificity Code
Ala	DLLFGIAVLK
Asn	DLTKLGEVGK
Asp	DLTKVGHIGK
Cys	DHESDVGITK
Gln	DAQDLGVVDK
Glu	DAWHFGGVDK
Ile	DGFFLGVVYK
Leu	DAWFLGNVVK
Phe	DAWTIAAVCK
Pro	DVQLIAHVVK
Ser	DVWHLSLIDK
Thr	DFWNIGMVHK
Tyr	DGTITAEVAK
Val	DAFWIGGTFK

Figure 2.1: **Modular structure of polyketide synthases and nonribosomal peptide synthetases.** **a)** Domain architecture of initiator modules, canonical modules, and termination modules of polyketide synthesis. The mechanism for ketoacyl synthesis by the KS domain in the canonical module is shown. The KS domain transfers an acetyl group from the ACP domain from the previous module and catalyzes a Claisen condensation with malonyl-ACP of its own module. **b)** The three types of reduction reactions in PKS enzymology. The KR domain synthesizes a β-hydroxyacyl-ACP from a β-ketoyacyl-ACP, followed by α-β-enoyl-ACP formation by the DH domain, and lastly an acyl-ACP formed by the ER domain. **c)** Domain architecture of initiator modules, canonical modules, epimerization modules, and termination modules in nonribosomal peptide synthesis. The structure and stereochemistry of the nascent peptide is shown. An MLP is shown interacting with some adenylation domains, which do not always require an MLP for full activity. **d)** Peptide bond formation catalyzed by the C domain. The amine group on the acceptor amino acid acts as the nucleophile and attacks the carboxyl carbon of the donor amino acid. **e)** Identity of active site residues in NRPS A domains. The active site is part of a compact subdomain that contains the specificity code residues. The consensus codes for 14 amino acids are shown. The first and last residues, D and K, are invariable and essential for A domain activity.

catalyzes the covalent linkage of a starter or extender unit to an acyl carrier protein or thiolation (ACP or T) domain within the same module. The subsrate binds directly to a phosphopantetheine arm that is attached to the ACP domain by a phosphopantetheinyl transferase (PPTase) [53]. AT domains can be encoded within the same peptide as the rest of the catalytic domains (*cis*-AT) or as a standalone protein that may esterify ACP domains in multiple modules (*trans*-AT) [54]. The AT domain is commonly considered the "gatekeeping" domain of polyketide biosynthesis. Crystal structures have revealed the binding-site residues responsible for AT specificity for malonyl-CoA or methylmalonyl-CoA [55, 56, 57].

Next, a ketosynthase (KS) domain catalyzes a Claisen condensation between the extender unit and an acyl unit from a previous module, resulting in a β-ketoacyl unit. The β carbon on this substrate can be reduced to varying degrees by a ketoreductase (KR) domain, dehydratase (DH) domain, and finally an enoylreductase (ER) domain, in that order (Figure 2.1b). The extent of reduction is determined by the omission or inactivation of the responsible domains within the module. In modules with only KR activity, the KR domain determines the stereochemistry of the β-hydroxy carbon. In cases where the extender unit is a methylmalonyl moiety or other chiral malonyl derivative, the stereochemistry of the α carbon of the condensation product is also controlled by the KR domain, including some KR domains that have no reductase activity and act solely as an epimerase [58, 59].

Starter modules contain fewer catalytic domains because they only serve to activate an acyl-CoA or acyl-ACP. A common structure is a loading AT domain or CoA ligase domain followed by an ACP (Figure 2.1a). Alternatively, a three-domain loading module contains a modified KS domain (in addition to AT and ACP domains) that can activate malonyl-CoA and its structural derivatives and catalyze a decarboxylation reaction after activation [60]. A third type of starter module has an adenylation domain that esterifies free carboxylic acids onto the ACP, similar to NRPS loading modules [61]. The termination module contains a thioesterase (TE) domain, which cleaves the thioester bond between the polyketide and ACP (Figure 2.1a). It is common for this domain to cat-

alyze a cyclization reaction between the free carboxylic acid and a hydroxyl group in the polyketide.

Even within the type I class of PKSs, modules can deviate significantly from this sequence of reactions, particularly for *trans*-AT PKSs [62]. Other PKSs do not fall into the assembly line format of the type I class. The domains in type II PKSs are encoded as standalone enzymes and form a minimal PKS containing two ketosynthases and an ACP that iteratively incorporates malonyl-CoA, and supplementary domains can further modify the polyketide's structure. This class of PKSs synthesizes many aromatic polyketides, including antibiotics like tetracycline [63]. Type III PKSs also catalyze iterative reactions, but exist only as a homodimer of KS monomers. Despite this relatively simple enzyme architecture, type III PKSs are capable of a diverse set of elongation and cyclization reactions [64].

There is a similar level of diversity in NRPS enzymology, but our discussion focuses primarily on enzymes with an assembly line format. In canonical nonribosomal peptide synthesis, each module activates an amino acid and catalyzes peptide bond formation with amino acids from adjacent modules, similar to ribosomal protein synthesis. These enzymes are capable of extraordinary structural diversity because they are not limited to proteinogenic amino acids. Researchers have identified up to 800 non-proteinogenic amino acids, and NRPSs can also incorporate aryl acids [65].

A typical NRPS module contains three domains, an adenylation (A) domain, condensation (C) domain, and a peptidyl-carrier protein or thiolation (PCP or T) domain (Figure 2.1c). The A domain catalyzes the ATP-dependent activation of an L-amino acid and converts into an amino acyl-AMP. This is considered the gatekeeping reaction in NRPS enzymology. Based on the structure of an A domain bound to its amino acid substrate, the residues conferring amino acid specificity were determined [66]. Researchers quickly discovered that these residues closely mapped to the enzymes's substrate, and these residues were eventually called the "specificity code" (Figure 2.1e). There are 10 residues that interact with the amino acid substrate, 8 of which are indicative of the domain's specificity. There are several prediction algorithms that rely primarily on these

residues to predict amino acid specificity, which have been incorporated into natural product discovery platforms [67, 48].

Adenylation domains often require interactions with an MbtH-like protein or another activating protein, which is encoded separately from the NRPS [68, 69]. After A-domain activation, the activated substrate is then attached to the phosphopantetheine moiety of the PCP domain. The C domain catalyzes peptide bond formation between this amino acid and the amino acid bound to the PCP domain in the previous module (Figure 2.1d). Adenylation domains commonly activate L-amino acids, modules sometimes contain an epimerization (E) domain that enables the incorporation of D-amino acids. The amino acid is epimerized by the E domain after peptide bond formation by the C domain [70]. Alternatively, hybrid C/E domains that yield similar results have also been reported, and there are a few examples of A domains that recognize D-amino acids directly [71, 72].

Starter modules in NRPSs are similar to those of PKSs in that they have fewer modules, usually an A domain followed by a PCP domain (Figure 2.1c). These modules activate free acids, but alternative starter modules with starter condensation domains or CoA ligase domains can activate CoA-bound metabolites [73, 74]. Like polyketide biosynthesis, termination modules contain a TE domain that releases the nascent peptide from the NRPS. PKSs and NRPSs sometimes contain a type II TE, either as an attached domain or as a standalone enzyme, that regenerates misprimed carrier protein domains, preventing stalling of the assembly line [75].

NRPSs that deviate from the assembly line architecture have also been described. Type II NRPSs feature standalone enzymes responsible for amino acid activation and peptide bond formation [76]. These pathways incorporate enzymatic steps not usually seen in modular NRPSs, such as dehydrogenases and halogenases. Analogous to type II PKSs, there are some iterative NRPS modules, although their domain architecture is not as distinct from linear NRPSs, iterative NRPSs are capable of oligomerization based on the specificity of the TE domain [77].

2.3.2 Genetic and protein engineering of PKS and NRPS components

There is considerable interest in engineering PKSs and NRPSs to alter substrate specificity, increasing their potential as a source of novel natural products [78, 79]. However, these large enzymes pose several challenges to manipulating the synthesis of their products. Genetic manipulation of the BGCs is not trivial because clusters often range from 20 to 100 kb in size (with some >100 kb) and contain repetitive DNA sequences due to the modular structure of their encoded megaenzymes [80].

Many techniques designed for specifically cloning large BGCs encoding PKSs and NRPSs are available, allowing researchers to reliably transfer BGCs to heterologous hosts. Traditionally, researchers have cloned BGCs of interest by constructing libraries from genomic DNA and screening for the correct clone with PCR [35]. Targeted methods for capturing BGCs from genomic DNA are a more efficient option. These methods take advantage of restriction enzymes or programmable nucleases that can cleave desired DNA segments from the chromosome in vitro [81, 82]. Researchers can then use homologous recombination and/or ligation cloning techniques to directly capture BGCs of interest in an expression vector [83]. RecET direct cloning and yeast transformation-associated recombination (TAR) cloning are commonly used to clone large BGCs in *E. coli* and Saccharomyces cerevisiae, respectively [84, 85]. DNA assembly methods have also been developed for constructing large BGCs from multiple DNA fragments, which allows researchers to refactor BGCs of interest. Large DNA molecules can be assembled in vitro or in vivo using Golden Gate cloning, single-stranded annealing, and homologous recombination in yeast [86, 87, 88]. Methods for modifying plasmids containing large BGCs have recently been developed for engineering modular enzymes through point mutations and domain swapping [89, 90].

With up to five domains within a canonical PKS module, there are many options for modifying polyketide biosynthesis through protein engineering. The 6-deoxyerythronolide synthase (DEBS) was the model PKS for the earliest attempts [91, 50]. Directly inspired by the modular structure of PKSs, researchers performed whole module swaps between

homologous enzymes to synthesize erythromycin derivatives [92, 93]. In a later study, combinatorial synthesis of PKSs containing loading modules follwed by two extender modules yielded biosynthesis of 71 unique triketide lactones from a total of 154 DNA constructs [79]. While this result showed great promise for PKS engineering via module substitution, this systematic approach revealed some mechanisms of failure, such as non-optimal linker interactions between adjacent modules [94]. Chimeric PKSs also consistently produce smaller amounts of product due to poor protein stability and protein-protein interactions between non-cognate modules.

Specifically swapping loading modules is an attractive engineering strategy due to their broad substrate specificity [95]. These modules also enable semi-synthetic strategies for polyketide derivitization because the can incoporate aromatic or alkyne substrates, resulting in functionalized polyketides [96, 97, 98]. When constructing PKS chimeras, researchers usually use defined module or domain boundaries, but recent reports have challenged these seemingly intuitive positions (Figure 2.2a). Chemler *et al.* discovered that hybrid PKSs with unconventional module boundaries can function [99]. Notably, novel polyketides were synthesized from constructs with module boundaries falling within AT and KS domains. More recently, novel module boundaries at the KS-AT di-domain boundary, instead of the traditional ACP-KS boundary, improved the activity of chimeric PKSs *in vitro* [100].

Researchers have also been able to make polyketide modifications by swapping individual domains (Figure 2.2b). AT domain swaps can enable the incorporation of different starter and extender units, and traditionally these modifications were used to switch between malonyl-CoA, methylmalonyl-CoA, and ethylmalonyl-CoA [91, 101]. Alternatively, inserting or deleting reductive domains can alter the oxidation state of carbon centers throughout polyketide structures [93, 102, 103]. KR domain swaps or mutagenesis are also an opportunity to modify the stereochemistry of the α- and β-carbons of the nascent polyketide [104, 105, 106].

There is considerable interest in manipulating polyketide biosynthesis by engineering AT domains. In one of the earliest examples of PKS engineering, Oliynyk *et al.*

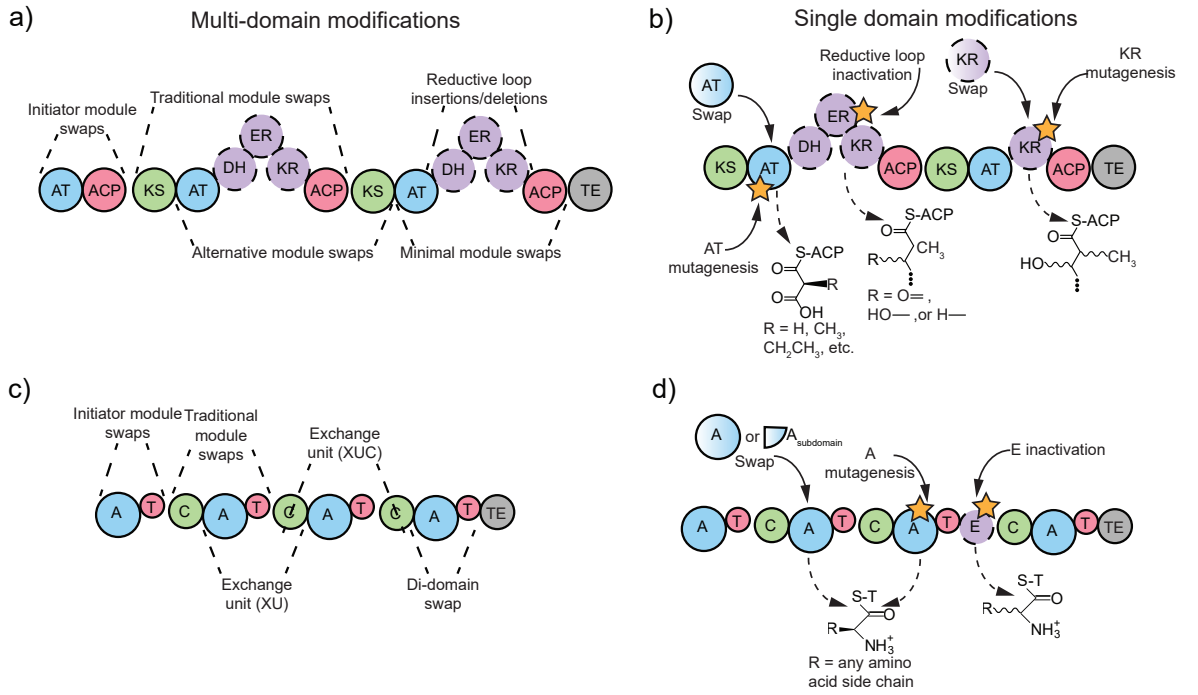


Figure 2.2: **Engineering strategies in modular PKSs and NRPSs. a)** Multi-domain substitutions in PKSs. Several module boundaries have been defined in PKS engineering. **b)** Single domain modifications in PKS engineering. AT domain swaps and mutagenesis enable activation of alternative extender units. Inactivating individual domains in the reductive loops decreases the level of reduction of the nascent polyketide. KR domain swaps and mutations alters the stereochemistry of the α- and β-carbons of the extender unit. **c)** Multi-domain substitutions in NRPSs. Several module boundaries have been defined in NRPS engineering. **d)** Single domain modifications in NRPSs. A domain swaps and substitutions alter the activated amino acid substrate. E domain inactivation switches the amino acid stereochemistry from the D-isomer to the L-isomer.

successfully synthesized novel triketide lactones lacking a methyl group by replacing a methylmalonyl-CoA-specific AT domain with a malonyl-CoA-specific AT domain [91]. In later studies, researchers discovered that most PKSs with heterologous AT domains have significantly reduced productivity [107]. Biochemical analyses of some of these hybrid PKSs suggested that this reduced activity is due to impaired elongation by the KS and ACP domains within the module of the swapped AT domain [108]. However, a recent study found that the bottleneck imposed by KS domains in chimeric PKSs is alleviated by a few point mutations [109]. Despite these issues, AT domain swaps are still a common PKS engineering strategy today. Alternatively, *trans*-AT domains have also been explored for installing alternative extender units into polyketides. After inactivating the native *cis*-AT, researchers successfully synthesized polyketide derivatives after introducing a heterologous *trans*-AT [110]. However, more information on how these enzymes

interact with ACP domains is needed before this becomes a universal engineering strategy [111, 112].

Issues with domain swaps motivated researchers to investigate factors conferring AT domain specificity. It was immediately clear that AT domains that activate malonyl-CoA contained a HAFH sequence motif, and methylmalonyl-CoA domains contained a YASH motif [113, 114]. However, these motifs were not the only residues controlling substrate specificity, as motif mutations usually yielded promiscuous AT domains. In some cases, this promiscuity is advantageous and has enabled the incorporation of unsaturated extender units by mutated AT domains, resulting in alkyne-tagged polyketides [115]. Other sequence motifs have been identified from molecular simulations and bioinformatics, and these findings have enabled the construction of modified AT domains with specificity shifted towards the desired substrate [116]. Site-directed mutatagenesis of AT domains has great promise for PKS engineering because this strategy avoids issues with protein stability that arise in domain-swapped PKSs.

Protein engineering efforts in NRPS systems have seen similar challenges. Module and domain swaps often result in unstable proteins, and switching domain specificity through site-directed mutagenesis has been limited to structurally similar substrates. Substitution strategies showed early promise when researchers successfully synthesized peptide derivatives from A-T di-domain swaps and fusing whole modules (C-A-T) together (Figure 2.2c) [117, 118]. While the A domain was initially considered the sole substrate-determining domain in NRPS enzymology, *in vitro* studies did suggest C domains can be highly specific for its cognate amino acid, suggesting that substituting C-A or C-A-T units are more likely to be successful than doing the same with A and A-T units [119, 120]. For example, A domain substitutions in a L-threonine-specific module in the pyoverdine synthetase of *P. aeruginosa* only yielded wild-type pyoverdine [121]. Even non-threonine-specific A domains did not alter the final product. However, substituting the entire C-A di-domain region resulted in pyoverdine derivatives with L-lysine and L-serine substitutions. While these results do suggest that C domains have some role in determining substrate specificity, the loss of important C-A domain-domain inter-

actions could also be the mechanism of failure for A domain substitutions [122]. Indeed, one study found that mutations throughout the A domain improved activity of hybrid NRPSs with a heterologous A domain [123]. These mutations were likely important for protein stability and protein-protein interactions with adjacent domains.

Considerable progress in NRPS engineering has been achieved by researchers thinking beyond the canonical C-A-T module boundaries. A remarkable example defined exchange units (XU) around A-T-C tri-domain regions that yielded functional chimeric NRPSs (Figure 2.2c) [124]. The authors hypothesized that the C-terminus of the C-A domain linker was the optimal exchange boundary because there were no apparent C-A domain-domain interactions disrupted. One limitation of this strategy was that the C domain had to have the same substrate specificity as the downstream exchange unit, again suggesting that C domains may limit the scope of A domain substitutions in NRPS engineering. The same research group published a potential solution by defining yet another exchange unit (XUC), with the boundaries falling within the C domain (Figure 2.2c) [125]. These boundaries yielded hybrid NRPSs capable of producing peptide derivatives in similar amounts as the previous publication, and C domain specificity was not a concern in these substitutions because the C subdomains, donor and acceptor, did not interact with A domains of a different substrate specificity.

A recent report on NRPS engineering challenged the dogma on the C domain's role in A domain substitutions. Calcott *et al.* demonstrated that A domain substitutions were more successful if the C-A boundary was shifted towards the N-terminus of the C-A linker [126]. They found that this strategy produced more functional chimeric NRPSs than C-A swaps and A domain swaps from earlier studies and that it was possible to construct functional swaps with enzymes that shared relatively low sequence homology. Interestingly, the C-A boundary reported by Bozhuyuk *et al.* for swapping A-T-C exchange units, a strategy that had issues with C-domain specificity in hybrid NRPSs, was on the opposite terminus of the C-A linker [124].

Another type of domain substitution strategy was inspired by recombination events that have occured naturally in NRPS evolution [127]. *In vitro* experiments demonstrated

that swapping a compact subdomain containing 9 of the 10 active site residues results in a chimeric A domain with a matching substrate specificity to the donor A domain (Figure 2.2d) [128, 78]. This strategy allows a complete switch in A domain specificity without compromising protein-protein interactions between adjacent domains. Further *in vitro* studies observed that NRPSs with chimeric A domains may be limited by C domain activity, but as observed in earlier studies, directed evolution can improve total activity and tighten substrate specificity for the chimeras [129, 130, 131].

Lastly, most other NRPS engineering efforts have focused on mutations in the A domain active site. This strategy has primarily focused on the 10 active site residues that make up the remarkable "specificity code" [66]. Even though naturally-occuring bacterial A domains rarely break the specificity code rules, swapping the code residues is not enough to switch substrate specificity [67]. The first example of successful specificity code mutations resulted in specificity swaps from L-glutamate to L-glutamine and from L-aspartate to L-asparagine, with both changes requiring only one mutation [132]. Almost twenty years later, Kaniusaite *et al.* reported engineered A domains that activated 4-hydroxyphenylglycine instead of the original and structurally similar substrate, 3,5-dihydroxyphenylglycine [133]. These mutants only required 1-2 key mutations in the active site for altering subtrate specificity.

Directed evolution approaches on the specificity code have successfully altered substrate specificity in some A domains, but the variants are usually promiscuous and still activate the original substrate [134, 20]. Some researchers have embraced this feature after discovering a single mutation in a L-tyrosine-specific A domain that enabled activation of functionalized tyrosine derivatives [135]. This discovery recently enabled the production of peptides modified with alkyne, halide, and benzoyl moieties, which are compatible with well-known "click" reactions [136].

Advancements in PKS engineering have closely mirrored those in NRPS engineering. Considerable progress has been made in determining optimal module and domain boundaries when constructing homology substitutions in assembly line enzymes. There is enough knowledge on substrate specificity of gatekeeping domains to reliably predict

natural product structures from sequence information alone, but a universal strategy for changing specificity has not been described. In order to truly realize the "lego-ization" of polyketide and nonribosomal peptide synthesis, researchers will need to identify strategies for maintaining inter-domain interactions and substrate channeling in chimeric enzymes and expand our knowledge of substrate recognition beyond sequence motifs.

2.3.3 Host-specific challenges in natural product discovery and production

PKS and NRPS activity is dependent on accessory proteins that are present in the host. PKSs and NRPSs are only functional in the presence of a compatible phosphopantetheinyl transferase (PPTase), which is responsible for incorporating an essential post-translational modification to carrier protein domains [137]. In addition, NRPS adenylation domains, which are responsible for activating amino acid substrates for incorporation into the peptide product, often require the presence of an MbtH-like protein (MLP) in order to be fully functional [68]. MbtH is a protein encoded in the BGC for synthesis of the mycobactin siderophore from *Mycobacterium tuberculosis* and forms the basis for the MLP superfamily [138]. The exact function of MLPs is still an active area of research, but researchers have observed that they bind to their cognate adenylation domains and improve their solubility and substrate affinity [139, 140]. These auxiliary proteins can be located outside the gene clusters encoding putative NRPSs/PKSs thereby imparting a requirement to co-express native or promiscuous variants in heterologous hosts.

A major bottleneck to discovering novel natural products is the fact that the overwhelming majority of genetic material in environmental samples originates from strains that have not been cultured in the laboratory [141]. In order to retrieve BGCs from these strains, researchers have turned to metagenomic libraries and/or commercial DNA synthesis of refactored BGCs prior to introducing them into a heterologous host (Fig. 2.3) [39, 142]. Even at $0.10 per base pair, the cost of synthesizing large clusters is out of reach for most academic laboratories and a limitation to industrial efforts. Similarly, the cost of screening metagenomic libraries in search of clones harboring the desired cluster

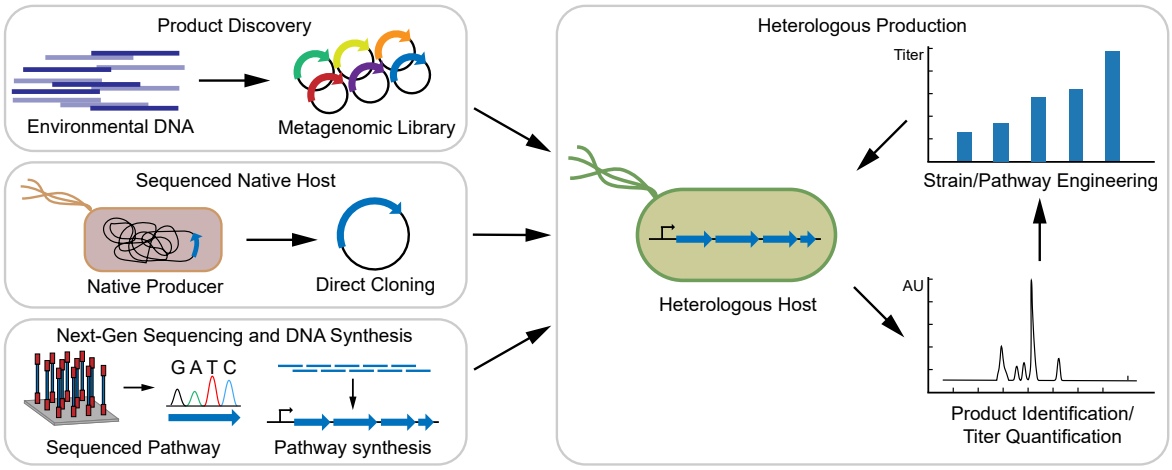


Figure 2.3: **Strategies for discovering novel natural products using heterologous expression of BGCs.** Candidate BGCs can be sourced from environmental DNA, genomic DNA from the native host, or *de novo* DNA synthesis. The ideal heterologous host has characteristics that allow for rapid strain engineering and robust cell growth and production of natural products.

can be equally large and depend on either PCR, sequencing, or functional assays [51]. Functional screens include bacterial and/or fungal growth inhibition, detection of pigmented compounds, biosensor activation, and high-throughput analytical chemistry (e.g. HPLC) [142, 143, 144]. Early screenings of metagenomic libraries were mostly random, but sequencing environmental DNA samples has allowed researchers to identify libraries with BGCs related to those of secondary metabolites of interest [145]. Heterologous hosts chosen to express BGCs from metagenomic libraries can influence which natural products and corresponding BGCs are discovered because some promoters from the library may be poorly expressed in various hosts [146]. However, researchers have found that introducing foreign transcriptional regulators to heterologous hosts can improve expression from these promoters and therefore increase natural product production [147, 148].

Another challenge associated with isolating bioactive natural products from native hosts that are difficult to cultivate is producing them in amounts sufficient for further characterization and ultimately clinical trials [149, 150]. While many natural products can be synthesized chemically, their structural complexity requires the use of many chemical steps that result in poor overall yield [151]. Biological production can simplify this process by reducing multiple chemical reactions to a single bioreactor, if a suitable host can be developed. Even when native bacteria containing a BGC of interest can be cul-

tivated, other issues can prevent production of secondary metabolites at levels required for subsequent functional testing. Native hosts often grow slowly and have strict nutrient requirements, limiting the ability to scale up production from native hosts. Many clusters are also not expressed in normal laboratory conditions and the conditions required for activating the expression of many BGCs remain unknown [152]. Instead, researchers are using synthetic biology strategies for activating expression of these clusters in the native host if tools are available. That said, genetic tools are less likely to be available for species isolated from the environment, making transfer of the BGC into a heterologous host a necessary strategy [153]. A heterologous host provides the opportunity to use a production strain that has characterized genetic tools and is more suited for industrial-scale production.

Synthetic biology provides the necessary tools to manipulate BGCs and gain access to previously unknown natural products. Useful genetic tools and methods for expressing natural product clusters are characterized transformation protocols, promoters, ribosome binding sites (RBSs), terminators, replicative or integrative vectors for introducing heterologous genes, and methods for editing the genome [154]. To date, most advancements in bacterial synthetic biology have been applied to *E. coli*, but recently there has been a shift towards the development of genetic tools for non-model bacteria, including those from taxonomic groups known to synthesize polyketides and non-ribosomal peptides [155, 156]. Characterized promoters, RBSs, and terminators enable the predictable expression of heterologous BGCs. Methods for engineering the chromosome allow researchers to modify endogenous pathways. This strategy requires markerless and scarless methods that use a counterselection system. Earlier methods relied on counterselection genes that conferred sensitivity to a metabolite, such as *sacB* and sucrose or *upp* and 5-fluorouracil [157, 158]. These methods require multiple steps to generate the desired strain and may require modifications to the genome a priori for the counterselection to be functional, so researchers have turned to counterselections based on CRISPR-associated nucleases [159, 160]. Scarless and markerless genome editing, including CRISPR-based methods, enables the generation of multiple mutations to the chromosome, allowing re-

searchers to optimize multiple metabolic pathways in heterologous hosts for improved polyketide and nonribosomal peptide production [161, 162, 163].

2.4 Ideal traits of heterologous hosts for producing polyketides and non-ribosomal peptides

The ideal heterologous host for producing polyketides and non-ribosomal peptides should have a growth rate suitable for industrial fermentations, a characterized synthetic biology toolbox, and a sufficient supply of precursor metabolites (Fig. 2.4). Important genomic features include GC content and codon usage similar to the native host, unless the BGC can be codon optimized for the heterologous host. A lack of native BGCs reduces competition for limited metabolite supplies and simplifies downstream purifications [164]. Alternatively, competing BGCs can be deleted from the chromosome using genome editing methods described in this review. Particularly in the case of antibiotics production, heterologous hosts should also be resistant to the product of interest and have the ability to secrete the product from the cell [165]. Bacteria that are capable of unicellular growth are more suited to growth in liquid cultures compared to mycelial cultures that can increase the media viscosity and lower oxygen transfer rates [166]. Ideal hosts also have the metabolic flexibility to convert common feedstocks to the product of interest, in contrast to many native hosts that have atypical and expensive nutrient requirements.

2.4.1 Domestication of *P. putida* with synthetic biology

As synthetic biology advancements have shifted focus towards non-model organisms, the soil bacterium, *P. putida* has received considerable attention. Tool development for this species is rapidly expanding and there are multiple options for each type of tool. The explosion in *P. putida* synthetic biology even prompted some researchers to write an in-depth review that compares available tools and calls for researchers to apply them towards metabolic engineering goals [167]. The desire to engineer *P. putida* for producing chemicals of interest arises from its robust growth at high cell densities, resistance

to solvents and other industrial stresses, and ability to catabolize non-traditional carbon sources, such as lignin-derived monomers and biomass by-products [168, 169, 170]. *P. putida* has a lot of potential as a production host for natural products, as Pseudomonads commonly synthesize their own [171]. *P. putida* has a relatively high GC content (61.5%) in its chromosome, similar to that of common natural product hosts, such as Actinomycetes and myxobacteria [172, 173]. This microbe is also capable of functionally expressing assembly line enzymes because it naturally secretes a siderophore called pyoverdine, which is synthesized by NRPSs encoded in its genome [73]. Lastly, there are considerably more genetic tools available for *P. putida* and it has a faster growth rate in laboratory conditions compared to other heterologous hosts. These characteristics provide a great opportunity for *P. putida* to be developed specifically for producing natural products from other bacteria.

The minimal set of bacterial genetic tools is multiple promoters, RBS sequences, and plasmids with replicative origins. Like other Gram-negative bacteria, plasmids used in *P. putida* have broad-host range (BHR) origins that replicate plasmid DNA independently of host replication processes [174]. The most commonly used BHR origins for *P. putida* are the RK2, BBR1, and RSF1010 origins [175, 176, 177]. A fourth origin that has been used primarily for experiments in *P. aeruginosa*, RO1600, has recently been used successfully in *P. putida* as well [178].

Transcriptional control in *P. putida* is achievable with constitutive and inducible promoters. Libraries of constitutive promoters have been created through saturation mutagenesis, and the Anderson promoter library also predictably expresses genes in *P. putida* [179, 180]. Several inducible promoters discovered from Pseudomonads have been characterized and shown robust gene expression in *P. putida*, including P_m induced with 3-methylbenzoate [181, 182], P_{sal} induced with salicylate [181, 182], and P_{alkB} induced with dicyclopropylketone [181, 183]. Researchers have also characterized promoters from non-pseudomonads in *P. putida*, including rhamnose (P_{rhaB}; [181, 184]) and arabinose (P_{araB}; [181, 185]), which both exhibit tight repression in the absence of the inducer. The *lac* family of promoters have also been used for heterologous expression in *P. putida*,

although the fold-induction for some systems is relatively low compared to their performance in *E. coli* [186]. While these promoters are used often, there has been little effort towards optimizing inducible promoter systems for *P. putida*, particularly in reducing basal expression or increasing the fold-induction. One recent study demonstrated the benefits of this research with their work on P_m in *P. putida* [178]. By integrating the regulatory protein, XylS, onto the chromosome with a constitutive promoter, the authors reduced leaky expression and increased the fold-induction 170-fold. Expression from inducible promoters could benefit from other strategies for optimization, such as protein degradation tags and small RNA repression [187]. CRISPR interference techniques could be use more broadly to attenuate chromosomal gene expression as well [188, 189].

Investigations into the next step in gene expression, translation, has not been a major focus in *P. putida* tools. The field has progressed towards most researchers using the RBS calculator developed by the Salis research group to design RBS sequences, but no publication has experimentally verified these predictions in *P. putida* [190, 191, 192]. An RBS library was recently characterized in *P. putida*, where researchers mutagenized the Shine-Dalgarno sequence to determine its effect on heterologous gene expression [193]. While researchers across the field have successfully translated diverse proteins in *P. putida*, a greater understanding of RBS design would improve synthetic operon design and enable the construction of more complex genetic circuits.

Genome editing is an essential tool in synthetic biology and metabolic engineering. Genomic modifications are required for shaping cellular metabolism, and chromosomal integrations are often more stable than replicative plasmids for heterologous gene expression. Methods based on homologous recombination have been a staple in *P. putida* genome editing. With only one plasmid containing homology arms and a counterselection gene (*upp*; 5-fluorouracil or *sacB*; sucrose), researchers have generated large chromosomal deletions and heterologous gene insertions [158, 194]. Sucrose-based counterselections are preferred in *P. putida* because anti-metabolites like 5-fluorouracil are known to be mutagenic in bacteria [195]. However, recent genome editing methods have relied on CRISPR-associated (Cas) nucleases for counterselections. Several labs have developed

Cas-assisted ssDNA recombineering methods, including Cas9 and Cas12a [196, 197, 198]. Cas-assisted homologous recombination with dsDNA has only been reported once (outside of our lab), but this technique used an I-SceI counterselection in addition to Cas9 [199]. Methods with ssDNA recombineering have enabled the generation of chromosomal point mutations with and without Cas-based counterselections, offering more versatility in the *P. putida* genome editing toolkit [200, 193].

Researchers have developed methods specifically for integrating exogenous DNA in *P. putida* as well. Tn7 transposition is a common technique for integrating DNA downstream of the *glmS* gene in Gram-negative bacteria [201]. Phage integrases can be used in *P. putida* to integrate DNA into locations containing a "landing pad" introduced beforehand by homologous recombination [179, 202]. With such a multitude of genome editing tools, virtually any chromosomal edit is possible in *P. putida*. However, these genome editing techniques have not been adapted for multiplexed or high-throughput applications, which is currently limiting progress in metabolic engineering projects with *P. putida*.

Metabolic engineering in *P. putida* has focused primarily on lignin valorization into multiple products, including poly(hydroxyalkanoates), adipic acid, and muconic acid [203, 204, 205]. Natural products engineering in *P. putida* has mostly consisted of integrating various biosynthetic gene clusters, but in one case the authors also introduced genes to synthesise methylmalonyl-CoA for polyketide synthesis [206]. The most remarkable metabolic engineering examples in the literature take advantage of omic analyses to determine optimization strategies. This approach has allowed researchers to quantify pathway enzymes *in vivo* and identify catabolic pathways reducing product titers [207, 208].

Any major improvements in the state of *P. putida* metabolic engineering will require systematic approaches to designing DNA constructs and engineered strains. An exceptional study showcasing these elements reported a muconate biosensor optimized for *P. putida* [209]. The authors constructed a library of promoter and RBS mutations and used fluorescence-activated cell sorting (FACS) to isolate a mutant with high expression when cells grew in the presence of muconate precusors and minimal expression without. This

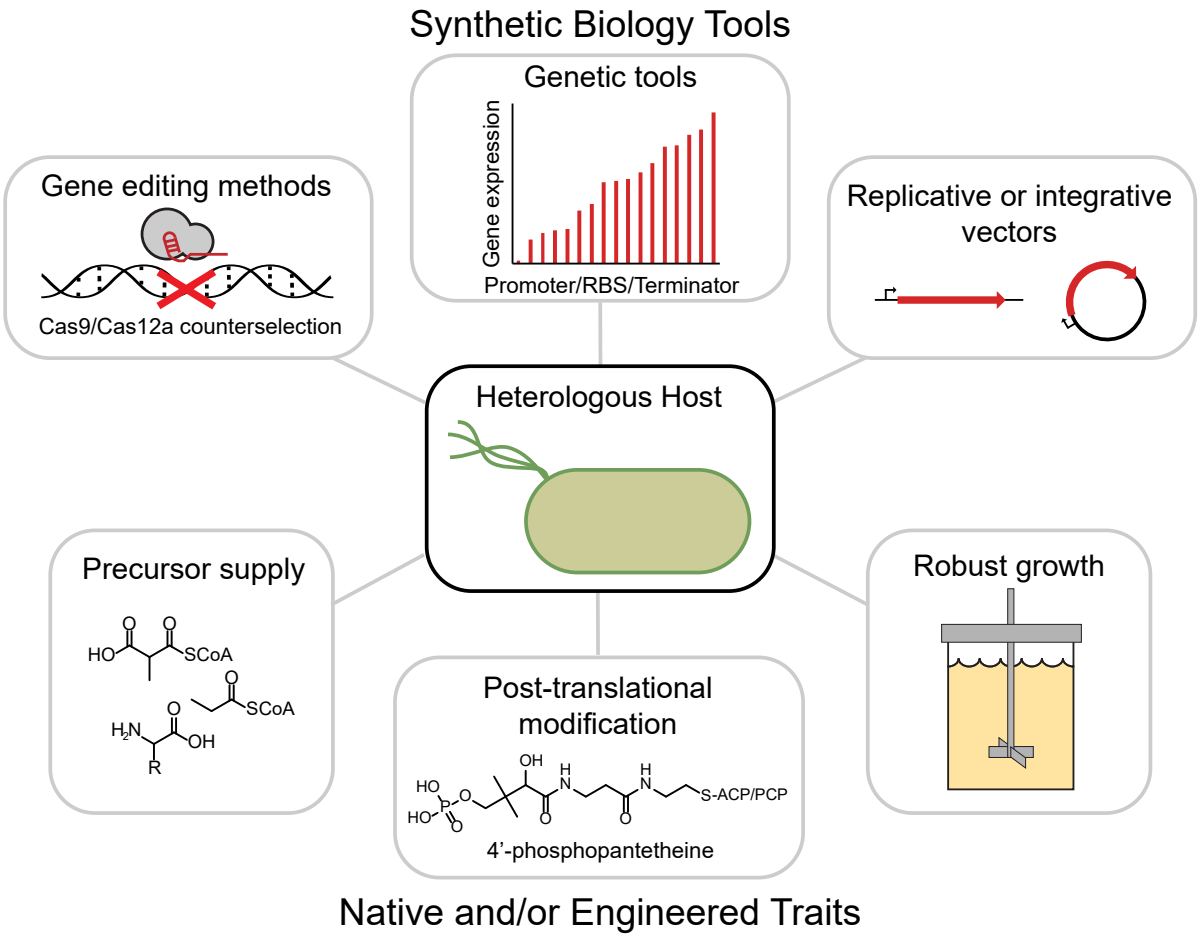


Figure 2.4: **Synthetic biology tools and physiological traits required for production of polyketides and non-ribosomal peptides from a heterologous host.** Necessary synthetic biology tools include genome editing methods, DNA parts for controlling gene expression, and replicative or integrative vectors. Physiological traits include synthesis of required metabolite precursors, post-translational modification of acyl- or peptidyl-carrier proteins, and robust cell growth.

biosensor was then used to screen a library of strains created from adaptive laboratory evolution (ALE) for improved muconate production, revealing key genomic mutations that informed the next steps in strain engineering.

2.4.2 Comparison of bacterial heterologous hosts for natural product investigations

There are multiple bacterial hosts that researchers have used for studying natural products. Table 2.1 provides a comparison of characteristics for species that have successfully produced heterologous polyketides and non-ribosomal peptides: *Streptomyces lividans*, *Streptomyces albus*, *Streptomyces venezuelae*, *E. coli*, *Myxococcus xanthus*, *Pseudomonas*

Table 2.1: Comparison of phylogenetically diverse heterologous hosts for polyketide and non-ribosomal peptide production. –, not reported

Strain	Phylum or class	Genome Size (Mbp)	GC content (%)	Doubling time	Oxygen requirement	Recombinant DNA introduction (max size)
S. lividans TK24	Actinobacteria	8.34	72.24	4.2 h	Obligate aerobe	Electroporation, conjugation (90 kb) [210]
S. albus J1074[a]	Actinobacteria	6.84	73.3	-	Obligate aerobe	Electroporation, conjugation (120 kb) [211]
S. venezuelae ATCC15439	Actinobacteria	9.05	71.7	1 h	Obligate aerobe	Electrporation (32 kb) [212], conjugation
E. coli BL21(DE3)	Gamma-proteobacteria	4.56	50.8	20 min	Facultative anaerobe	Chemical transformation, electroporation, conjugation (>100 kb)
P. putida KT2440	Gamma-proteobacteria	6.18	61.5	45 min	Obligate aerobe	Electroporation (30 kb), conjugation (60 kb) [213]
M. xanthus DK1622	Delta-proteobacteria	9.14	68.9	4-5 h	Obligate aerobe	Electroporation (60 kb) [213]
B. subtilis 168	Firmicutes	4.22	43.5	20 min	Facultative anaerobe	Electroporation, natural transformation
Anabaena sp. PCC 7120	Cyanobacteria	6.41	41.3	14 h	Photoautotroph	Conjugation (42 kb) [214]
Synechocystis sp. PCC 6803	Cyanobacteria	3.95	47.35	12 h	Photoautotroph	Conjugation

[a] *S. albus* J1074 was recently reclassified as *Streptomyces albidoflavus*.

putida, *Bacillus subtilis*, *Anabaena* sp. PCC 7120 (hereafter PCC 7120), and *Synechocystis* sp. PCC 6803 (hereafter PCC 6803). These strains are representative of most of the species of bacteria that researchers have used for heterologous production of polyketides and non-ribosomal peptides.

S. lividans and *S. albus* are two of the most common *Streptomyces* species used to study polyketides and non-ribosomal peptides from the natural product-rich actinobacteria. Compared to *E. coli* and *B. subtilis*, streptomycetes typically have slower growth rates and therefore low productivity rates for secondary metabolites [215]. *S. venezuelae* was recently identified as an alternative heterologous *Streptomyces* host because its doubling time is about half of that of *Streptomyces coelicolor* [216, 217]. While not as plentiful as actinobacteria, a considerable number of bioactive natural products have been isolated from myxobacteria [28]. This bacterial class is also plagued with slow growth rates, with the model myxobacterium, *Myxococcus xanthus*, exhibiting a doubling time of 4–5 hours [218].

P. putida has been proposed as a "hybrid" heterologous host for the production of polyketides and nonribosomal peptides. Its growth rate is similar to that of *E. coli*, it has a higher GC content similar to that of actinobacteria and myxobacteria, and pseudomonads encode many natural product BGCs in their genomes thereby providing native auxiliary proteins [171, 219]. *B. subtilis* is another fast-growing heterologous host, and *Bacillus* species are a rich source of natural products [220]. The low GC content of *B. subtilis* makes it an attractive candidate for the heterologous expression of BGCs

from other low-GC bacteria and a difficult host for expressing the more common high GC content clusters from actinobacteria.

Cyanobacteria are an intriguing alternative to heterotrophic microorganisms for bioprocesses because they can convert sunlight and CO_2 into chemicals of interest [221]. PCC 7210 and PCC 6803 are both attractive cyanobacterial heterologous hosts for natural products and have similar doubling times of 14 and 12 hours, respectively [222]. Alternative cyanobacterial hosts have faster growth rates, including *Synechococcus elongatus* PCC 7942, which has a doubling time of 5 hours, and a related strain, *S. elongatus* UTEX 2973, grows twice as fast [223]. Heterologous production of a polyketide or nonribosomal peptide has not yet been reported in UTEX 2973 or another fast-growing model species *Synnechococcus* sp. PCC 7002, which contains one native PKS [52].

Researchers have characterized promoters for most of the heterologous hosts mentioned in Table 2.1, including *Streptomyces* species [217, 224]. *P.putida* [179, 225], *B. subtilis* [226], PCC 7120 [227], PCC 6803 [228], and of course, *E. coli* [180]. Most myxobacterial genetic tools have been developed for *M. xanthus*, and heterologous expression in this species has involved several constitutive promoters, but a promoter library or a reliable inducible promoter has not been reported in the literature [229]. Therefore, *M. xanthus* is limited in transcriptional control of heterologous genes compared to other candidate hosts.

Methods for integrating BGCs into the chromosome of heterologous hosts include transposition, site-specific phage recombinases, and homologous recombination. All three of these tools are available for *P. putida* and have been demonstrated for the chromosomal integration of BGCs [179, 230, 231]. Integration via phage recombinases is a convenient method for introducing heterologous genes to *Streptomyces* because most species contain at least one phage integration site [217]. Transposition is the most reliable method for stably integrating large (>50 kb) BGCs into the chromosome of *M. xanthus*, whereas phage integration cannot be used for constructs of this size [213, 232]. RecA-mediated recombination is an efficient method for integrating exogenous DNA into the chromosome of *B. subtilis* [233]. Even though there are few examples of cyanobacteria being used as

a heterologous host for BGCs, genetic tools have been developed for the heterologous production of natural products in these species [234, 235, 17, 236]. Lastly, CRISPR-based genome editing methods are available for all of the heterologous hosts discussed in this review [161, 162, 196, 198, 237, 238, 239, 240].

Common substrates for PKSs and NRPSs include primary metabolites, such as acetyl-CoA, malonyl-CoA, and proteinogenic amino acids. However, many BGCs require substrates that are either not intrinsically synthesized or are synthesized in small quantities by some heterologous hosts [40, 41, 241]. This issue can universally be addressed by expressing a promiscuous coenzyme A ligase in the heterologous host and supplying the necessary substrate in the growth media [242, 243]. There has also been success in engineering the de novo synthesis of metabolites, such as methylmalonyl-CoA, ethylmalonyl-CoA, and non-proteinogenic amino acids, by expressing heterologous pathways or by overexpressing native pathways [206, 244, 245]. These strategies have also been used to heterologously synthesize natural products functionalized with azide and alkyne groups, enabling semi-synthetic strategies for producing natural product derivatives that can be further diversified chemically [135, 246].

As mentioned previously, PPTases are essential for PKS and NRPS activity. PPTase genes are often located in BGCs encoding PKSs and NRPSs, but in the case that they are not, heterologous hosts must express a broad-range PPTase in order for the heterologous PKSs and NRPSs to be active [137]. PPTases are commonly found in actinobacteria [247], and *E. coli* is the only heterologous host mentioned above that does not encode a PPTase in its genome. Pfeifer *et al.* solved this issue by introducing *sfp*, a gene encoding a broad-range PPTase from *B. subtilis* [248, 53]. Multiple cyanobacteria express genes encoding PPTases, and the PPTase from PCC 7120 appears to possess broad-range activity, but the one from PCC 6803 does not [249, 250]. Both *M. xanthus* and *P. putida* have at least one *sfp*-type PPTase, and the one from *P. putida* has demonstrated a broad substrate range [137, 251, 252]. Unlike PPTases, MLPs from phylogenetically distant hosts are not interchangeable [253, 254]. Often the MLP is encoded in the same BGC as the NRPS, but a considerable amount of BGCs do not encode an MLP, so care must be taken to

provide a compatible MLP if they are to be transferred to a heterologous host [255].

2.5 Recent Examples of Heterologous Production of Bioactive Polyketides and Non-Ribosomal Peptides

Beginning in 2013, there has been a consistent increase in the number of studies using bacterial heterologous expression to study or engineer the biosynthesis of polyketides and non-ribosomal peptides (Fig. 2.5a). Unsurprisingly, actinobacteria are the most abundant source of sequenced BGCs encoding PKSs and NRPSs, and *Streptomyces* species have been the most common hosts for the heterologous expression of these BGCs (Fig. 2.5b and c). They are also the most common heterologous hosts for the discovery of novel polyketides and non-ribosomal peptides through the heterologous expression of BGCs from metagenomic libraries and cryptic BGCs (Fig. 2.5d). *E. coli* is the second most popular heterologous host, and it is most commonly engineered for improved production of polyketides and non-ribosomal peptides natively produced by non-proteobacteria, showing that researchers are taking advantage of the state-of-the-art synthetic biology tools available for *E. coli*. Studies using myxobacteria, pseudomonads, and *B. subtilis* as the heterologous host are less common, and in contrast to those using *Streptomyces*, most are attempting to engineer the production of the polyketides and/or non-ribosomal peptides of interest.

Except for *E. coli*, heterologous hosts are most often used to express BGCs from strains of the same phylum (Fig. 2.5c). Related strains are more likely to share characteristics important for heterologous gene expression, such as codon usage and GC content [256]. This phylogenetic dependence is a limiting factor for the study of BGCs from cyanobacteria. Even though the number of bioactive secondary metabolites isolated from cyanobacteria has increased in recent years [28], utilizing cyanobacteria for the heterologous production of polyketides and non-ribosomal peptides is still rare, and optimal heterologous hosts for the production of cyanobacterial secondary metabolites are still

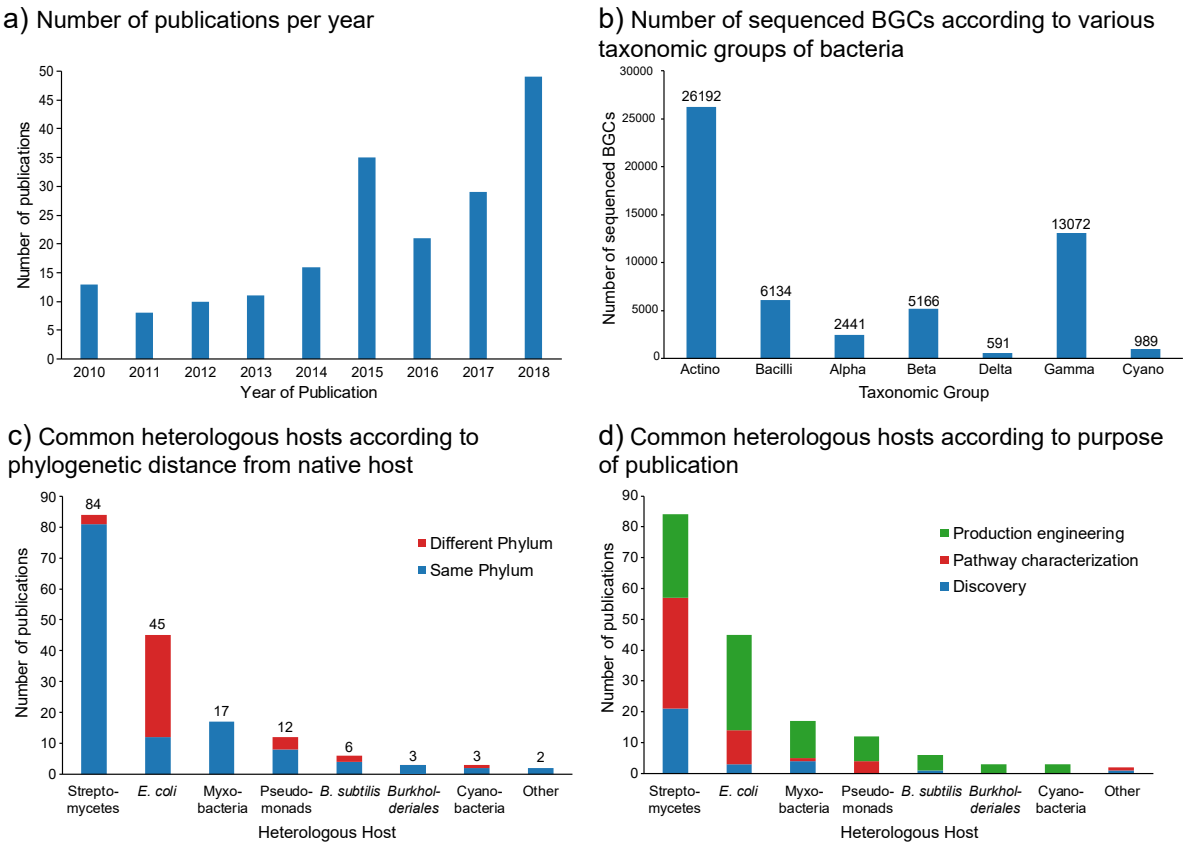


Figure 2.5: **Trends impacting heterologous expression of BGCs in bacteria.** **a)** Number of publications reporting heterologous expression of polyketide and non-ribosomal peptide BGCs per year. **b)** Number of sequenced BGCs encoding PKSs and/or NRPSs in the AntiSmash database. Data labels represent value on bar chart. All groups represent a bacterial class, except for the phylum cyanobacteria. Values were determined by building a query on the AntiSmash database searching for clusters of the types "nrps" and/or "pks" for each taxonomic group of interest. The AntiSmash database was accessed on January 6, 2019. **c)** Number of publications reporting heterologous expression of BGCs according to heterologous host and phylogenetic distance between the native host and heterologous host. Data labels represent total number of publications. **d)** Number of publications reporting heterologous expression of BGCs according to heterologous host and publication purpose. Data in panels C and D are limited to articles published from 2013 to 2019. Publications were identified via advanced searches in PubMed and Web of Science and supplemented by a manual literature review. The PubMed query was "(((heterologous[Title/Abstract] AND (expression[Title/Abstract] OR production)[Title/Abstract])) AND (polyketide OR nonribosomal peptide OR non-ribosomal peptide)) NOT (fungi[Title/Abstract] OR fungal[Title/Abstract] OR plant[Title/Abstract])". The Web of Science query was "ALL = (heterologous AND (expression OR production)) AND ALL = (polyketide OR nonribosomal peptide OR non-ribosomal peptide) NOT TS = (fungal OR fungi OR plant)". Irrelevant publications were manually removed from the search results.

to be determined. There is a relatively low number of sequenced genomes and BGCs from myxobacteria and cyanobacteria compared to other taxonomic groups (Fig. 2.5b), hindering the identification and heterologous expression of new BGCs from these groups [257, 258].

In the past two years, three publications have reported the capabilities of *Schlegelella brevitalea* sp. nov. DSM 7029 (originally classified as Polyangium brachysporum, hereafter DSM 7029 [259]) as an alternative heterologous host for the production of myxobacterial natural products [260, 261, 262]. However, yields were 1–2 orders of magnitude lower compared to heterologous product yields from *M. xanthus*. Currently, *M. xanthus* has yielded higher titers for myxobacterial natural products, but DSM 7029 could be developed for heterologously expressing BGCs from other betaproteobacteria, another natural product-rich class of bacteria [263, 264].

In the following sections, we discuss individual studies that report the heterologous production of nine bioactive polyketides and non-ribosomal peptides from the aforementioned heterologous hosts. Throughout our discussion, we highlight the synthetic biology techniques and metabolic engineering strategies mentioned in section 3 that researchers have used to improve heterologous production. The structures of these products are depicted in Fig. 2.6 and their properties are summarized in Table 2.2.

2.5.1 Heterologous Production in Streptomycetes

Actinobacteria have been the source for several prominent antibacterial compounds, such as daptomycin and chloramphenicol. Roughly 35% of all known microbial natural products with antinfective, antitumor, and antiviral activities were discovered from actinobacteria [28]. The development of model *Streptomyces* species as heterologous hosts will be essential for the discovery and production of actinobacterial natural products, as they will likely continue to be the major source of novel bacterial secondary metabolites. Recently, researchers have used genetically tractable *Streptomyces* species for the heterologous production of the actinobacterial natural products, neoantimycins and oxytetracycline. Neoantimycins are hybrid polyketides/non-ribosomal peptides produced by several

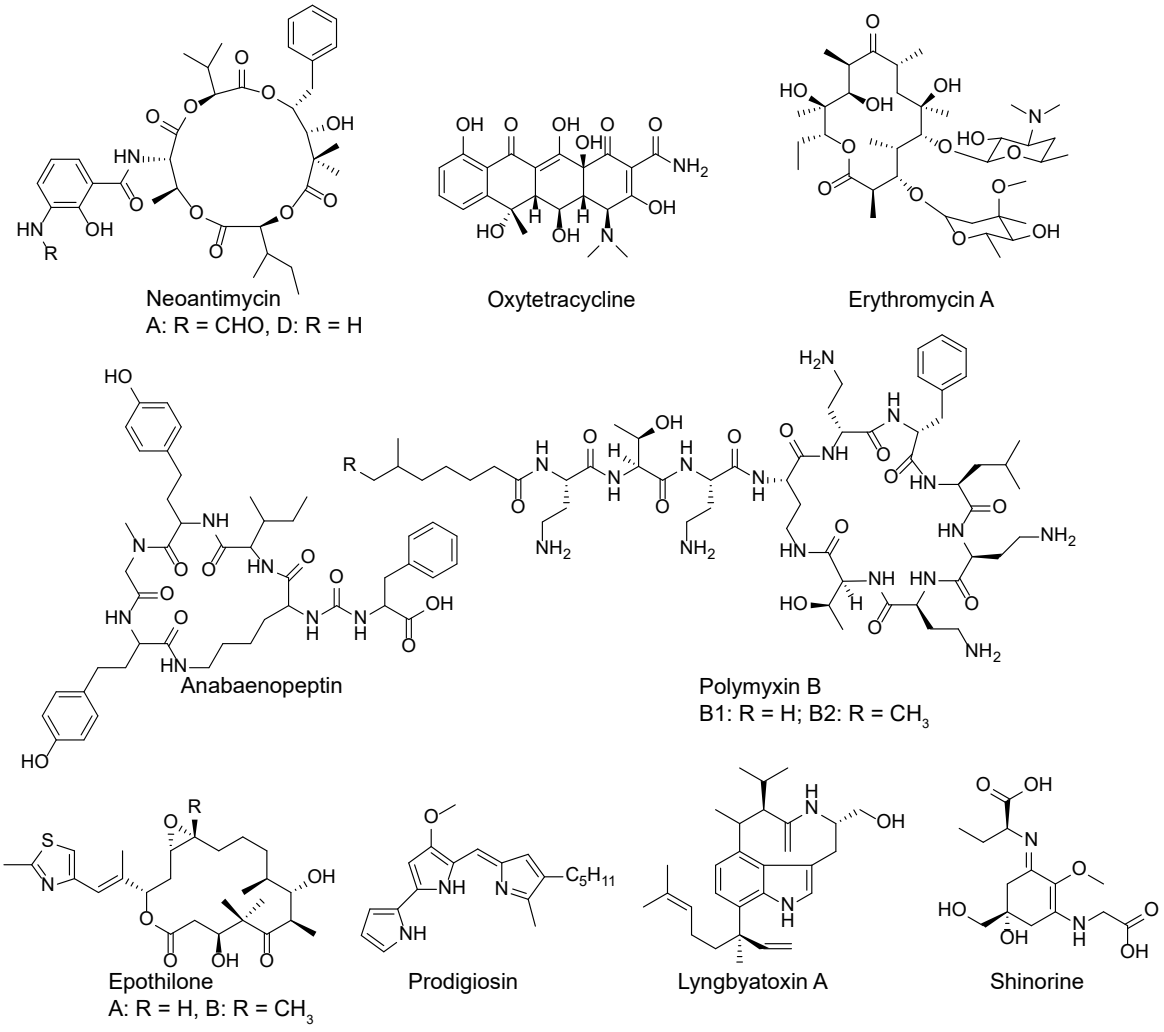


Figure 2.6: **Structures of select bioactive polyketides and non-ribosomal peptides produced by heterologous hosts.**

actinobacteria, and several derivatives have been found to induce apoptosis in cancer cells. [265, 266] Neoantimycin A specifically inhibits the K-Ras GTPase that is implicated in the development of most pancreatic, colorectal, and lung cancers [267]. Oxytetracycline is a precursor for the semi-synthesis of tetracycline-based antibiotics [268]. Tetracycline and its derivatives interfere with the association of aminoacyl-tRNAs with the ribosome in bacteria thereby inhibiting protein synthesis [269].

The BGCs producing neoantimycins from *Streptomyces conglobatus* and *Streptomyces orinoci* were recently characterized through heterologous expression of the clusters in *S. albus* [270, 271]. Heterologous expression of the BGC from *S. conglobatus* yielded a novel derivative of neoantimycin that was up to 10 times more active toward cancer cell lines than cisplatin, a common anti-cancer drug. Cas9-assisted genome editing was used to

Table 2.2: Bioactive polyketides (PK) and non-ribosomal peptides (NRP) produced through heterologous expression in bacteria in recent studies

Heterologous host	**Molecule**	**Native host**	**Class**	**Bioactivity**	**Titer**
S. albus	Neoantimycin	*Streptomyces conglobatus*	PK/NRP	Anticancer	not determined [270, 272]
S. venezuelae	Oxytetracycline	*Streptomyces rimosus*	PK	Antibiotic	430 mg L^{-1} [268]
E. coli	Erythromycin	*Saccharopolyspora erythrea*	PK	Antibiotic	40 mg L^{-1} [273]
E. coli	Anabaenopeptin	*Nostoc punctiforme*	NRP	Protease inhibitor	n. d. [274]
M. xanthus	Epothilone	*Sorangium cellulosum*	PK	Anticancer	21 mg L^{-1} [275]
P. putida	Prodigiosin	*Serratia marcescens*	PK/NRP	Antibiotic, antitumor, immunosuppressive	150 mg L^{-1} [276]
B. subtilis	Polymyxin	*Paenibacillus polymyxa*	NRP	Antibiotic	200 mg L^{-1} [277]
Anabaena sp. PCC 7120	Lyngbyatoxin A	*Moorea producens*	NRP	Cytotoxin	3.2 mg L^{-1} [214]
Synechocystis sp. PCC 6803	Shinorine	*Fischerella* sp. PCC 9339	NRP	Antioxidant, sunscreen	0.71 mg L^{-1} [278]

introduce part of the neoantimycin BGC from *S. orinoci* into a related antimycin BGC on the chromosome of *S. albus* to generate a chimeric pathway that produces several neoantimycin derivatives at levels comparable to *S. orinoci*.

The fast growth rate of *S. venezuelae* makes it an attractive host for the heterologous production of actinobacterial polyketides and non-ribosomal peptides. Yin *et al.* successfully engineered *S. venezuelae* to heterologously produce oxytetracycline. Initially, the genes in the oxytetracycline BGC did not express in *S. venezuelae*, resulting in no oxytetracycline production. The authors improved transcription of the heterologous pathway by introducing a transcriptional activator from the native producer. They also increased intracellular levels of malonyl-CoA by overexpressing an acetyl-CoA carboxylase from *S. coelicolor*. The final oxytetracycline titer of 430 mg L^{-1} is comparable to that of the native host, *Streptomyces rimosus*, but *S. venezuelae* can achieve this titer in two days, while the native host takes eight, resulting in a four-fold increase in productivity through heterologous production.

2.5.2 Heterologous Production in *E. coli*

The advanced synthetic biology tools characterized in *E. coli* and the wealth of knowledge of *E. coli* metabolism has allowed researchers to engineer *E. coli* for the heterologous production of a select few natural products. Recent compounds include the actinobacterial antibiotic, erythromycin, and anabaenopeptins, which have been discovered from multiple cyanobacteria [279, 280]. Erythromycin inhibits protein synthesis in bacteria

by binding to the 50S subunit of the ribosome and preventing translocation along the RNA transcript, and this antibiotic is still used today to treat bacterial infections [281]. The cyanobacterial natural products, anabaenopeptins, exhibit protease inhibitor activity, specifically against protein phosphatase I, an enzyme that regulates transcription of HIV-1 [280, 282, 283].

Erythromycin is one of the most studied actinobacterial polyketides, partly because its synthase is the first PKS to be functionally expressed in *E. coli*[248]. The entire BGC for erythromycin synthesis from Saccharopolyspora erythrea was first successfully reconstituted in *E. coli* by Zhang *et al.*, achieving titers up to 10 mg L^{-1} [284]. The availability of advanced synthetic biology tools and metabolic knowledge for *E. coli* has led to gradual increases in titers. Strategies for increasing erythromycin titers include engineering native metabolic pathways to increase intracellular concentrations of propionyl-CoA, introducing heterologous enzymes from S. coelicolor for methylmalonyl-CoA biosynthesis, and optimizing the expression plasmids encoding the erythromycin BGC [285, 286]. The most recent study achieved an erythromycin titer of 40 mg L^{-1} [273].

A recently developed strategy for cloning BGCs into heterologous hosts resulted in the heterologous production of cyanobacterial non-ribosomal peptides in *E. coli* [274]. Greunke *et al.* used Direct Pathway Cloning (DiPAC) to refactor the BGC for anabaenopeptin with *E. coli*-specific promoters and introduced the optimized pathway into *E. coli* via two replicative plasmids. The authors did not report an exact value for anabaenopeptin production, but HPLC/MS analysis revealed that their engineered strain of *E. coli* produced over 100 times more anabaenopeptins in 48 hours than what the native producer produced in 50 days.

These two studies demonstrate how well-characterized genetic tools can facilitate the heterologous expression of BGCs from phylogenetically distant bacteria. The favorable growth characteristics also make *E. coli* an attractive host for producing cyanobacterial natural products because it can reach higher cell densities and production titers in a much shorter time compared to cyanobacteria. However, some studies suggest that *E. coli* cannot be a widely useful heterologous host for natural products due to apparent

toxicity from the expression of heterologous BGCs [84, 287].

2.5.3 Heterologous Production in *M. xanthus*

Of the heterologous hosts discussed in this review, *M. xanthus* appears to have the least developed synthetic biology toolbox available. However, several researchers have been able to improve the production of myxobacterial compounds in this heterologous host by optimizing transcription of the corresponding BGC. The antitumor compounds, epothilones have been a common target for heterologous production in *M. xanthus*. Epothilones bind to the β-tubulin subunit of microtubules, leading to their polymerization and inducing apoptosis in cancer cells [288].

Heterologous production of epothilones in *M. xanthus* was first reported by Julien and Shah with titers approaching 0.5 mg L^{-1} [289]. Since then, several studies have attempted to improve the production of epothilone derivatives. Yue et al. discovered that several native promoters from *M. xanthus* are transcriptionally active in different stages of growth, resulting in low epothilone titers [290]. They constructed several tandem-repeat promoter variants to drive the expression of the epothilone BGC and improved epothilone titers two-fold. Peng *et al.* achieved similar results using CRISPR/dCas9 activation of the epothilone BGC. They fused the ω transcription factor to a catalytically inactive Cas9 (dCas9) and screened several sgRNA targets for improved epothilone production, demonstrating the advantages of transcriptional control in the heterologous production of natural products.

A recent study details an impressive attempt to completely refactor the epothilone BGC from *S. cellulosum* for heterologous expression in *M. xanthus* [291]. The authors codon optimized individual genes for expression in *M. xanthus* and distributed them into four operons using an established promoter, RBS, and terminator commonly used in *M. xanthus*. While epothilone production using the artificial pathway was much lower than other studies using the native pathway (100 μg L^{-1} vs. 20 mg L^{-1}), this effort shows a great step forward towards refactoring complex natural product BGCs in *M. xanthus*. Refactoring pathways into synthetic genetic circuits often requires testing multiple designs

and can reduce production compared to the native pathway [38]. Future iterations of the artificial epothilone pathway should lead to improvements in heterologous production.

2.5.4 Heterologous Production in *B. subtilis*

Despite the advantageous growth characteristics of *B. subtilis*, it is limited as a heterologous host for polyketides and non-ribosomal peptides because of the low GC content of its chromosome. For example, a refactored pathway for the synthesis of 6-deoxyerythronolide B, the precursor to erythromycin, in *B. subtilis* resulted in a titer of 2.6 μg L^{-1} [292], much lower than what has been achieved in *E. coli* [248]. Several natural products have been discovered from low-GC bacteria, including other Bacilli and some cyanobacteria. One example of these compounds is a class of antibiotics called polymyxins. These compounds are bactericidal towards Gram-negative bacteria, and their primary mode of action is increasing cell wall permeability [293, 294].

Kim *et al.* demonstrated the utility of *B. subtilis* for heterologously producing non-ribosomal peptides from other low-GC Bacilli by engineering a strain of *B. subtilis* to produce polymyxins, antibiotics natively produced by *Paenibacillus polymyxa* [295]. Integrating the polymyxin BGC into the chromosome of *B. subtilis* yielded a strain capable of producing up to 200 mg L^{-1} of polymyxins. The authors also succeeded in engineering strains towards the selective production of polymyxins B and E by replacing domains in the polymyxin NRPS with homologous domains with the desired substrate specificity.

2.5.5 Heterologous Production in Cyanobacteria

The development of cyanobacteria for the heterologous production of polyketides and non-ribosomal peptides is still in its infancy. However, researchers are poised to make rapid improvements in this field with the availability of well-characterized synthetic biology tools for cyanobacteria [296]. Roulet *et al.* have developed genetic tools specifically for polyketide production in *S. elongatus* PCC 7942 and applied them to the heterologous production of multimethyl-branched esters [234]. Two exceptional studies, described be-

low, demonstrate the capabilities of PCC 7120 and PCC 6803 for producing two bioactive non-ribosomal peptides, lyngbyatoxin A and shinorine, respectively. Lyngbyatoxin A is a cytotoxin and acts as a tumor promoter in mammalian cells by activating protein kinase C, an enzyme that regulates cell proliferation and differentiation [297]. Shinorine is a UV-protective natural product that also exhibits antioxidant activity by activating the Keap1–Nrf2 pathway that is responsible for regulating the production of antioxidant enzymes in human cells and is therefore a common target for treating diseases caused by oxidative stress [298].

PCC 7120 was used as a heterologous host for the production of lyngbyatoxin A, a cytotoxin naturally produced by *Moorea producens* [214]. The authors introduced the BGC for lyngbyatoxin A into PCC 7120 through a replicative vector. Initially, PCC 7120 was able to express the BGC using the native promoter from *M. producens* and produce lyngbyatoxin A, but the authors improved production by altering the nitrogen source and replacing the native promoter with other promoters characterized for the heterologous host. Lyngbyatoxin A production was maximized and increased 13-fold to 3.2 mg L^{-1} when the strain was grown on nitrate and a strong constitutive promoter was used to drive expression of the BGC.

Yang *et al.* describe the only study so far on the heterologous production of a non-ribosomal peptide-based natural product in PCC 6803 [278]. The authors introduced a BGC from *Fischerella* sp. PCC 9339 responsible for synthesizing shinorine using a replicative vector that co-expressed the shinorine BGC and the broad-range PPTase from PCC 7120 [299]. Initially, imbalanced expression levels of individual genes in the BGC led to the accumulation of pathway intermediates and limited shinorine production. The authors removed this imbalance and improved production by inserting strong constitutive promoters upstream of the poorly expressed genes in the BGC, resulting in a titer of 0.71 mg L^{-1} of shinorine.

Remarkably, both heterologous hosts were able to generate natural product titers comparable to those of the native producers. The use of well-characterized promoters in these two studies demonstrates the importance of transcriptional control when expressing

heterologous BGCs.

2.5.6 Heterologous Production in *P. putida*

P. putida has been a popular choice for the heterologous production of polyketides and non-ribosomal peptides from proteobacteria, particularly those from myxobacteria. One early study demonstrated that *P. putida* was capable of producing myxochromide S with titers up to 5 times higher than the native myxobacterial producer, *Stigmatella aurantiaca* [300]. However, more recent studies have reported low titers of heterologous products compared to alternative hosts [213, 229, 301]. Despite the limited success of *P. putida* as a heterologous host for natural products, many genetic tools have been developed for it within the last few years, so heterologous production from this species could likely be improved if researchers take advantage of the latest tools.

One exceptional study reports the heterologous production of prodigiosin, a pigmented polyketide/non-ribosomal peptide natively synthesized by *Serratia marcescens* [302]. Prodigiosin is an antibiotic that inhibits transcription in *E. coli*, but not *P. putida* [165, 303]. Recently, there has been increased interest in prodigiosin due to its reported anticancer and immunosuppressant activities. It has displayed apoptotic and antimetastatic affects against multiple cancer cell lines [304, 305], and it is a T cell-specific immunosuppressant that inhibited the development of immune disease in mouse models [306].

Domröse *et al.* introduced the BGC for prodigiosin synthesis through Tn5 transposition. The authors took advantage of the deep red color generated by prodigiosin and constructed a transposable vector with a promoterless copy of the BGC. After screening for transposon mutants that appeared as red colonies, they identified several mutants that had integrated the gene cluster downstream of a strong chromosomal promoter, resulting in prodigiosin titers of up to 150 mg L^{-1} [276]. This strategy enabled the discovery of strong native promoters for the heterologous expression of BGCs and improved prodigiosin production compared to a previous study that used a synthetic promoter [307].

P. putida as a heterologous host from 2018-2021

In recent years, researchers have demonstrated *P. putida*'s versatility as a heterologous host for natural products, primarily through sucessfully producing natural products from several proteobacteria and through metabolic engineering to increase production titers. Zhang *et al.* used *P. putida* to elucidate the biosynthesis of a class of lipodepsipeptides, the thalassospiramides, natively synthesized by marine α-proteobacteria [308]. Using a site-specific recombinase, IntB13, the authors integrated two BGCs into the chromosome of *P. putida* EM383, a derivative of KT2440 optimized for heterologous gene expression, and were able to produce titers similar to or greater than the native strains. They also found that *P. putida*'s native PPTase was able to activate the PKS and NRPS enzymes of the two BGCs, but production was improved if the PPTase from one of the native hosts was co-expressed with the BGCs. Thalassospiramide biosynthesis involves hydrid PKS/NRPS enzymes with multiple modules, and these results demonstrate that *P. putida* can functionally express mega-enzymes that *E. coli* cannot.

Another study systematically compared a suite of γ-proteobacteria, including *P. putida*, in the expression of BGCs from *Photorhabdus luminescens*. The authors designed a universal genome editing method called chassis-independent recombinase-assisted genome engineering (CRAGE) that they used to integrate BGCs into multiple bacteria and express them using T7 RNA polymerase [309]. *P. putida* and other *Pseudomonas* species consistently produced a lower amount of metabolites compared to the heterologous hosts that were more phylogenetically related to *P. luminescens*, challenging the notion that a bacterial host could be developed as a universal heterologous host for natural products from diverse bacteria.

Researchers have begun to investigate metabolic engineering strategies to improve polyketide and nonribosomal peptide production in *P. putida*. These efforts often take advantage of pigmented natural products that provide a rapid screen. Flaviolin is a pigmented polyketide, and researchers have used this molecule as a biosensor for malonyl-CoA concentration and screened for conditions or mutations that increase malonyl-CoA availability [310, 311]. *P. putida* was recently engineered to produce robust titers of in-

digoidine, a nonribosomal peptide synthesized from L-glutamine [15]. The authors used a genome-scale metabolic model of *P. putida* to identify 14 gene targets to downregulate. Multiplexed CRISPR interference was used to repress nine of these genes, and the engineered strain achieved 26 g L^{-1} indigoidine in batch and fed-batch cultures.

2.6 Future Directions of Heterologous Production

Researchers have put considerable effort towards identifying alternative heterologous hosts for the discovery and production of polyketides and non-ribosomal peptides. *Streptomyces* species remain the dominant candidates, and this trend will hold in the foreseeable future as actinobacteria continue to be rich sources for bioactive natural products. Currently, heterologous hosts are most successful when expressing BGCs from phylogenetically similar species. However, for non-model heterologous hosts that are related to natural product-rich species, synthetic biology tools are limited and it is considerably more difficult to engineer these hosts for improved heterologous production. The development of new synthetic biology tools needs to continue for these bacteria in order for heterologous hosts to become a more viable option for producing polyketides and non-ribosomal peptides. However, the model heterologous hosts for some bacterial taxa, such as myxobacteria and cyanobacteria, have slower growth rates and are more difficult to handle compared to other heterologous hosts.

Refactoring bacterial pathways in heterologous hosts has become a common strategy for pathways that do not produce polyketides and non-ribosomal peptides [38, 312], but when researchers do heterologously express PKSs and NRPSs, the native BGC is often left fully intact, requiring the use of potentially nonoptimal promoters, RBSs, and terminators. Attempts to refactor these BGCs are usually limited to introducing characterized promoters for 1–2 operons. This minimal approach can be improved upon by taking advantage of tools for designing RBSs and investigating effects of termination on gene expression in heterologous hosts [190, 313].

While genetic tools remain poorly characterized for many natural product-rich bacte-

ria, phylogenetically distant species with better developed tools could become ideal hosts, provided heterologous BGCs are refactored for optimal expression. Successful attempts to refactor BGCs from actinobacteria and cyanobacteria for heterologous expression in *E. coli* demonstrates the utility in this strategy and suggests that the ideal heterologous host for producing polyketides and non-ribosomal peptides can be an unrelated species [87, 314]. However, issues that synthetic biology cannot immediately address, such as observed toxic effects from expressing heterologous pathways in *E. coli* [84, 287], require that alternative hosts are developed for these projects. Improved synthetic biology tools will be necessary for non-model organisms if pathway refactoring is to be fully realized as a strategy for improving the heterologous production of polyketides and non-ribosomal peptides in these hosts.

2.7 Research Needs

Characterized genetic tools for heterologous hosts

Engineering metabolic pathways for natural products synthesis is often more challenging than engineering primary metabolism because these pathways are generally less understood and the gene clusters encoding them are large in size and therefore difficult to manipulate. These issues are compounded by a lack of well-characterized genetic tools for microorganisms that are able to functionally express these pathways. Whether the microbe is the native host or a heterologous host, researchers have a small set of genetic parts that they can use to modify a pathway of interest. Ideally, there should at least be a small library of constitutive promoters and a set of inducible promoters for a metabolic engineering chassis. Steps to ensure that heterologous proteins are being translated efficiently are also needed. Investigations into determining optimal RBS sequences are rare for non-model bacteria, and research into codon usage and optimization for bacteria has only been based on findings for *E. coli* so far.

Genetic and metabolic engineering of natural product pathways

Most metabolic engineering attempts for natural products have focused on *Streptomyces* hosts, where researchers have devised universal strategies for improving production titers, such as promoter engineering and deleting competing pathways. Individual natural product pathways are usually complex and variable, so examples of engineering reactions within a pathway of interest are less common. Complete pathway refactoring, where every gene is re-organized into synthetic operons, has been applied to some pathways, but not for pathways producing polyketides or nonribsomal peptides. In most cases, pathways of interest have not evolved for overproduction, so a re-working of metabolism should be expected when designing a production host, but efforts so far have not used most of the engineering tools available for bacterial hosts.

Protein engineering of assembly line enzymes in natural product synthesis

Genetic information has shown that PKSs and NRPSs have evolved through domain and module substitutions and duplications. It is possible to some extent to repeat this process in the lab with synthetic biology techniques, but there is not yet a defined set of rules for creating chimeric enzymes. There are no established domain boundaries for designing functional chimeras, so optimizing substitutions within these enzymes is laborious because multiple variants need to be analyzed. Active site mutagenesis of gatekeeping domains avoids issues with protein stability of chimeric enzymes, but our current understanding of substrate specificity does not yet permit reliable specificity swaps in modular enzymes. Directed evolution has provided a path forward for generating enzymes with altered specificity, but these studies have not provided enough sequence-function information to inform future rational modifications.

Natural products are still a viable resource for drug discovery. Next-generation sequencing has provided countless gene clusters with the potential to make a bioactive compound, but many steps are required to generate novel compounds from this information. Synthetic biology and metabolic engineering have so far been crucial components for this new drug discovery paradigm. The available tools from these fields allow re-

searchers to clone and express new pathways, modify biosynthetic enzymes to generate natural product derivatives, and engineer production strains to produce drug candidates at larger scales.

2.8 Acknowledgements

This work was supported by a research grant from the National Science Foundation (MCB-1716594). TC is the recipient of a National Institutes of Health Biotechnology Training Program Fellowship (NIGMS 5 T32 GM08349).

Chapter 3

Genetic Tools for Reliable Gene Expression and Recombineering in *P. putida*

Authors and Contributors:

- **Taylor B. Cook** Designed and performed experiments, analyzed data, made figures, and wrote the manuscript.
- **Dr. Jacqueline M. Rand** Designed experiments, wrote the manuscript.
- **Wasti Nurani** Performed experiments.
- **Dylan K. Courtney** Performed experiments.
- **Sophia A. Liu** Performed experiments.
- **Dr. Brian F. Pfleger** Designed the experiments, edited and reviewed the manuscript.

This chapter was originally written as an article in the *Journal of Industrial Microbiology and Biotechnology* [225].

3.1 Abstract

Pseudomonas putida is a promising bacterial host for producing natural products, such as polyketides and nonribosomal peptides. In these types of projects, researchers need a genetic toolbox consisting of plasmids, characterized promoters,and techniques for rapidly editing the genome. Past reports described constitutive promoter libraries, a suite of broad host range plasmids that replicate in *P. putida*, and genome-editing methods. To augment those tools, we have characterized a set of inducible promoters and discovered that IPTG-inducible promoter systems have poor dynamic range due to overexpression of the LacI repressor. By replacing the promoter driving *lacI* expression with weaker promoters, we increased the fold induction of an IPTG-inducible promoter in *P. putida* KT2440 to 80-fold. Upon discovering that gene expression from a plasmid was unpredictable when using a high-copy mutant of the BBR1 origin, we determined the copy numbers of several broad host range origins and found that plasmid copy numbers are significantly higher in *P. putida* KT2440 than in the synthetic biology workhorse, *Escherichia coli*. Lastly, we developed a λRed/Cas9 recombineering method in *P. putida* KT2440 using the genetic tools that we characterized. This method enabled the creation of scarless mutations without the need for performing classic two-step integration and marker removal protocols that depend on selection and counterselection genes. With the method, we generated four scarless deletions, three of which we were unable to create using a previously established genome-editing technique.

3.2 Introduction

Pseudomonas putida is being developed into a prominent metabolic engineering chassis for industrial biotechnology applications [315]. *P. putida* naturally has a relatively high guanine–cytosine (GC) content and is capable of expressing complex biosynthetic clusters, such as polyketide synthases and nonribosomal peptide synthetases, making it an ideal host for the production of secondary metabolites derived from GC-rich bacteria [316, 156]. In addition to its production capabilities, *P. putida* is HV1 certified with a relatively high tolerance toward industrial solvents [317]. It has a completely sequenced genome, allowing it to be genetically tractable [318]. For *P. putida* to reach its potential as a major chassis for industrial chemical production, it requires a robust set of synthetic biology tools to allow for genetic manipulations and modulations of protein expression.

The main synthetic biology tools needed to engineer *P.putida* include a reliable set of promoters, both constitutive and inducible, a set of origin of replications that can allow for stable plasmid maintenance, and robust and efficient genome-editing techniques [319]. A few constitutive promoter libraries have been characterized in *P. putida* for chromosomal expression, with one study reporting a dynamic range around three orders of magnitude [201] and a second study finding a 72-fold range of expression [179], demonstrating that a wide range of expression on the chromosome can be achieved.

A variety of inducible promoters natively found in *Pseudomonas* species have demonstrated in *P. putida* relatively high levels of expression upon induction, including P_m induced with 3-methylbenzoate [181, 182], P_{sal} induced with salicylate [181, 182], and P_{alkB} induced with dicyclopropylketone [181, 183]. Other heterologous inducible systems have been tested in *P. putida* with varying success. Induction systems using rhamnose (P_{rhaB}; [181, 184]), arabinose (P_{araB}; [181, 185]), methyl ethyl ketone (P_{mekA}; [181, 185]), and mannitol (P_{mtlE}; [320]) demonstrated induction levels comparable to the native systems, though basal level expression varied, with P_{mtlE} exhibiting higher levels of basal expression and (P_{araB} acting as a tightly regulated system. The more commonly used *Escherichia coli* induction systems (P_{tet} and P_{lac} variants) have also been tested in *P. putida*, but with varying degrees of success. One study demonstrated 38-fold induction

with the P_{tet} system using four times the reported maximum necessary anhydrotetracycline (aTc) concentration for *E. coli* [185, 176], and a second study reported using P_{tet} with tetracycline as the inducer for tubulysin production but did not report quantitative protein expression levels, making it hard to judge the efficacy of P_{tet} in *P. putida* [301]. The dynamic range was generally low for systems involving the lac operon, with P_{trc} having only a four-fold change in expression upon induction [186]. P_{lacUV5} was seen to have induced fluorescence levels within the error for the uninduced samples; however, they did report that altering the ribosome binding site (RBS) and 5' untranslated region (5' UTR) for P_{lacUV5} caused a 26-fold increase in expression [181].

Many of the standard plasmid origins commonly used in *E. coli* are narrow host range origins incapable of replication in *P. putida*; so a variety of broad host range (BHR) origins have been developed [321, 322, 323]. The RK2 origin is a member of the IncP incompatibility group and requires an origin of replication sequence and a replication initiation protein encoded by *trfA* to function [175]. It has been shown to be stably maintained in *P. putida* and has been measured to be a low-copy plasmid in *E. coli* [324]. RSF1010 is a high-copy BHR origin routinely used for plasmid maintenance in *P. putida* [325, 177]. It is part of the IncQ incompatibility group, though it requires three genes to successfully replicate (*repA*, *repB*, *repC*), making it a larger origin than RK2 [321, 323]. pBBR1 is an interesting BHR origin that does not seem to belong to any of the standard incompatibility groups and only requires a single gene (*rep*) for replication, allowing it to be one of the smaller BHR origins [326, 327]. Even with its small size, pBBR1 has been shown to replicate in *P. putida* and is generally considered a low- to medium-copy plasmid, though the copy number can be altered by mutations in the *rep* gene [328, 169, 329]. There are also origins that have been isolated from *Pseudomonas*-specific plasmids (pR01600, pVS1, pNI10), but they must be used in conjunction with other host-specific origins to generate shuttle vectors for *P. putida* [319, 330].

There are a multitude of techniques for editing the genome of *P. putida*. Early methods for gene deletions retained the selection marker on the chromosome, limiting the overall number of genomic mutations possible in one strain [331]. To combat this issue, site-

specific recombinases such as Cre-*loxP* have been developed for use in Gram-negative bacteria to allow for efficient removal of the selectable marker [332, 333]. Cre-*loxP* has been used to efficiently create gene deletions in *P. putida* KT2440 when paired with the λRed recombinases from the λ bacteriophage, but deletions generated with the Cre-*loxP* system leaves a *loxP* scar that can affect future recombination events, also limiting the total number of gene deletions [334].

Markerless and scarless gene-editing methods remove any selectable markers or genomic scars, but they require the use of a counterselection system. In many Gram-negative bacteria, sucrose sensitivity can be conferred by the *sacB* gene thereby allowing it to be used as a counterselection marker; however, *sacB* is known to have a high mutation rate that can cause an increase in the false positive rate of mutants [335, 336]. A counterselection for scarless gene deletions based on the antimetabolite 5-fluorouracil was developed, but this technique requires a strain containing the seemingly innocuous deletion of *upp*, the gene-encoding uracil phosphoribosyltransferase, for functionality [158]. Additionally, most of these methods involve the integration of a suicide plasmid into the chromosome through a single-crossover recombination. This co-integrate can either resolve back to the wild-type sequence or to the desired knockout, resulting in a theoretical efficiency of 50%.

A counterselection based on endonuclease I-SceI allows the use of wild-type *P. putida* for scarless modifications, but this method also selects for the resolution of an integrated suicide plasmid [337]. Another recently published method found a way to improve the transformation efficiency of sequences integrated into the genome using a serine recombinase system, but this requires the presence of the bacteriophage integrase on the genome and limits its overall utility in genome editing [179].

CRISPR/Cas systems have revolutionized genome editing by providing a fully programmable system, the most common examples using DNA cleavage by the Cas9 nuclease from *Streptococcus pyogenes* as the counterselection [338]. Cas9 uses a single guide RNA (sgRNA) to target virtually any DNA sequence [159] and has been used successfully for genome editing in both Gram-negative and Gram-positive bacteria [162, 339, 340].

CRISPR/Cas9 cleavage has recently been demonstrated for enhancing the efficiency of single-stranded DNA recombineering in *P. putida* [196].

Here, we report our characterization of some of the common synthetic biology tools in *P. putida*. We have tested commonly used *E. coli* induction systems (P_{trc}, $P_{LlacO\text{-}1}$, P_{tet} and P_{araB}) in *P. putida* to determine induction curves and maximum fold induction possible. For the P_{trc} and $P_{LlacO\text{-}1}$ promoter systems, we adjusted the expression of *lacI* in an attempt to improve the fold induction for those systems. We also quantified the plasmid copy number of the BHR origins RK2, pBBR1, and RSF1010 and highlighted the differences in copy number between *E. coli* and *P. putida*. Finally, we are reporting a Cas9-assisted recombineering system for *P. putida* with genome-editing efficiencies approaching 100%.

3.3 Materials and methods

3.3.1 Plasmids, Bacterial Strains, and Growth Conditions

The plasmids and bacterial strains used in this study are shown in Table 3.1. *E. coli* MG1655 and *P. putida* KT2440 were grown in LB medium at 37 and 30 °C, respectively. LB medium was supplemented with kanamycin (50 μg mL^{-1}, Kan50), gentamycin (33 μg mL^{-1}, Gent30), tetracyline (10 μg mL^{-1}, Tet10 for *E. coli*; 25 μg mL $^{-1}$, Tet25 for *P. putida*), or 5-FU (20 μg mL^{-1}). 5-FU was purchased from Sigma-Aldrich (F6627-1G).

3.3.2 Induction Curves in *P. putida*

P. putida KT2440 strains containing GFPuv expression plasmids were grown overnight in LB Kan50. These overnight strains were inoculated in 5 mL LB Kan50 supplemented with varying amounts of inducer: IPTG (0–2.5 mM), anhydrotetracycline (0–400 ng/mL), or l-arabinose (0–2% w/v). The optical density at 600 nm and fluorescence (excitation: 400 nm, emission: 510 nm) of samples diluted 1:10 in fresh LB media were measured using a Tecan Infinite M1000 plate reader. Analytical flow cytometry was used to measure the

Table 3.1: Strains and plasmids used in this work.

Strain or plasmid	Genotype and relevant characteristics	Source
Strains		
E. coli MG1655	K-12 F- λ- *ilvG- rfb-50 rph-1*	ECGSC
P. putida KT2440	Wild-type	ATCC 47054
P. putida KTU	KT2440 Δ*upp*	[158]
Plasmids		
pBBR1k-GFPuv	IPTG-inducible P_{trc} expressing *gfpuv* on BBR1 origin	This study
pBBR2k-GFPuv	Anhydrotetracycline-inducible P_{tet} expressing *gfpuv* on BBR1 origin	This study
pBBR5k-GFPuv	IPTG-inducible P_{lacUV5} expressing *gfpuv* on BBR1 origin	This study
pBBR6k-GFPuv	IPTG-inducible $P_{LlacO-1}$ expressing *gfpuv* on BBR1 origin	This study
pBBR8k-GFPuv	L-arabinose-inducible P_{araB} expressing *gfpuv* on BBR1 origin	This study
pBBR1k-con-GFPuv	pBBR1k-GFPuv with *lacI* removed	This study
pBBR6k-con-GFPuv	pBBR6k-GFPuv with *lacI* removed	This study
pBBR1k-J23101-GFPuv	pBBR1k-GFPuv with Anderson promoter J23101 expressing *lacI*	This study
pBBR1k-J23107-GFPuv	pBBR1k-GFPuv with Anderson promoter J23107 expressing *lacI*	This study
pBBR6k-J23101-GFPuv	pBBR6k-GFPuv with Anderson promoter J23101 expressing *lacI*	This study
pBBR6k-J23107-GFPuv	pBBR6k-GFPuv with Anderson promoter J23107 expressing *lacI*	This study
pRK2-AraE	Constitutively expressed *araE* on RK2 origin	This study
pBbB1k-GFPuv	pBBR1k-GFPuv with BBR1-UP origin	[176]
pBb(B5)1k-GFPuv	pBRR1k-GFPuv with B5 origin	This study
pBb(RK2)1k-GFPuv	pBRR1k-GFPuv with RK2 origin	This study
pBb(B5)8k-sfGFP	L-arabinose-inducible P_{araB} expressing *sfgfp* on B5 origin	This study
pBBRR1MCS-2	Empty vector control with KanR and BBR1 origin	[322]
pCas9	Recombineering plasmid with constitutively expressed *cas9* and P_{araB} expressing $\alpha\beta\gamma$	This study
pJOE	*P. putida* KTU suicide plasmid	[158]
pJOE-lvaA	*P. putida* KTU suicide plasmid with repair template for *lvaA* knockout	This study
pJOE-fpvA	*P. putida* KTU suicide plasmid with repair template for *fpvA* knockout	This study
pJOE-pvdJ	*P. putida* KTU suicide plasmid with repair template for *pvdJ* knockout	This study
pgRNAtet-lvaA	Guide RNA plasmid targeting *lvaA*	This study
pgRNAtet-gcvP	Guide RNA plasmid targeting *gcvP*	This study
pgRNAtet-fpvA	Guide RNA plasmid targeting *fpvA*	This study
pgRNAtet-pvdJ	Guide RNA plasmid targeting *pvdJ*	This study

fluorescence of cells induced for GFPuv expression. Samples were washed with TBS buffer and diluted 1:10 in fresh TBS buffer before analysis in a BD FACSCalibur. Fluorescence intensities were quantified using the Flowing Software package.

3.3.3 Quantifying Plasmid Copy Number

The copy numbers of five different broad host range origins were quantified in *E. coli* MG1655 and *P. putida* KT2440 using quantitative PCR. Individual colonies of *P. putida* KT2440 strains containing a plasmid were inoculated in LB Kan50 and grown overnight in biological triplicate. Once the cultures reached stationary phase, their genomic and total DNA was extracted using phenol:chloroform:isoamyl alcohol as reported previously, with a few modifications [341, 176]. 1 mL of culture was resuspended in 400 μL 50 mM

Tris/50 mM EDTA, pH 8. Cells were lysed by the addition of 8 μL 50 mg mL^{-1} lysozyme, followed by incubation at 37 °C for 30 min. Lysis was continued by adding 4 μL 10% SDS and 8 μL 20 mg mL^{-1} proteinase K solution, mixing the samples with a 22-gauge needle, and incubating at 50 °C for 30 min. Proteinase K was then heat inactivated by incubating the sample at 75 °C for 10 min, and RNA was digested by adding 2 μL of 10 mg mL^{-1} RNase A solution and incubating at 37 °C for 30 min. The DNA was extracted by mixing the samples with 425 μL of 25:24:1 phenol:chloroform:isoamyl alcohol, vortexing vigorously for 1 min, and letting the samples sit at room temperature for a few minutes. The samples were centrifuged at 14,000g for 5 min at 4 °C. The upper aqueous phase was transferred to a new tube with a wide-opening pipet tip. DNA extraction was continued by adding 400 μL of chloroform and vortexing and centrifuging the samples as before. The upper aqueous phase was transferred to a new tube. The DNA was further purified by precipitation with 1 volume isopropanol, centrifugation of the samples at maximum speed for 30 min at 4 °C, washing with 500 μL of 70% ethanol, and rehydrating the DNA by incubating the samples in 200 μL of nuclease-free water at 65 °C for 1 h.

For determining plasmid copy numbers per cell, individual colonies of *E. coli* MG1655 and *P. putida* KT2440 strains containing a plasmid were inoculated in LB Kan50 and grown overnight in biological triplicate. Once the cultures reached stationary phase, they were diluted in distilled/deionized water to a concentration of 1000 cells μL^{-1} (using an experimentally determined conversion factor OD 0.4 = 105 cells μL^{-1}).

Quantitative PCR was performed in a 10 μL mixture containing 1X SYBR Green PCR Master Mix (Bio-Rad), 150 nM forward primer, 150 nM reverse primer, and 1 μL of sample. Primers used to determine template concentration were specific to *rrsA* (for *E. coli*), *lvaC* (for *P. putida*), and *gfpuv* (for plasmid DNA). Each sample was prepared in technical duplicate. The standard curve for chromosome concentration was made with a dilution series of a *P. putida* KT2440 genomic DNA extraction from 10^5 to 1 chromosomes μL^{-1}, the standard curve for cell concentration was made with a dilution series of wild-type cells from 10^4 to 1 cell μL^{-1}, and the standard curve for plasmid concentration was made with a dilution series of purified plasmid DNA from 2×10^5

to 20 plasmids μL^{-1}. Plasmid and genomic DNA concentrations were measured using the Qubit 3.0 Fluorometer. The reactions were prepared on an AriaMx skirted 96-well plate (Agilent Technologies) and the plate was sealed with an adhesive cover (Bio-Rad). Reactions were run on an AriaMx Real-Time PCR System (Agilent Technologies) using the following cycling conditions: 3 min at 95 °C, followed by 35 cycles of 15 s at 95 °C, 30 s at 55 °C, and 1 min at 72 °C. Melt curves were generated by increasing the temperature from 55 to 95 °C using increments of 5 °C every 5 s. After each run, C_q values were exported to Excel. Technical duplicates with a standard deviation in C_q value greater than 0.3 were not used for analysis. Standard curves were constructed by plotting C_q values vs. the log of chromosome/cell/plasmid concentration. Each standard curve used four to five data points and had an R^2 value of at least 0.99. The plasmid copy number was calculated by determining the chromosome/cell/plasmid concentration in each sample using the standard curves and then dividing the plasmid concentration by the chromosome/cell concentration.

3.3.4 Genome Editing in *P. putida*

For the two-step λRed/Cas9 recombineering protocol, *P. putida* KT2440 containing pCas9 was transformed with pJOE by electroporation and selected on LB Gent30, Kan50. One of the transformants was inoculated in LB Gent30, Kan50 and grown overnight at 30 °C. Once the cells reached stationary phase, the λRed genes were induced with 0.5% l-arabinose for 15 min. These cultures were used to prepare electrocompetent cells by washing twice with 10% glycerol and resuspending in 100 μL 10% glycerol for every 1 mL of culture. These cells were transformed with about 100 ng pgRNA by electroporation and allowed to recover in 1 mL LB for 2 h at 30 °C. The recovered cells were selected on LB Gent30, Tet25. For the one-step protocol, *P. putida* KT2440 containing pCas9 was used to prepare electrocompetent cells in the same way as the two-step protocol. These cells were transformed with about 100 ng pgRNA and about 500 ng pJOE by electroporation and recovered cells were selected on LB Gent30, Tet25. Genome editing with the 5-FU counterselection was completed as described in Graf and Altenbuchner (2011)

[158]. Transformants from all three methods were screened for the desired knockout using colony PCR with primers flanking the gene of interest. All positive hits were later screened after re-streaking for isolated colonies with a secondary colony PCR to check for the presence of wild-type. Plasmids were cured from *P. putida* by growing the cells overnight in LB media without antibiotics and plating on LB agar. Single colonies were screened for loss of antibiotic resistance.

3.4 Results

3.4.1 Inducible gene expression in *P. putida*

The P_{lac}, P_{tet}, and P_{araB} promoter systems have been used extensively in *E. coli* and other bacteria [176], but they are not well characterized for gene expression in *P. putida*. We investigated gene expression from the following promoter systems from the BglBrick vector database: two P_{lac} promoters (1k and 6k), one P_{tet} promoter (2k), and one P_{araB} promoter (8k). The P_{lac} promoter systems use the P_{trc} (1k) and the $P_{LlacO\text{-}1}$ (6k) promoters. The 1k promoter system also differs from the 6k system in that *lacI* is expressed in the same direction as the gene of interest (Fig. 3.1a). P_{trc} is also a stronger promoter than $P_{LlacO\text{-}1}$ [176]. We identified which promoter systems had high expression levels at maximum induction in *P. putida* and characterized them

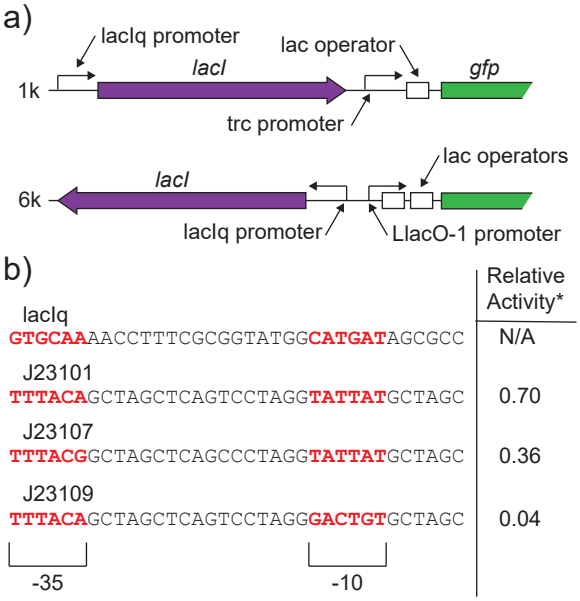


Figure 3.1: **IPTG-inducible promoter systems from the BglBrick vector database. a)** Genetic structure of the 1k and 6k promoter systems, highlighting the promoters used and the direction of *lacI* expression. **b)** Sequences of P_{lacIq} and the three Anderson promoters used to modify the 1k and 6k promoter systems. The -10 and -35 motifs are highlighted in red. (Asterisk) Relative activity is the promoter strength in *E. coli* with respect to the reference promoter in the Anderson promoter library, J23100.

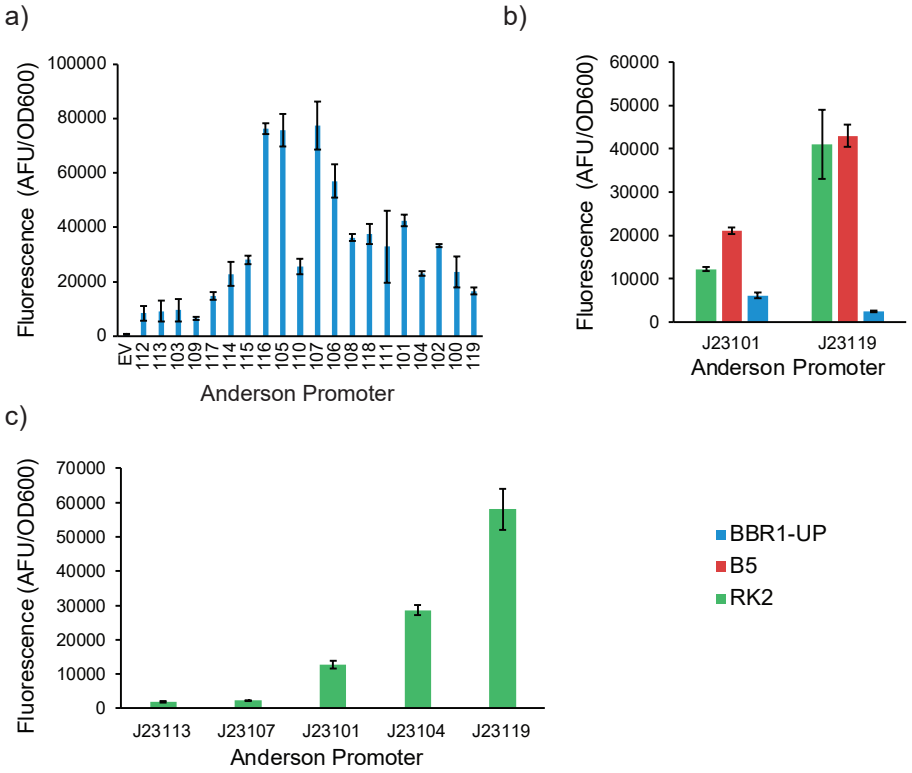


Figure 3.2: **Differences in expression of a fluorescent reporter protein from plasmids with the BBR1-UP and BBR1-B5 origins. a)** Activity of Anderson promoters when using the BBR1-UP origin. Promoters are listed in order of decreasing activity in *E. coli*. EV = empty vector control. Error bars represent one standard deviation. **b)** Activity of a sample of Anderson promoters when using the BBR1-UP,BRR1-B5, and RK2 origins. **c)** Promoter activity of an extended sample of Anderson promoters on the RK2 origin. Promoters are listed in order of increasing activity.

by collecting induction curves. We collected the data for these promoter systems after 24 h of growth using the wild-type BBR1 origin and the gfpuv gene as a fluorescent reporter. The BglBrick vector database uses a high-copy mutant of the BBR1 origin, referred to here as BBR1-UP, but our previous work demonstrated that gene expression from plasmids with this origin in *P. putida* was unreliable, especially when using strong promoters (Figure 3.2). *P. putida* cells replicating plasmids with BBR1-UP also had a growth defect (Figure 3.3).

Of the *lac* promoters, the 1k system had the highest expression levels at maximum induction, but at high inducer concentrations, gene expression began to decrease (Fig. 3.5a). The 6k promoter system had extremely poor induction, but $P_{LlacO\text{-}1}$ produces an easily detectable amount of GFPuv in the absence of *lacI* expression (Fig. 3.4b). We hypothesized that *lacI* expression is too high in *P. putida* for the 6k system and typical levels of inducer cannot de-repress the promoter. We confirmed that the promoter driving

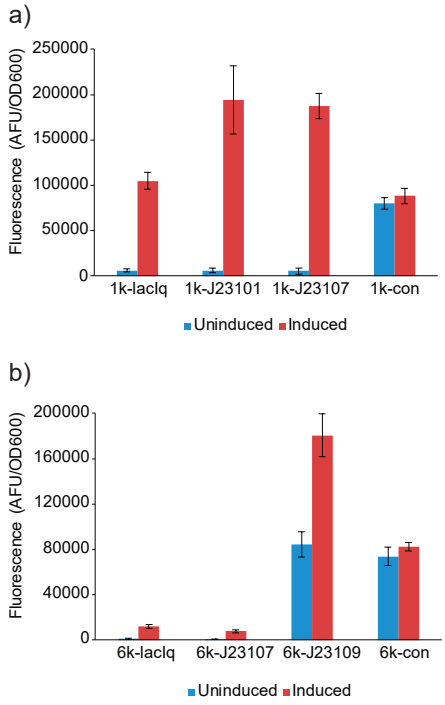


Figure 3.4: **Improving expression at maximum induction for 1k and 6k promoter systems.** **a)** Fluorescence for modified 1k promoter systems. **b)** Fluorescence for modified 6k promoter systems. The samples are listed in order of decreasing *lacI* expression, from left to right. Samples 1k-con and 6k-con are constitutive versions of the 1k and 6k systems, respectively. Error bars represent one standard deviation. Fluorescence values are adjusted so that the empty vector control has a value of 0.

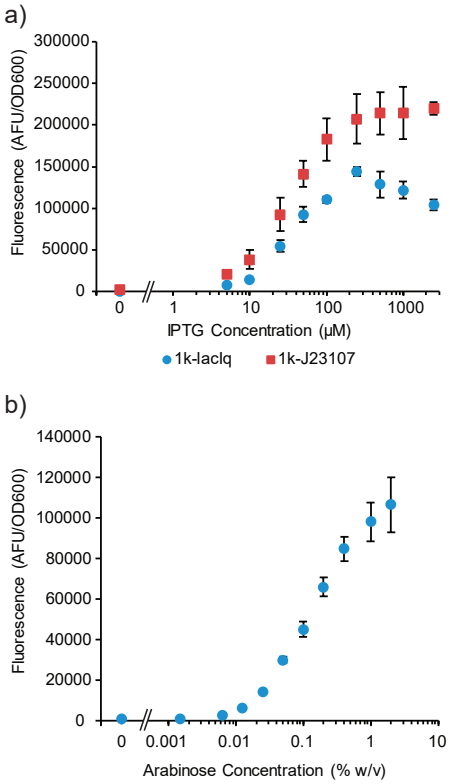


Figure 3.5: **Induction curves for BglBrick promoter systems in *P. putida*.** **a)** Induction curves for 1k promoter system with P_{lacIq} and J23107 promoters expressing *lacI*. **b)** Induction curve for 8k system. Error bars represent one standard deviation. Fluorescence values are adjusted so that the empty vector control has a value of 0.

lacI expression, P_{lacIq}, is much stronger than a sample of promoters from the Anderson promoter library [180] (Fig. A.1), and we then modified the 1k and 6k systems by switching P_{lacIq} with these weaker constitutive promoters (Fig. 3.4b). Gene expression at maximum induction for the modified 1k system was almost double that of the original system and had an 80-fold induction (Figs. 3.5a, 3.4a). Gene expression also maintained a steady plateau at high inducer concentrations (Fig. 3.5a). Increasing expression levels at maximum induction for the 6k system required one of the weakest promoters in the Anderson library, but basal expression increased because *lacI* expression was too low (Fig. 3.4b).

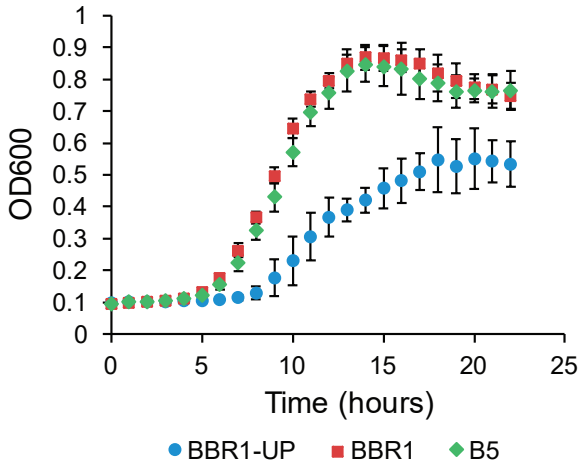


Figure 3.3: **Growth curve of *P. putida* KT2440 replicating broad-host range plasmids.** *P. putida* KT2440 cells replicating pBbB1k-GFPuv (BBR1-UP), pBBR1k-GFPuv (BBR1), or pBb(B5)1k-GFPuv (B5) grown at 30 °C, shaking in LB in a Tecan Infinite M200 plate reader.

Of the remaining promoter systems, induction was detected from P_{araB}, but P_{tet} did not appear to be functional (Fig. 3.5b, Figure A.2). P_{araB} provided titratable gene expression over a wide range of l-arabinose concentrations and had a 120-fold induction. However, maximum induction for this promoter system required up to 2% w/v l-arabinose instead of the 0.2% w/v required in *E. coli*. *P. putida* does not have an annotated transporter for l-arabinose [342], so we tested the effect of an arabinose transporter, AraE, on gene expression in *P. putida* cells expressing a fluorescent protein under P_{araB} (Figure A.3a). When *araE* was constitutively expressed with a weak promoter, maximum induction of P_{araB} was possible with lower inducer concentrations. Furthermore, *P. putida* demonstrated homogeneous expression of fluorescent protein with and without *araE expression*, as judged by flow cytometry (Figure A.3b).

3.4.2 Copy number of broad host range origins

During our previous work using the BBR1-UP origin, we found that it had poor gene expression and plasmid stability in *P. putida*. This work also resulted in the discovery of BBR1-B5, a mutant of BBR1-UP with an early stop codon in the *rep* gene, and we found that gene expression was more consistent when using the BBR1-B5 mutant and other BHR origins (Figure 3.2). We hypothesized that this mutation lowered the copy number of BBR1-UP, resulting in improved reliability in gene expression. We determined the copy number of five BHR origins in *E. coli* MG1655 and *P. putida* KT2440 using quantitative PCR (Table 3.2 and Table A.1). Three of the origins were variants of BBR1: wild-type BBR1, BBR1-UP, and BBR1-B5. We also included the origins RK2 and RSF1010 for comparison. For determining plasmid concentration in quantitative

Table 3.2: Copy number per chromosomal equivalent of broad host range plasmids in *E. coli* and *P. putida* KT2440 in stationary phase.

Origin	***E. coli* copy number**	***P. putida* copy number**
BBR1	3-4 [176]	30 ± 7
BBR1-UP	17-20 [176]	120 ± 20
BBR1-B5	2-4	70 ± 20
RK2	4-7 [343]	30 ± 10
RSF1010	7-10 [344]	130 ± 40

PCR, primers targeted the gene *gfpuv*. *LvaC*, which is located approximately 3,180,000 bp from the chromosome's replication origin, was the target for determining chromosome concentration. For all origins, the copy numbers were one order of magnitude higher in *P. putida* than in *E. coli* [343, 176, 344]. Both the BBR1 and RK2 origins, which are considered low-copy origins in *E. coli*, have around 30 copies per chromosomal equivalent in *P. putida*. The relative copy numbers of the origins are similar between *E. coli* and *P. putida*, except for RSF1010, which has a copy number similar to that of BBR1-UP in *P. putida*. We confirmed that the BBR1-B5 variant has a reduced copy number, which may explain the improved reliability in gene expression when using this origin.

3.4.3 λRed/Cas9 recombineering in *P. putida*

Our laboratory has had varying success generating knockouts in *P. putida* using existing methods for genomic deletions. Several knockouts were never constructed due to extremely low editing efficiency (Table 3.4). To remedy this issue, we developed a λRed/Cas9 recombineering protocol for generating genomic deletions in *P. putida*. This recombineering method uses an RK2-based plasmid that expresses *cas9* from a constitutive, weak promoter (pCas9). The λRed recombinases increase the efficiency of homologous recombination in *P. putida* [334], so we also included the $\alpha\beta\gamma$ operon expressed by the inducible P_{araB} promoter. While optimizing this system, we attempted Cas9 recombineering without expression of the $\alpha\beta\gamma$ operon, but these experiments did not yield any colonies (data not shown). The sgRNA is expressed constitutively from a second plasmid (pgRNAtet), which uses the high-copy BBR1-UP origin. The sgRNA is designed to tar-

Table 3.3: Summary of knockouts in *P. putida* KT2440 generated from λRed/Cas recombineering.

Gene	Function	Location on chromosome	Length of deletion (bp)	One-step Cas9		Two-step Cas9	
				CFU/mL	Fraction correct	CFU/mL	Fraction correct
lvaA	Phosphotransferase	3 001 201	1032	260	0/30	4300	27/30
pvdJ	Nonribosomal peptide synthetase	1 394 561	7848	570	01/30	4400	30/30
fpvA	Siderophore receptor	1 414 088	2391	70	3/30	3100	30/31
gcvP-I	Glycine dehydrogenase	5 052 324	2844	N/A	1/4	3000	27/31

get the sequence of the chromosome that will be removed upon successful generation of the knockout. To reduce the chances of off-target effects from Cas9/sgRNA expression, we used the CasOT off-target searching tool to identify any off-target sites for potential sgRNAs [345]. The repair template that integrates into the chromosome to generate the knockout is located on a suicide vector (pJOE) [158]. We designed the repair templates for each knockout so that there are 500-1000 bp of homology on either end of the gene of interest and so that the majority of the gene is removed except the start codon and the last 10-20 codons.

To generate each knockout, we first integrated pJOE into the *P. putida* KT2440 chromosome via electroporation into a strain replicating pCas9 (Fig. 3.6a). We used the resulting transformants to prepare electrocompetent cells, during which we induced λRed expression with l-arabinose. Upon introduction of the corresponding pgRNAtet via electroporation, the Cas9/sgRNA complex creates a double-stranded break in the chromosome and the λRed proteins repair the chromosome via homologous recombination. We used this two-step procedure to generate four knockouts in *P. putida* KT2440 with efficiencies around 85–100% (Table 2). After confirming each knockout, we cured out pCas9 and pgRNAtet by growing liquid cultures overnight in LB media with no antibiotics and plating the cells on LB agar. We screened individual colonies for loss of antibiotic resistance and found that pgRNAtet was easily cured out, but the majority of cells maintained pCas9.

To streamline this method, we attempted to generate knockouts with a one-step protocol that involved a co-transformation of pJOE and pgRNAtet (Fig. 3.6b). Unsurprisingly, the number of transformants and the overall editing efficiency decreased for all four knockouts (Table 3.3). The rate of detection for the knockout in the initial colony PCR screen

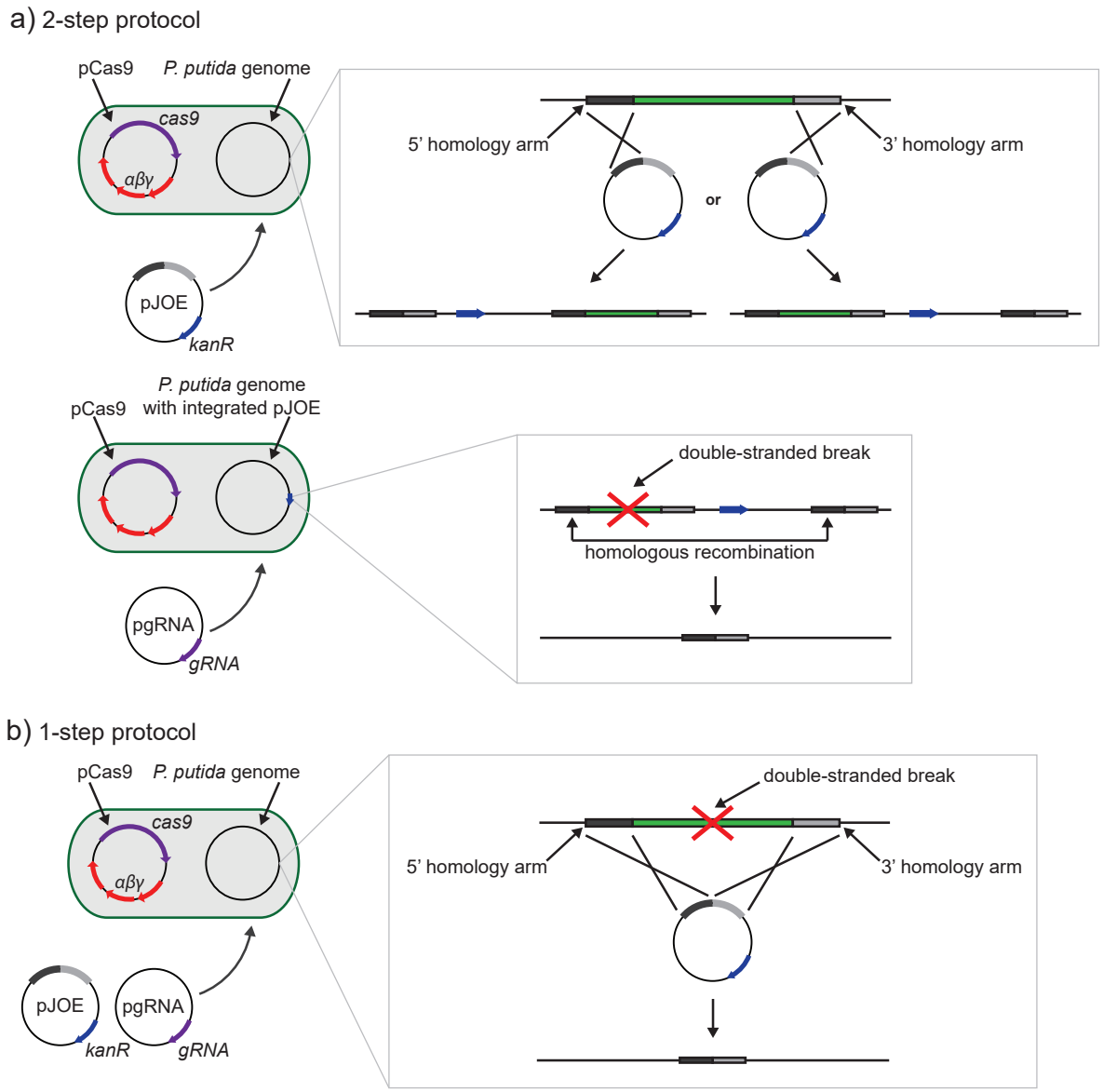


Figure 3.6: **Design and strategy for two-step and one-step λRed/Cas9 recombineering in *P. putida* KT2440. a)** The two-step protocol involves a chromosomal integration of pJOE, which carries the repair template for the desired knockout, followed by transformation of the corresponding pgRNA. **b)** The one-step protocol involves a co-transformation of pJOE and pgRNA, where pJOE is used to repair the double-stranded break created by Cas9/sgRNA.

Table 3.4: Comparison of different genome editing methods in *P. putida*.

Gene	1-step Cas9		2-step Cas9		5-FU Counterselection	
	Fraction correct	Percent correct	Fraction correct	Percent correct	Fraction correct	Percent correct
lvaA	0/30	0%	27/30	90%	29/30	97%
pvdJ	1/30	3%	30/30	100%	0/45	0%
fpvA	3/30	10%	30/31	97%	0/51	0%
gcvP-I	1/4	25%	27/31	87%	0/30	0%

was similar to that of the two-step protocol, but after streaking the transformants on selective media for single colonies, we found that most samples lost the designed deletion as well as the wild-type sequence (Figure A.5).

3.5 Discussion

The inducible promoter systems that we tested are native to *E. coli* [346, 347, 348], so one should expact that they behave differently in *P. putida*, an organism that does not catabolize l-arabinose or lactose [349, 342]. Expression from P_{araB} in *E. coli* results in a mixed population at low inducer concentrations because of an "autocatalytic" induction mechanism involving AraE [350]. Constitutive expression of *araE* in *E. coli* enables homogeneous expression from P_{araB} [351]. Adapting this strategy to *P. putida* allowed maximum induction at l-arabinose concentrations lower than 2% (Figure S4a). That said, the concentrations needed for maximum induction were less than 0.01% w/v in the presence of *araE* expression, making it more difficult to titrate gene expression from P_{araB} and suggesting that *araE* expression is too high. Optimizing *araE* expression may require lowering the gene copy via chromosomal integration or using an alternative transporter with a lower affinity to l-arabinose [352]. IPTG transport does not appear to be inhibitory to inducing P_{lac} promoters in *P. putida* because maximum induction can be easily achieved using concentrations around 1 mM. For the 1k and 6k promoter systems, high expression levels of *lacI* from the P_{lacIq} promoter reduced gene expression at maximum induction compared to the modified systems. The steady decrease in expression in the induction curve for the 1k system at high inducer concentrations suggests that the reduced expression was due to overproduction of protein, which is similar to the low gene expression observed from strong promoters on high-copy plasmids (Fig. 3.5a, Figure 3.2a). Increasing gene expression from the 6k system required lower *lacI* expression, but *lacI* expression for the 6k system must be further optimized so that basal (a.k.a. "leaky") expression is lower. We did not detect any induction from the P_{tet}/aTc inducible promoter system (Figure A.2). Reducing expression of the TetR repressor could be a successful

strategy for improving induction from this promoter system in *P. putida*. Other reports have shown considerable induction from P_{tet} [185, 176], so this system requires further investigation to determine its utility in *P. putida*. RBS and 5' UTR sequences also appear to be important for gene expression in *P. putida*, and differences in these sequences may explain the variability of P_{tet} in different reports [315, 156].

It is important to consider gene copy number when expressing heterologous genes in bacteria. Gene copy number in *E. coli* can range from one on bacterial artificial chromosomes (BACs) to 500 on high-copy pUC vectors [353, 354]. Protein production from high-copy plasmids can lead to a metabolic burden and reduce cell growth rate [355], so a complete genetic toolbox for any organism should have low-copy options available. Our quantitative PCR results show that none of the origins tested can be considered low-copy. The inducible promoter systems encoded on BBR1-based plasmids had poor induction due to overexpression of the transcription factors. Lowering the copy number for these promoter systems may improve gene expression by reducing the intracellular concentration of the transcription factors. For example, the copy number of RK2 is dependent on intracellular concentrations of its replication protein, TrfA, at low concentrations, so generating low-copy BHR origins in *P. putida* may be possible by lowering expression of the replication proteins [356]. *Pseudomonas*-specific origins that have been shown to be low-copy in other pseudomonads may also have similar properties in *P. putida* [357]. A guaranteed option for low-copy heterologous gene expression in *P. putida* is through chromosomal integrations. Either the transcription factor or the entire promoter system could be encoded on the chromosome, depending on the promoter strength. The BBR1 and RK2 origins offer the lowest plasmid copy number of BHR origins commonly used in *P. putida*, so the probability of overproducing proteins could be limited by using these origins with weak promoters and a single gene or small operon. Even though BBR1-UP leads to unreliable protein production, the efficacy of Cas9/sgRNA activity in our λRed/Cas9 recombineering protocol shows that this origin can be used to reliably express sgRNAs and potentially other non-protein gene products.

The near-100% editing efficiency of Cas9-assisted recombineering in *P. putida* relies

on the counterselection provided by the Cas9/sgRNA complex and the presence of the λRed proteins to facilitate homologous recombination. The most efficient techniques for editing the genome of *P. putida* involve a single-crossover recombination event between the chromosome and a suicide vector [158, 337]. The counterselections for these methods do not select against the removal of the wild-type sequence, but rather the removal of the suicide vector that can recombine back to the wild-type sequence, resulting in a theoretical editing efficiency of 50%. In practice, deletions may have efficiencies well below 10% and are not reliably identified in a low-throughput genetic screen (Table 3.4). Existing methods rely on native recombination pathways in *P. putida*, and therefore they require overnight incubations to resolve the suicide vector. If a knockout causes a growth defect, then any cells that maintain the wild-type sequence will take up a larger percentage of the population as cells replicate, further decreasing the editing efficiency. Designing a sgRNA to target the wild-type sequence for a double-stranded break prevents wild-type cells from outgrowing knockout mutants, improving the editing efficiency. The λRed genes can also mediate homologous recombination more rapidly than *P. putida*'s native recombination pathways, so the cells can resolve the suicide vector from the chromosome within a short 2-h recovery rather than an overnight incubation. That said, the lower editing efficiencies of the one-step protocol suggest that the λRed proteins can facilitate other recombination events in absence of the repair template. These events persist because *P. putida* is polyploid, and a correctly edited copy of the chromosome can coexist with other incorrect copies until they segregate toward a common DNA sequence. If the repair template was only incorporated into a fraction of the copies of the chromosome, then it is unlikely to survive segregation. This problem does not occur for the two-step protocol because the repair template is already incorporated into every copy of the chromosome after transformation of the pJOE suicide plasmid.

Due to the versatility of Cas9 and BHR origins, this recombineering method could be used to generate deletions in other pseudomonads. The genome-editing technique based on the I-SceI endonuclease developed by de Lorenzo and colleagues originally for *P. putida* was also used successfully to generate deletions in *Pseudomonas syringae* and

Pseudomonas fluorescens [337]. Homologous recombination with the λRed proteins and other related recombination systems has also been demonstrated in multiple pseudomonads [358, 359, 360]. There are several Cas9-assisted recombineering systems available for *E. coli* that could be adapted for editing the genomes of other Gram-negative bacteria by replacing the *E. coli* plasmid origins with BHR origins [239, 361].

We demonstrated this technique's utility in generating knockouts, but it could easily be adapted for integrating heterologous DNA into the chromosome by including the sequence of interest in the repair template between the regions of homology. Altenbuchner and colleagues integrated a pathway for vanillin production onto the chromosome of *P. putida* using the 5-FU counterselection developed originally for gene deletions [158, 362]. There are several efficient chromosomal integration systems developed for *P. putida* [179, 201]. However, these methods are site specific, so they cannot be used to integrate pathways onto multiple loci or modify endogenous pathways. The latter feature is especially necessary for activating cryptic gene clusters in pseudomonads by modifying regulatory elements directly on the chromosome [363, 364].

The results described here establish a common set of genetic tools for use in *P. putida*. The P_{lac} family of promoters and P_{araB} are commonly used in *E. coli*, but they behave considerably differently in *P. putida*. Gene expression from P_{araB} is the most similar between the two bacteria, but maximum induction in *P. putida* requires an order of magnitude higher concentration of inducer than in *E. coli*. Reducing the level of expression of the LacI repressor improved induction from IPTG-inducible promoters, and this strategy could be used to improve other promoter systems with a poor fold induction, such as the P_{tet} system. However, other factors effecting gene expression, such as RBS and 5' UTR sequence, should also be considered when optimizing gene expression in *P. putida*. The copy numbers of BHR origins are significantly higher in *P. putida*, which leads to plasmid instability and unreliable protein production from constitutive and inducible promoters. This fact limits the availability of plasmid-based genetic tools for *P. putida*; therefore, chromosomal integration should be considered when expressing heterologous genes for its relatively high stability and lower copy number. Taking advantage of CRISPR/Cas9 for

editing the genome allows for much more efficient strain development over alternative methods. Shown here as a tool for generating knockouts, λRed/Cas9 recombineering can also be adapted for introducing heterologous genes virtually anywhere on the chromosome or modifying endogenous pathways, providing an alternative platform for metabolic engineering when plasmid-based gene expression may not be optimal.

3.6 Acknowledgments

This study was supported by research grants from the National Science Foundation (CBET-114678, MCB-1716594) and US-AID (PEER 3-195). T.B.C. and D.K.C. are recipients of NIH Biotechnology Training Program Fellowships (NIGMS 5 T32 GM08349). J.M.R. was supported by an NSF Graduate Research Fellowship (DGE1256259). S.A.L. is the recipient of a fellowship from the Promega Corporation through the Dane County Youth Apprenticeship Program in Biotechnology. The authors would like to thank Dr. Yalun Arafin and Dr. Fransiskus Ivan for their assistance with the project.

Chapter 4

Heterologous expression of biosynthetic gene clusters in *P. putida*

Authors and Contributors:

- **Taylor B. Cook** Designed and constructed production strains, determined heterologous product titers, purified glidobactin A standard.
- **Dr. Tyler B. Jacobson** Performed LC-MS analysis.
- **Maya V. Venkataraman** Constructed production strains and quantified prodigiosin titers.
- **Dr. Heike Hofstetter** Performed ^{1}H-NMR analysis.
- **Erin A. Conley** Performed capreomycin purifications and disk diffusion assays.
- **Dr. Daniel Amador-Noguez** Provided guidance, edited, and reviewed the manuscript.
- **Dr. Michael G. Thomas** Provided guidance, edited, and reviewed the manuscript.
- **Dr. Brian F. Pfleger** Provided guidance, edited, and reviewed the manuscript.

4.1 Abstract

Polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS) comprise biosynthetic pathways that provide access to diverse, often bioactive natural products. Metabolic engineering can improve production metrics to support characterization and drug-development studies, but often native hosts are difficult to genetically manipulate and/or culture. For this reason, heterologous expression is a common strategy for natural product discovery and characterization. Many bacteria have been developed to express heterologous biosynthetic gene clusters (BGCs) for producing polyketides and nonribosomal peptides. In this article, we describe tools for using *Pseudomonas putida*, a Gram-negative soil bacterium, as a heterologous host for producing natural products. Pseudomonads are known to produce many natural products, but *P. putida* production titers have been inconsistent in the literature and often low compared to other hosts. In recent years, synthetic biology tools for engineering *P. putida* have greatly improved, but their application towards production of natural products is limited. To demonstrate the potential of *P. putida* as a heterologous host, we introduced BGCs encoding the synthesis of prodigiosin and glidobactin A, two bioactive natural products synthesized from a combination of PKS and NRPS enzymology. Engineered strains exhibited robust production of both compounds after a single chromosomal integration of the corresponding BGC. Next, we took advantage of a set of genome-editing tools to increase titers by modifying transcription and translation of the BGCs and increasing the availability of auxiliary proteins required for PKS and NRPS activity. Lastly, we discovered genetic modifications to *P. putida* that affect natural product synthesis, including a strategy for removing a carbon sink that improves product titers. These efforts resulted in production strains capable of producing 1.1 g/L prodigiosin and 470 mg/L glidobactin A.

4.2 Introduction

For most of the last century, natural products have been an essential source for the discovery of novel drugs for treating human diseases [365]. Natural products are secondary metabolites produced by living organisms that are often unnecessary for growth but provide advantages to the host organism in certain environments. These compounds evolved for specific interactions with biomolecules, and as a result, they often exhibit medically relevant properties, such as antibiotic or anti-cancer activities [4]. Two of the most intriguing classes of natural products are polyketides and nonribosomal peptides. These compounds are often synthesized in an assembly line fashion by modular enzymes called type I polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS) [366]. Multiple enzymatic domains are encoded in single large peptides, and chemical diversity is achieved through the addition, deletion, and modification of individual domains [367]. The biosynthesis potential of these assembly-lines has long been touted as a panacea for accessing diverse molecular structures, but this potential remains mostly unrealized due to challenges in engineering enzymes to meet required synthesis metrics.

Bacteria, particularly Actinomycetes, are a rich source of natural products, including polyketides and nonribosomal peptides. The number of putative biosynthetic gene clusters (BGCs) has accelerated with the rise of next-generation sequencing [368], as has the number without a validated product [369]. Connecting putative BGCs to purified compounds requires several challenges to be overcome. Putative BGC-containing hosts can be difficult to cultivate in a laboratory setting, and even if cultivation is achieved, the conditions required for inducing production of the desired natural product may be unknown [152]. Alternatively, BGCs can be transferred to a heterologous host and expressed from modern synthetic biology vectors. This strategy by-passes growth and regulation problems associated with the native host by using a microorganism with well-characterized genetic tools that also grows robustly in laboratory conditions [84, 370, 371]. Once a BGC is introduced into a heterologous host, novel compounds can be isolated for structure elucidation and titers can be improved with common metabolic engineering strategies to enable *ex vivo* bioactivity assays [372, 309]. Most heterologous BGC expression projects

use *Streptomyces* species as hosts because actinomycetes are the most abundant source of bacterial natural products [25, 173]. However, *Streptomyces* hosts are more difficult to genetically modify than the synthetic biology workhorse, *Escherichia coli*. Several groups have developed *E. coli* strains for polyketide and nonribosomal peptide production [87, 248]. However, functional expression of these modular enzymes in *E. coli* is often difficult to achieve, so researchers have sought alternative hosts for producing heterologous polyketides and nonribosomal peptides [260, 373, 52, 371].

Pseudomonas putida KT2440, a Gram-negative soil microbe, has become a popular metabolic engineering chassis in recent years, mostly due to its ability to utilize lignin-derived compounds and its tolerance to organic solvents and stresses associated with industrial-scale cultivations [219, 203]. *P. putida* also has several traits that make it an attractive alternative host for producing natural products. Its genome has a relatively high GC content (61.5%) similar to that of actinomycete-derived natural product BGCs. *P. putida* natively expresses a promiscuous phosphopantetheinyl transferase (PPTase) capable of activating heterologous acyl- and peptidyl-carrier proteins, and it has excellent growth properties amenable to large-scale cultivations [168, 137]. There has also been an explosion in synthetic biology tools developed for *P. putida*, including genetic parts for expressing heterologous proteins and constructing genomic edits [196, 179, 167]. Despite these advantages, *P. putida* has performed inconsistently in the literature as a heterologous host for producing polyketides and nonribosomal peptides, where it either achieves titers on par or greater than the native host [374, 300] or it performs poorly compared to alternative heterologous hosts [301, 309]. Most reports have investigated the effect of transcriptional control on heterologous BGC expression in *P. putida* [276, 375, 213]. Other variables that affect production titers, such as GC content of the BGC, translational efficiency, and specific activity of PKSs and NRPSs, have not been systematically explored in *P. putida*.

We selected prodigiosin and glidobactin A, two natural products natively produced by Gram-negative bacteria, to serve as model products for exploring factors that influence biosynthesis in *P. putida*. The BGCs for prodigiosin and glidobactin A biosynthesis both

contain PKS and NRPS components, and both compounds possess medically relevant properties, such as anti-cancer activity [303, 306, 376, 377, 304]. Prodigiosin is a tripyrrole secondary metabolite synthesized by many strains of *Serratia marcescens* [378]. It is a red pigment, providing an easily detectable phenotype for screening libraries for factors that increase prodigiosin production. The highest reported production titer of prodigiosin from a heterologous host is 150 mg/L from *P. putida* [276]. BGCs for glidobactin synthesis have been identified and characterized from *Schlegelella brevitalea* DSM7029 and *Photorhabdus luminescens* subsp. *laumondii* TT01 [375, 379]. The two glidobactin BGCs differ in their GC content and in the presence of different genes required for biosynthesis, providing a case study that can reveal variables important for heterologous production in *P. putida*. The highest heterologous titer for this molecule was achieved after expression of the BGC from *P. luminescens* in *Xenorhabdus doucetiae*, which yielded 177 mg/L glidobactin A while production from *P. putida* was not detected [309].

In this work, we designed a genome editing pipeline for integrating and expressing heterologous BGCs in *P. putida*. The BGCs responsible for prodigiosin (from *S. marcescens* ATCC274) and glidobactin (from *S. brevitalea* DSM7029 and *P. luminescens* subsp. *laumondii* TT01) biosynthesis serve as models for investigating the effects of transcription and translation on production titers. Focusing on glidobactin A biosynthesis, we compare strategies for improving the functional expression of modular PKSs and NRPSs. We also take advantage of the visible phenotype associated with prodigiosin production to devise a screen for discovering mutants with improved production titers of heterologous products. Through these efforts, we demonstrate that *P. putida* is capable of heterologous production titers in 100 mg - 1 g/L quantities from all three BGCs.

4.3 Materials & Methods

4.3.1 Plasmids, bacterial strains, and growth conditions

The bacterial strains used in this study are shown in Table 4.1. Vectors constructed for various genome editing methods are available through Addgene. All *E. coli* strains were grown in LB medium at 37°C. *P. putida* KT2440 and derivative strains were grown in LB medium at 30°C. LB medium was supplemented with kanamycin (50 μg/mL, Kan50), gentamycin (30 μg/mL, Gent30), tetracyline (10 μg/mL, LB Tet10 for *E. coli*, 25 μg/mL, Tet25 for *P. putida*), and irgasan (25 μg/mL, Irg25) when necessary. *Serratia marcescens* ATCC274 and *Photorhabdus luminescens* subsp. *laumondii* TT01 were grown in LB medium at 30°C. *Schlegelella brevitalea* sp. nov. DSM7029 was grown at 30°C on CY-Agar (3g/L casitone, 1.4 g/L $CaCl_2 \bullet 2H_2O$, 1.0 g/L yeast extract, 15 g/L agar) and in CYCG medium (6 g/L casitone, 1.4 g/L $CaCl_2 \bullet 2H_2O$, 2.0 g/L yeast extract, 25 g/L glycerol) [380]. All liquid cultures were shaken at 250 rpm during incubation.

For secondary metabolite production, *P. putida* was grown in RK medium [381]. The medium was prepared by mixing 13.3 g potassium phosphate monobasic (KH2PO4), 4.0 g ammonium phosphate $(NH_4)_2PO_4$, 1.7 g citric acid, 0.1 g Fe(III) ammonium citrate, and 25 g glycerol or 25 g D-glucose to 800 mL deionized/distilled water. To this solution, 10 mL of sterile 100X RK batch trace minerals and 10 mL of sterile 120 g/L $MgSO_4 \bullet 7H_2O$ were added. The pH was adjusted to 6.7 with 5M NaOH and the volume was adjusted to 1 L. The 100X trace minerals solution was prepared by adding 0.42 g EDTA, 0.125 g $CoCl_2 \bullet 6H_2O$, 0.75 g $MnCl_2 \bullet 4H_2O$, 0.06 g $CuCl_2$, 0.15 g H_3BO_3, 0.125 g $Na_2(MoO_4)$ $\bullet 2H_2O$, and 0.65 g $Zn(CH_3COO)_2 \bullet 2H_2O$ (zinc acetate) to 300 mL deionized/distilled water and adjusting to a final volume of 500 mL. All media components and the final media formulation were sterilized by filtration. RK medium was supplemented with kanamycin (25 μg/mL, Kan25) when necessary. Some early production experiments also used Terrific Broth, supplemented with 5 g/L or 25 g/L glycerol.

Table 4.1: Strains used in this work.

Strain	**Genotype and relevant characteristics**	**Source**
Escherichia coli		
DH5α	F- Φ80*lacZ*ΔM15 Δ*(lacZYA-argF)* U169 *reacAI endAI hsdR17* (r_{k-} m_{k+}) *phoA supE*44 *thi*-1 *gyrA*96 *relA*1 λ-	Invitrogen
HB101	F- *thi*-1 *hsd*20 (r_{B-} m_{B-}) *supE*44 *recA*13 *ara*-14 *leuB*6 *proA*2 *lacY*1 *galK*2 *rpsL*20 (str^r) *xyl*-5 *mtl*-1	CGSC
GB2005	DH10β Δ*fhuA* Δ*ybcC* Δ*recET*	[84]
GB05-dir	GB2005 Δ*ybcC*::P_{BAD}*ETgA*	[84]
S17-1 λpir	TpR SmR *recA thi pro hsdR*-M+RP4: 2-Tc:Mu: Km Tn7 λ*pir*	[230]
BW29427	RP4-2 (TetS *kan*1360::FRT) *thrB*1004 Δ*lacZ*58(M15) Δ*dapA*1341::[*erm pir*$^+$] *rpsL*(str^r) *thi- hsdS- pro-*	CGSC
Serratia marcescens		
ATCC274	Wild-type	NRRL B-23389
Photorhabdus luminescens		
subsp. *laumondii* TT01	Wild-type	DSMZ 15139
Schlegelella brevitalea		
sp. nov.	Wild-type	DSMZ 7029
Pseudomonas putida		
KT2440	Wild-type	ATCC 47054
KTU	KT2440 Δ*upp*	[158]
Δ*pvdL*	KT2440 Δ*upp* Δ*pvdL*	[225]
pvd-	KT2440 Δ*upp* Δ*pvdL* Δ*pvdIJD*	This work
GFP01	pvd- Δ*pvdIJD*::P_{rha}-*gfpuv*	This work
GFP02	pvd- Δ*pvdIJD*::P_{rha}-*sfgfp*	This work
GFP03	pvd- Δ*pvdIJD*::P_{trc}-*gfpuv*	This work
GFP04	pvd- Δ*pvdIJD*::P_{trc}-*sfgfp*	This work
PIG01	pvd- Δ*pvdIJD*::P_{rha}-*pig*	This work
PIG02	pvd- Δ*pvdIJD*::P_{trc}-*pig*	This work
PIG03	pvd- Δ*pvdIJD*::P_{rha}-*pig* Δ*phaC1ZC2*	This work
PIG04	pvd- Δ*pvdIJD*::P_{rha}-*pig* Δ*glpR*	This work
PIG05	pvd- Δ*pvdIJD*::P_{rha}-*pig* Δ*glpR* Δ*phaC1ZC2*	This work
PIG06	pvd- Δ*pvdIJD*::P_{trc}-*pig* Δ*glpR*	This work
PIG07	pvd- Δ*pvdIJD*::P_{trc}-*pig* Δ*glpR* Δ*phaC1ZC2*	This work
PIG08	pvd- Δ*pvdIJD*::P_{rha}-*pig* Δ*cyoABCDE*	This work
PIG09	pvd- Δ*pvdIJD*::P_{trc}-*pig* Δ*cyoABCDE*	This work
LUM01	pvd- Δ*pvdIJD*::P_{rha}-*lum*	This work
LUM02	pvd- Δ*pvdIJD*::P_{trc}-*lum*	This work
LUM03	pvd- Δ*pvdIJD*::P_{rha}-*lum PP1183*::J23111	This work
LUM04	pvd- Δ*pvdIJD*::P_{trc}-*lum PP1183*::J23111	This work
LUM05	pvd- Δ*pvdIJD*::P_{trc}-*lum* Δ*pvdL*::P_{tac}-*sfp*	This work
LUM06	pvd- Δ*pvdIJD*::P_{trc}-*lum-SNP*	This work
LUM07	pvd- Δ*pvdIJD*::P_{trc}-*lum-ATG*	This work
LUM08	pvd- Δ*pvdIJD*::P_{trc}-*lum-ATG* Δ*pvdL*::P_{tac}-*sfp*	This work
LUM09	pvd- Δ*pvdIJD*::P_{rha}-*lum* Δ*glpR* Δ*phaC1ZC2*	This work
LUM10	pvd- Δ*pvdIJD*::P_{trc}-*lum-ATG* Δ*pvdL*::P_{tac}-*sfp* Δ*glpR* Δ*phaC1ZC2*	This work
LUM11	pvd- Δ*pvdIJD*::P_{rha}-*lum* ΔPP_3808	This work
GLB01	pvd- Δ*pvdIJD*::P_{rha}-*glb*	This work
GLB02	pvd- Δ*pvdIJD*::P_{trc}-*glb*	This work
GLB03	pvd- Δ*pvdIJD*::P_{trc}-*glb*V2	This work
GLB04	pvd- Δ*pvdIJD*::P_{trc}-*glb*V2 Δ*PP3808*	This work
GLB05	pvd- Δ*pvdIJD*::P_{trc}-*glb*V2 Δ*PP3808*::J23104-*glbE*	This work
GLB06	pvd- Δ*pvdIJD*::P_{rha}-*glb* Δ*PP3808*::J23104-*glbE* Δ*pvdL*::P_{tac}-*sfp*	This work
GLB07	pvd- Δ*pvdIJD*::P_{trc}-*glb*V2-*SNP* Δ*PP3808*::J23104-*glbE*	This work
GLB08	pvd- Δ*pvdIJD*::P_{trc}-*glb*V2-*ATG* Δ*PP3808*::J23104-*glbE*	This work
GLB09	pvd- Δ*pvdIJD*::P_{trc}-*glb*V2-*ATG* Δ*PP3808*::J23104-*glbE* Δ*pvdL*::P_{tac}-*sfp*	This work
ANT01	pvd- Δ*pvdIJD*::P_{rha}-*ant*	This work
CMN01	pvd- Δ*pvdIJD*::P_{rha}-*cmn*	This work
ARY01	pvd- Δ*pvdIJD*::P_{rha}-*ary1* Δ*pvdL*::P_{tac}-*ary2*	This work

4.3.2 Plasmid construction

Plasmids used in this work are shown in Table 4.2. Most plasmids described in this work were constructed using Gibson assembly as described previously [382]. The plasmids pNVLT-pvdIJD-Prha-pig, pNVLT-pvdIJD-Prha-lum, and pNVLT-Prha-glb were constructed using RecET direct cloning [84, 80]. Genomic DNA was purified from *S. marcescens*, *P. luminescens*, and *S. brevitalea* using phenol:chloroform:isoamyl alcohol extraction as reported previously [341, 225]. Pure genomic DNA was prepared from 10mL of liquid culture, and the extraction protocol was scaled up accordingly. To "release" the BGC of interest, 20 μg of genomic DNA was digested with up to two restriction enzymes in 400-μL reactions and then purified by ethanol precipitation. Capture vectors were linearized by restriction digest in 100-μL reactions containing 2 μg of plasmid DNA and then purified by gel extraction. Restriction digest reactions were incubated at 37°C for 2 hours. To prepare DNA for construction of pNVLT-pvdIJD-Prha-pig, *S. marcescens* genomic DNA was digested with AflII and the capture vector, pNVLT-pvdIJD-Prha-prepig, was digested with KpnI-HF. For pNVLT-pvdIJD-Prha-lum, *P. luminescens* genomic DNA was digested with PacI and MluI and the capture vector, pNVLT-pvdIJD-Prha-prelum, was digested with BsaI-HFv2. For pNVLT-pvdIJD-Prha-glb, *S. brevitalea* genomic DNA was digested with DraI and AvrII, and the capture vector, pNVLT-pvdIJD-Prha-preglb, was digested with BsaI-HFv2.

To prepare competent cells for RecET direct cloning, 5 mL of fresh LB were inoculated with 150-μL of overnight culture of *E. coli* GB05-dir. Cultures were incubated at 37°C for about 2 hours, and then induced with 100 μL 20% (w/v) L-arabinose. Growth continued for another 45 minutes and then the cultures were placed on ice. For every aliquot of competent cells needed, 1 mL of culture was centrifuged for 30 seconds at 11,000g and washed with 1 mL 10% (v/v) glycerol. The samples were washed with 1 mL 10% glycerol two more times, and then resuspended in 20 μL 10% glycerol. At least 500 ng of linearized vector and 5 μg digested genomic DNA were added to each aliquot of competent cells to a total approximate volume of 50 - 60 μL. The competent cells were then added to chilled 1-mm electroporation cuvettes and electroporated with a voltage of 1.8kV, yielding time

constants of 3 - 4 ms. After electroporation, cells were immediately mixed with 1 mL SOC media and incubated at 37°C for 90 minutes. All of the recovered cells were plated on LB Kan25 supplemented with 0.5% (w/v) D-glucose. Plates were incubated overnight at 37°C.

4.3.3 Genome editing in *P. putida*

For the 2-step λRed/Cas9 recombineering protocol, pNVLTv2 integration vectors were introduced into *P. putida* containing pCas9 by tri-parental conjugation, as described previously [383]. The helper strain was *E. coli* HB101 containing pRK600 and the donor strain was *E. coli* DH5α or GB2005 containing pNVLTv2. Conjugations were plated on LB Gent30, Kan50, Irg25. Conjugants were inoculated in LB Gent30, Kan50 and grown overnight at 30°C. The next day, the λRed genes on pCas9 were induced by adding 150 μL 20% L-arabinose to the cultures and incubating for another 45 minutes. To prepare electrocompetent cells, 500 μL of culture was washed twice with 1 mL 10% glycerol and resuspended in 50 μL 10% glycerol. About 100 ng pgRNAtet were added to each sample and the samples were added to a chilled 1-mm electroporation cuvette. Samples were electroporated with a voltage of 1.8kV and allowed to recover in 1 mL LB for 3 hours at 30°C. The recovered cells were selected on LB Gent30, Tet25 plates covered with aluminum foil to limit light exposure. Some electroporations were completed by washing and resuspending with 300 mM sucrose instead of 10% glycerol.

For oligo recombineering, *P. putida* containing pCas9-beta or pCas9-beta-mmr was used to prepare electrocompetent cells as described above. These cells were transformed with approximately 100 ng pgRNAkan and 1 μL of 1 μM oligo by electroporation. Cells were recovered in 1 mL LB and selected on LB Gent30, Kan50. Oligos used for this protocol are shown in Table 4.3. Genome editing with 5-FU counterselection was completed as described previously [158]. Transformants from all three methods were screened for the desired knockout using colony PCR with primers flanking the gene of interest.

Table 4.2: Plasmids used in this work

Plasmid	**Description**	**Source**
pBAM1	*tnpA* AmpR KanR *ori*R6K	[230]
pJOE	*upp* (from *P. putida* KT2440), KanR, ColE1 origin	[158]
pSEVA2514-rec2-mutL$_{E36K}^{PP}$	Oligo recombineering and mismatch repair suppression; source of $mutL_{E36K}$	[200]
pCas9	Cas9 and λRed expression plasmid for recombineering	[225]
pCas9-beta	Cas9 and β expression plasmid for single-stranded recombineering	This work
pCas9-beta-mmr	Cas9, β, and $mutL_{E36K}$ expression plasmid for single-stranded recombineering and mismatch repair supression	This work
pNVL	pJOE derivative, removed *upp*, replaced ColE1 origin with p15A origin	This work
pNVLTv2	pNVL derivative, removed AmpR, replaced *oriT* sequence	This work
pJOE-pvdL	Integration vector for deletion of *pvdL* (PP4243)	This work
pNVL-pvdIJD	Integration vector for deletion of *pvdIJD* (PP4219-4221)	This work
pNVLTv2-pvdIJD-Prha-GFPuv	Integration vector for insertion of P_{rha}-*GFPuv* into *pvdIJD* locus	This work
pNVLTv2-pvdIJD-Prha-sfGFP	Integration vector for insertion of P_{rha}-*sfGFP* into *pvdIJD* locus	This work
pNVLTv2-pvdIJD-Ptrc-GFPuv	Integration vector for insertion of P_{trc}-*GFPuv* into *pvdIJD* locus	This work
pNVLTv2-pvdIJD-Ptrc-sfGFP	Integration vector for insertion of P_{trc}-*sfGFP* into *pvdIJD* locus	This work
pNVLTv2-pvdIJD-Prha-base	Empty integration vector with P_{rha}	This work
pNVLTv2-pvdIJD-Ptrc-base	Empty integration vector with P_{trc}	This work
pNVLTv2-pvdIJD-Prha-prepig	Capture vector for *pig* cluster with P_{rha} expression	This work
pNVLTv2-pvdIJD-Ptrc-prepig	Capture vector for *pig* cluster with P_{trc} expression	This work
pNVLTv2-pvdIJD-Prha-prelum	Capture vector for *lum* cluster with P_{rha} expression	This work
pNVLTv2-pvdIJD-Prha-preglb	Capture vector for *glb* cluster with P_{rha} expression	This work
pNVLTv2-pvdIJD-Prha-pig	Integration vector for insertion of *pig* cluster with P_{rha} expression into *pvdIJD* locus	This work
pNVLTv2-pvdIJD-Prha-lum	Integration vector for insertion of *lum* cluster with P_{rha} expression into *pvdIJD* locus	This work
pNVLTv2-pvdIJD-Prha-glb	Integration vector for insertion of *glb* cluster with P_{rha} expression into *pvdIJD* locus	This work
pNVLTv2-pvdIJD-Prha-ant	Integration vector for insertion of *ant* cluster with P_{rha} expression into *pvdIJD* locus	This work
pNVLTv2-pvdIJD-Ptrc-pigA	Integration vector for swapping P_{rha} with P_{trc} in strains with P_{rha}-*pig* insertion	This work
pNVLTv2-pvdIJD-Ptrc-lumA	Integration vector for swapping P_{rha} with P_{trc} in strains with P_{rha}-*lum* insertion	This work
pNVLTv2-pvdIJD-Ptrc-glbB	Integration vector for swapping P_{rha} with P_{trc} in strains with P_{rha}-*glb* insertion	This work
pNVLTv2-pvdIJD-Ptrc-glbBV2	Integration vector for swapping P_{rha} with P_{trc} in strains with P_{rha}-*glb* insertion, uses RBSV2	This work
pNVLTv2-phaC1ZC2	Integration vector for deletion of *phaC1ZC2*	This work
pNVLTv2-PP3808-J23104	Integration vector for insertion of J23104 promoter upstream of PP3808	This work
pNVLTv2-PP3808-glbE	Integration vector for replacing PP3808 with *glbE*	This work
pNVLTv2-PP1183-J23111	Integration vector for inserting J23111 promoter upstream of PP1183	This work
pNVLTv2-pvdL-Ptac-sfp	Integration vector for inserting Ptac-sfp expression construct	This work
pNVLTv2-cyo	Integration vector for deleting *cyoABCDE* operon	This work
pgRNAtet-pvdI-BBR1-UP	Guide RNA plasmid targeting *pvdI*, TcR	This work
pgRNAtet-pvdIJD-BBR1-UP	Guide RNA plasmid targeting Δ*pvdIJD* sequence, TcR	This work
pgRNAtet-rhaRS	Guide RNA plasmid targeting *rhaRS*, TcR	This work
pgRNAtet-phaZ	Guide RNA plasmid targeting *phaZ*, TcR	This work
pgRNAtet-PP3808	Guide RNA plasmid targeting PP3808, TcR	This work
pgRNAkan-PP3808	Guide RNA plasmid targeting PP3808, KanR	This work
pgRNAtet-PP1183-P	Guide RNA plasmid targeting PP1183 promoter region, TcR	This work
pgRNAkan-sotB	Guide RNA plasmid targeting *sotB*, KanR	This work
pgRNAkan-glpR	Guide RNA plasmid targeting *glpR*, KanR	This work
pgRNAkan-phaG	Guide RNA plasmid targeting *phaG*, KanR	This work
pBBR8k-glbE	Expression plasmid, pBAD, *glbE*, KanR, BBR1 origin	This work

Table 4.3: DNA oligos used for ssDNA recombineering and RT-PCR in this work. Nucleotides with "*" between denote phosphorothiate linkages. The capitalized nucleotides in oligoes rTBC0971 and rTBC0972 indicate the nucleotides that differ from wild-type.

Oligo	Description	Sequence (5' to 3')
rTBC0623	*sotB* deletion oligo, lagging strand	AGCCTCTACTTGACCGACAAGGGGCCT GAGCGGCCTCGGCGGCAACAGGTCGCA CGCCAGCCACCCGCTTGCAGGGGGCC
rTBC0624	*sotB* deletion oligo, leading strand	GGCCCCCTGCAAGCGGGTGGCTGGCGT GCGACCTGTTGCCGCCGAGGCCGCTCA GGCCCCTTGTCGGTCAAGTAGAGGCT
rTBC0637	*glpR* deletion oligo, lagging strand	GCAGAATTTGTCCGGCAGCCCCAAAGG ACCGCCCATGAATTTGAACCAGAACAA GATCAGGCTTGAGGTGGTCTGAGGGC
rTBC0658	*phaG* deletion oligo, lagging strand	TGACGCGAATACCCTTTTTGCGCCAGG AGTCGATGACATGCAGCAGGGGCAGCA TGCATTTGCCATCTGAGCGGCTCGGC
rTBC0788	PP3808 deletion oligo, lagging strand	AGCAGGTTCAGCACGGTCACTGGGCGG CAGCTTCGGCCATCATTGCCTGTCTCC TTAAATGATGATGCCGCTGCGCGCCT
rTBC0846	PLU9380 RT-PCR forward primer	ATCGCGCCACCTTCTTATGC
rTBC0847	PLU9380 RT-PCR reverse primer	ATGATAGTGCGGACGCCAAC
rTBC0848	PLU9375 RT-PCR forward primer	GTGCGTCGTGTTTTCAACGA
rTBC0849	PLU9375 RT-PCR reverse primer	TTGCCGTTTGCATGGAGTTG
rTBC0850	PLU9370 RT-PCR forward primer	TTTCTTTCAGGCTCCACCGC
rTBC0851	PLU9370 RT-PCR reverse primer	AAGCCAAGAACACCGCTGAC
rTBC0852	PLU9365 RT-PCR forward primer	GGGTAGCGCGTGGTTATCTG
rTBC0853	PLU9365 RT-PCR reverse primer	ATCGTCAGTACGGCCAAGGA
rTBC0854	PLU9360 RT-PCR forward primer	GGGCCTTACTCCAATGGTAGG
rTBC0855	PLU9360 RT-PCR reverse primer	CCCAAATAACGCAAGCCTGG
rTBC0856	*glbB* RT-PCR forward primer	ATGCGGAAGGCATTCTCACG
rTBC0857	*glbB* RT-PCR reverse primer	TACCAGATGAGCCGGTGGAC
rTBC0858	*glbC* RT-PCR forward primer	GAGATGGCCTGGGAAGTCCT
rTBC0859	*glbC* RT-PCR reverse primer	GGCATCACGTTTTGCAGCAG
rTBC0860	*glbD* RT-PCR forward primer	ACACACCGGCCTTGTTCTTC
rTBC0861	*glbD* RT-PCR reverse primer	CTTGTTGCGGTGGAAGGTGT
rTBC0862	*glbE* RT-PCR forward primer	GTTTCTGGTCCTGGTCAACGC
rTBC0863	*glbE* RT-PCR reverse primer	CTCGATGTAGTCCAGGCAGG
rTBC0864	*glbF* RT-PCR forward primer	GAAGAGCGCCTCACCTATGC
rTBC0865	*glbF* RT-PCR reverse primer	GTCTTCAGGATCGCGAGCAG
rTBC0866	*glbG* RT-PCR forward primer	CGATGCAGGGCAAGAAGACC
rTBC0867	*glbG* RT-PCR reverse primer	CTCGTTCTCCCAGAAGACGC
rTBC0868	*glbH* RT-PCR forward primer	AACTGCATCGGTTGCCTCAG
rTBC0869	*glbH* RT-PCR reverse primer	ACCTCCTGCAGCGTTTTCTG
rTBC0971	*lum* ATG start codon recombineering oligo	g*t*tgagccggcgttggcaataagaatggaacattcgataaCgtgtttgaataattactcaTga aatttttcctaagataaaaagatttttgcagcttagtgattgccgtaaaagctat*c*a
rTBC0972	*glb* ATG start codon recombineering oligo	c*g*ccggtggtgtcgtggcgatcggggctggcggaggaCgagatgtcagaacgttggctcaTgg tggtcgattccggttctcgatcgatgaatgggtcaggtgggttgcgcgctggacg*c*a

4.3.4 Induction of chromosomal fluorescent reporters

Strains *P. putida* GFP01-04, all containing chromosomally integrated GFP constructs were grown overnight in LB media. The following day, colonies were inoculated at an OD600 of 0.1 in 100 μL fresh LB media in a 96-well plate with black walls and flat clear bottom. Induction curves were prepared with ranges of 0.001 - 0.5% (w/v) L-rhamnose and 0.001 - 1mM IPTG. Samples were incubated in a Tecan Infinite M1000 plate shaker at a temperature of 30°C and linear shaking. OD600 and fluorescence measurements were taken every hour, using excitation/emission values of 485/510 nm for samples expressing sfGFP, and 395/508 nm for GFPuv. Corrected fluorescence/OD values are reported at 24 hours.

4.3.5 Prodigiosin production and quantification

Prodigiosin production cultures were prepared by inoculating 25mL of fresh media to an OD650 = 0.05 using overnight cultures. Pre-cultures were prepared in LB media and were washed with PBS before inoculation. Production cultures were incubated at 30°C and while shaking at 250 rpm until they reached an OD650 = 0.5. Prodigiosin production was then induced by adding either 0.2% (w/v) L-rhamnose or 1 mM IPTG. For samples requiring a dodecane overlay, 10 mL of dodecane was added at this time. The cultures were incubated at 30°C until up to 48 hours after inoculation.

At the end of cultivation, cultures were transferred to 50-mL conical tubes of a known weight. Samples were then centrifuged for 15 minutes at 3,000g. If a dodecane overlay was used, then the upper dodecane layer was aspirated and transferred to a 15-mL conical tube of a known weight. The aqueous supernatant was decanted into a fresh 50-mL conical tube of known weight. The remaining pellet was resuspended in 10 mL acidified ethanol (4% (v/v) 1M HCl in ethanol). The volume and weight of aqueous supernatant and dodecane was determined for each sample. The pellets resuspended in acidified ethanol were then centrifuged for 15 minutes at 3,000g. The ethanol supernatant, the aqueous supernatant, and the dodecane were then diluted in acidified ethanol until the absorbance

at 535nm (A535) for each sample was within the linear range of prodigiosin quantification. The molar exctinction coefficient of prodigiosin (ϵ_{535} = 139,800 M^{-1} cm^{-1}, as determined previously [165]) was used to quantify the concentration of prodigiosin in the dodecane overlay, the supernatant, and the pellet for each culture. The total prodigiosin production was calculated from the recorded volumes and weights of the samples. The pellets were washed with 20 mL sterile water and lyophilized overnight to determine the dry cell weight for each culture.

Smaller prodigiosin cultures without a dodecane overlay were analyzed for production by adding two volumes acidified ethanol directly to the culture. The extraction was mixed well by pipetting and 1 mL was collected. The samples were vortexed for 10 minutes at 1500 rpm and then centrifuged at 17,000g for 5 minutes. The supernatant was diluted in acidified ethanol and prodigiosin was quantified by absorbance at 535 nm.

4.3.6 Tn5 knockout library of prodigiosin producers

Tn5 transposition was used to create libraries of random knockouts in the genomes of *P. putida* strains engineered to produce prodigiosin. The mini-transposon vector, pBAM1, was introduced into these strains via bi-parental conjugation. The donor strains were *E. coli* S17-1 λpir containing pBAM1 or *E. coli* BW29427 containing pBAM1 [230]. Transposon libraries were created in *P. putida* PIG01 and *P. putida* PIG02. The selection media for conjugation was LB Kan50 supplemented with 0.002-0.02% (w/v) L-rhamnose or RK Kan50 agar with 25 g/L glycerol. In conjugations where *E. coli* S17-1 λpir was the donor strain, the selection media was supplemented with Irg25. The screens were designed so that the majority of transformants would appear light pink on the selection media, and colonies of interest were screened visually for an increase or decrease in pigmentation. Candidate colonies were re-streaked on screening media to confirm their altered phenotype. The transposon location for each mutant was determined by arbitrary priming PCR, as described previously [230, 384].

4.3.7 Glidobactin A production and extraction

Strains that were engineered to produce glidobactin A were cultivated in RK medium with glycerol or glucose. Production cultures were prepared by inoculating 25 mL of fresh media to OD600 = 0.05. The cultures were incubated at 30°C while shaking at 250 rpm until they reached an OD600 = 0.5. Glidobactin A production was then induced with 0.2% (w/v) L-rhamnose or 1mM IPTG. The cultures continued to grow for up to 48 hours after inoculation.

Glidobactin A was extracted by adding 1 volume butanol to whole production cultures. After briefly shaking the culture/butanol mixtures by hand, solids were removed from the sides of the flask/tube using a spatula. The samples were shaken for 1 hour at 20°C, creating a butanol-water emulsion. One milliliter of the emulsion was collected from each sample immediately after swirling by hand and then transferred to centrifuge tubes. Samples were left on ice for at least 10 minutes and centrifuged for 10 minutes at 10,000g and 4°C. The butanol phase was transferred to a fresh tube and diluted 1:2 in butanol. This extract was filtered with a 0.22 μm nylon membrane in preparation for HPLC and LC-MS analysis. A second extraction protocol was initially used until we observed that glidobactin A was depositing on the inside of the glassware used for production cultures.

In the alternative extraction protocol, glidobactin A was extracted from production cultures as described previously, with a few modifications [309]. 1mL of each culture was collected in 2-mL microcentrifuge tubes and frozen at -80°C. The samples were then lyophilized overnight to remove water. The cells were then resuspended in 1 mL acetone and each suspension was sonicated for 1 minute with a pulsing pattern of 1 second on and 1 second off. The samples were vortexed for 20 min at 1500 rpm and then centrifuged for 5 minutes at 10,000g and 4°C. The supernatants were transferred to 10-mL glass centrifuge tubes. The acetone was removed using a SpeedVac for 60 minutes. The remaining pellet was resuspended in 1 mL ethyl acetate. The suspensions were sonicated, vortexed, and centrifuged in the same manner as before. The supernatants were transferred to the dried acetone extractions. The ethyl acetate was removed using the SpeedVac for 60 minutes. The remaining solids were resuspended in 500 μL methanol. The samples were filtered

with 0.22μm nylon membranes in preparation for HPLC and LC-MS analysis.

4.3.8 HPLC and LC-MS analysis of glidobactin A

HPLC analysis was completed with a Shimadzu HPLC system (Shimadzu Co., Columbia, MD, USA) equipped with a quaternary pump, autosampler, vacuum degasser, and fluorescence detector. HPLC separations were performed with an Agilent Eclipse Plus C18 column (2.1x50mm, with guard column). The injection volume was 2 μL and the flow rate was 0.4 mL/min. Samples were eluted with a gradient elution; mobile phase A was 5% acetonitrile in H2O with 0.1% formic acid and mobile phase B was 100% acetonitrile with 0.1% formic acid. For each injection, the column was equilibrated with 95% mobile phase A and 5% mobile phase B for 1 minute, then switched over to 25% mobile phase B over 2 minutes. Mobile phase B was increased to 75% over 12 minutes, and then increased to 100% over 2 minutes. The solvent was held to 100% mobile phase B over the next 3 minutes before re-equilibrating the column in 95% mobile phase A.

Samples were analyzed by LC-MS with a Vanquish UPLC coupled via electrospray ionization operating in positive mode to a Q-Exactive orbitrap high-resolution mass spectrometer (ThermoScientific). Separation was conducted on a 2.1 x 100mm Acquity UHPLC BEH C18 column with 1.7-μm particle size with a flow rate of 0.2 ml/min. Solvent A was 95:5 H2O-acetonitrile with 0.1% formic acid. Solvent B was acetonitrile with 0.1% formic acid. The following gradient was used: 0 to 2 min, 0% B; 2 to 14 min, linear gradient from 0 to 100% B; 14 to 16 min, 100% B; 16 to 17 min, linear gradient from 100 to 0% B; 17 to 22 min, 0% B. The mass spectrometry parameters were: full MS-SIM (single ion monitoring) scanning between 450 and 600m/z at a resolution of 140000 full width at half maximum (FWHM), automatic control gain (ACG) target of 1e6, and maximum injection time (IT) of 40ms. The MAVEN software suite was used for data analysis [385, 386].

4.3.9 Larger-scale extraction of glidobactin A and silica gel chromatography

While performing experiments for this chapter, we were unable to purchase a standard of glidobactin A for absolute quantitation. In order to prepare a purified glidobactin A standard, a crude extract from a 1-L culture was prepared and then purified through silica gel chromatography (Figure B.2) [376]. Pellet from a 1-L culture of *P. putida* LUM02 was extracted with 400 mL acetone. The acetone slurry was centrifuged and the acetone was decanted into an evaporating flask. The acetone was then removed using a rotary evaporator. The remaining pellet was extracted with 400 mL ethyl acetate. This slurry was centrifuged and decanted into the same evaporating flask containing the solid from the acetone extract. The ethyl acetate was removed using a rotary evaporator. The remaining solid from both extracts was dissolved in 150 mL methanol and then transferred to a new flask. The methanol was removed using the rotary evaporator and the remaining solid was kept at -20°C. A 20-mL silica gel column was prepared for purification of glidobactin A. The extract was applied to the column and eluted using a gradient elution from 100% ethyl acetate to 50% methanol and ethyl acetate. The fractions were monitored using thin-layer chromatography (solvent: 30% methanol/70% ethyl acetate).

The fractions containing glidobactin A were then pooled and analyzed by 1H-NMR (Figure B.5). A clear signal associated with the methyl group of the threonine residue was used to determine the quantity of glidobactin A in the standard. Other researchers have previously identified glidobactin derivatives from production cultures [387, 84], and through further analysis of the standard with LC-MS, we identified putative glidobactin derivatives that share the threonine residue in their structure, and therefore would contribute to the signal used to quantify glidobactin A (Table 4.4). Based on peak areas of the ions associated with these derivatives, about 15% of the signal used for quantification in NMR could be attributed to other glidobactin derivatives, and we incorporated this finding when determining the absolute titers of glidobactin A in our cultures.

Table 4.4: Secondary peaks predicted to contribute to 1H-NMR signal used to quantify glidobactin A. Predicted compounds were identified by comparing experimental m/z values to literature values. Further experiments were not performed to verify these identities. These identities were used to create a conservative estimate that 15% of the NMR signal used to quantify glidobactin A could be attributed to other glidobactin derivatives that have the same methyl group. Sum of peaks marked by * is equal to 10.97%.

Peak	*Predicted compound*	***m/z***	**Retention time**	**Intensity**	**Relative to peak 1**
1	Glidobactin A	521.333	6.93	1.91E+07	100.00%
2	Glidobactin A + 1	522.335	6.75	5.99E+06	31.36%
3	Glidobactin A + Na^+	543.314	6.79	1.20E+06	6.28%
4*	Luminmycin E	523.342	6.73	8.30E+05	4.35%
5	Glidobactin A + Na^+ + 1	544.321	6.84	3.38E+05	1.77%
6*	Glidobactin C	549.367	6.96	3.20E+05	1.68%
7*	?	540.307	6.72	2.80E+05	1.47%
8*	Glidobactin G	537.328	6.71	2.36E+05	1.24%
9*	?	495.318	6.31	2.20E+05	1.15%
10*	?	548.293	6.74	2.06E+05	1.08%

4.3.10 ^{1}H-NMR analysis of glidobactin A standard

^{1}H-NMR spectra were recorded at 24°C on a BrukerBiospin Avance III HD-600 MHz (1H 600.13 MHz, 13C 150.76 MHz) instrument equipped with TopSpin 3.2 software and 5 mm TCI-F (1H, 19F/13C/15N) cryoprobe fitted with a z-gradient. T1 relaxation for the compound was determined by inversion recovery to be 2.1s. Spectra were recorded using 3 mg of glidobactin in 500 μL DMSO-d_6. Quantitative 1H spectra were run with d1=15s and aq=2.7s using 1,4-NJbis(trichloromethyl) benzene (Acros Organics) as internal standard. Spectra were analyzed using MNova 14.2 (Mestrelab Research, Santiago de Compostela, Spain). Purity of the standard glidobactin sample was determined to be 79%.

4.3.11 Reverse transcription PCR

Oligos used for RT-PCR analysis are shown in Table 4.3. Production cultures of *P. putida* LUM01 and *P. putida* GLB01 were analyzed by RT-PCR. 400 μL of production culture were collected and RNA was extracted using the Qiagen RNeasy Mini Kit (Cat. 74104). Purified RNA was digested with DNase 1 following the Qiagen RNase-free DNase kit (Cat. 79254). The GoScript^{TM} Reverse Transcriptase kit from Promega was used to synthesize cDNA from the RNA samples (Cat. A5000). The cDNA was used as templates

for PCR using GoTaq Green Master Mix from Promega (Cat. M712). Primers used for GoTaq PCR of cDNA are described in Table 4.3. The thermocycler settings were as follows: 2:30 at 95°C, 30 cycles of 0:30 at 95°C, 0:30 at 55°C, and 1:00 at 72°C, followed by 5:00 at 72°C.

4.4 Results

4.4.1 Introducing heterologous BGCs into *P. putida.*

We initially focused on engineering a strain of *P. putida* to produce prodigiosin because the visible pigmentation of production cultures would allow for rapid detection of heterologous production from early production strains. The BGC for prodigiosin biosynthesis, designated *pig*, in *S. marcescens* ATCC274 is 20,948 bp in length and contains 14 genes expressed in a single operon (Figure 4.1a) [302]. Prodigiosin biosynthesis occurs as a bifurcated pathway where two precursors, 2-methyl-3-n-amyl-pyrrole (MAP) and 4-methoxy-2,2'-bipyrrole-5-carbaldehyde (MBC), are synthesized independently and then condensed in the final reaction that yields prodigiosin (Figure 4.1b). MAP is derived from 2-octenoyl-ACP/-CoA, and MBC is derived from L-proline, malonyl-CoA, and L-serine. L-proline is incorporated into the pathway by an NRPS (PigG and PigI), followed by the incorporation of malonyl-CoA by a PKS (PigJ and PigH) [378].

Replicative vectors in *P. putida* can sometimes induce a growth defect, which can have a negative effect on heterologous production [225, 388]. Therefore, we sought to optimize a genome editing protocol for integrating large DNA constructs into the chromosome. We previously developed a set of vectors for constructing scarless deletions in *P. putida* KT2440 using Cas9-assisted homologous recombination [225]. A replicative vector using the RK2 origin, pCas9, expressed Cas9 from Streptococcus pyogenes and the λRed system from λ bacteriophage, Exo, Beta, and Gam. The single guide RNA (sgRNA) was expressed from pgRNA, a replicative vector containing the BBR1 origin. The repair template was located in an integrative vector containing homology arms for

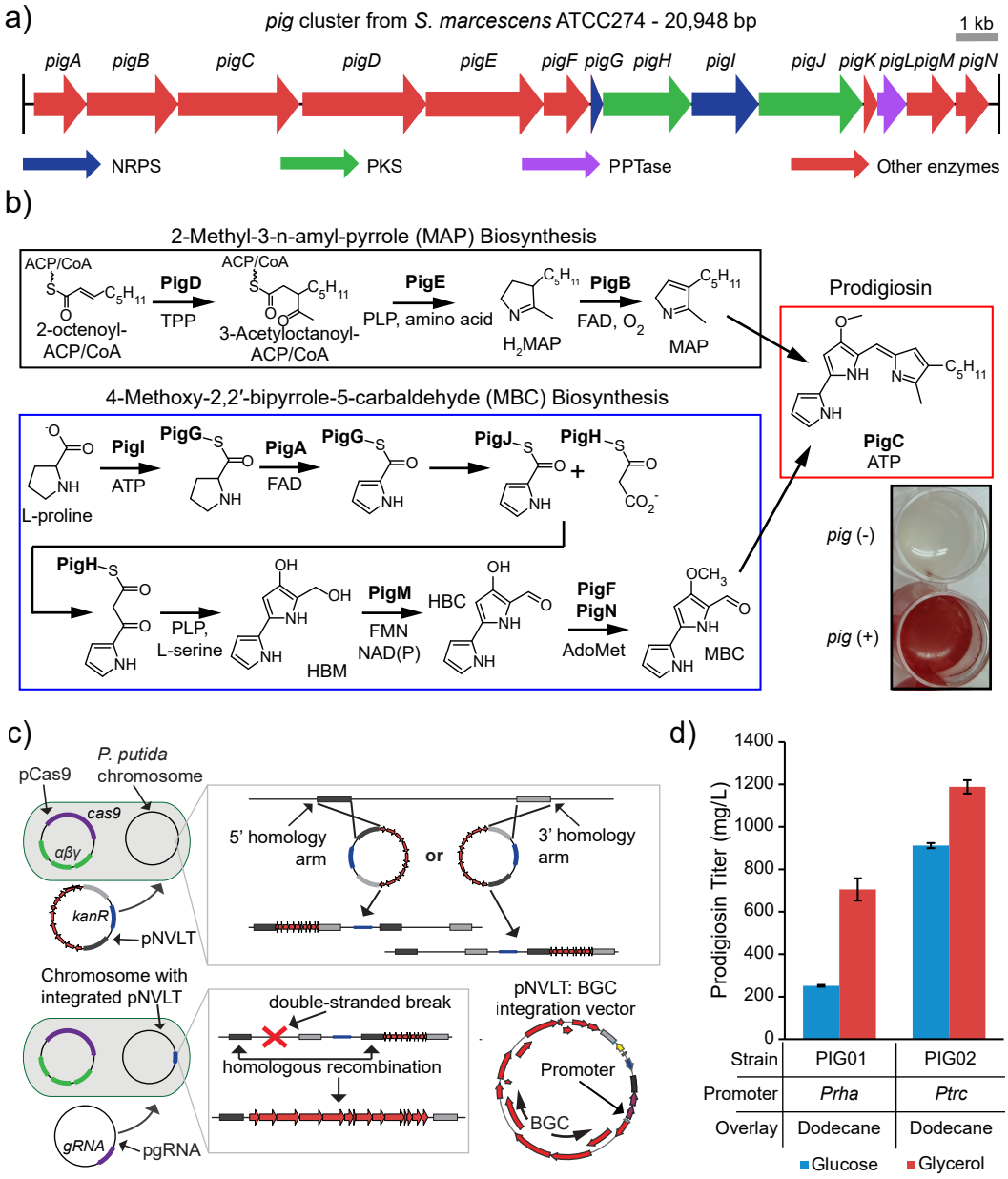


Figure 4.1: **Establishing a prodigiosin production strain.** **a)** Map of prodigiosin BGC from S. marcescens ATCC274. Colored arrows indicate the enzyme class according to the legend. **b)** Prodigiosin biosynthesis begins with intermediates in fatty acid metabolism and L-proline. 2-octenoyl-CoA/ACP is converted to MAP by PigDEB. L-proline, serine, and malonyl-CoA are converted to MBC by PigIA-JHMFN. The two intermediates are then fused by PigC to produce prodigiosin which generates a visible red color (inset) in culture. **c** Integration of BGCs into the chromosome of *P. putida*. A pNVLTv2 integration vector is introduced via conjugation into a strain of *P. putida* containing pCas9. The integration vector carries out a single-crossover recombination with the *P. putida* chromosome. The Cas9 counterselection is then enabled after electroporating pgRNA targeting the wild-type sequence to be replaced by the BGC. This selects for the double-crossover recombination of the integration vector and results in a markerless and stable integration of the BGC. **d)** Prodigiosin production from cultures grown in glucose or glycerol-based media under control of *rha* or *trc* promoters. All cultures were supplied with dodecane as a product-sink. Error bars represent standard deviation, n = 3 biological replicates. Differences in values between all samples were found to be statistically significant ($P<0.01$) by the Student's *t*-test.

the region to be deleted [158]. The biggest difference between our needs and the existing genome editing method was the size of the required repair template, here containing a full BGC. Cloning BGCs can lead to large plasmid sizes and the introduction of genes that may be toxic to common *E. coli* cloning strains [84]. Large plasmids are not as easily transformed by electroporation into *P. putida*, but conjugation is significantly more efficient [213]. To address these issues, we edited the integrative vector to replace the high-copy ColE1 origin with the medium-copy p15A origin and add a functional origin of transfer (oriT) sequence for performing conjugations. Unnecessary genes in the original vector were also removed to reduce the plasmid size, resulting in the integrative vector pNVLTv2.

Next, we deleted the endogenous BGC for pyoverdine biosynthesis to simplify the baseline strain, create a "neutral" integration site, and eliminate potential analytical challenges [231, 389, 73]. We had already created a strain from previous work with one gene deleted, (*P. putida* Δ*pvdL*), so we introduced homology arms for deleting the remaining three genes, *pvdIJD*, into pNVLTv2, resulting in pNVLTv2-pvdIJD. We designed a sgRNA plasmid targeting pvdI, pgRNAtet-pvdI, and successfully constructed the 24,920-bp deletion using the two-step Cas9-assisted protocol that our group has described previously, resulting in strain *P. putida* pvd-.

We selected two inducible promoter systems, the *rha* promoter (P_{rha}, induced by L-rhamnose) and the *trc* promoter (P_{trc}, induced by IPTG), to drive expression of heterologous BGCs at the *pvdIJD* locus. BGCs expressed as a single operon can be refactored for heterologous expression by replacing the native sequence directly upstream of the initial gene with a synthetic promoter and ribosome binding site (RBS) optimized for the heterologous host. Inducible promoters are ideal for these types of projects because they reduce BGC expression during cloning protocols and enable two-phase production cultures, where cultivation is separated into a growth stage and a production stage [390]. P_{rha} is tightly repressed by carbon catabolite repression in *E. coli*, but not in *P. putida*, making it an appropriate promoter for cloning BGCs in *E. coli* before transferring them into *P. putida* [184]. P_{trc} is one the strongest promoters that we have characterized in *P.*

putida, which allowed us to maximize expression of heterologous BGCs [225].

Experiments with each promoter linked to a fluorescent reporter construct confirmed that P_{trc} has higher basal and maximum expression than P_{rha} (Figure 4.2a). Next, we attempted to construct *pig* integration vectors with both promoters using RecET direct cloning [84, 80] in *E. coli* (Figure 4.1c). We isolated correct clones with P_{rha} only when transformation media was supplemented with glucose, conditions that silence P_{rha} through catabolite repression. Consistently, we were unable to obtain a P_{trc} version of the *pig* vector, presumably because leaky expression of the *pig* genes was toxic to *E. coli*. The expression construct including P_{rha} and the *pig* BGC (23,247 bp total) was integrated into the *P. putida* chromosome at the *pvdIJD* locus through the two-step protocol as described above, resulting in strain PIG01 (Figure 4.1c). Upon addition of up to 0.5% (w/v) L-rhamnose to the growth media, cultures of PIG01 produced a visible red pigment characteristic of prodigiosin production (Figure 4.1b).

4.4.2 Optimizing prodigiosin production

Once we established prodigiosin production in *P. putida*, we then compared various media formulations and their effects on prodigiosin titers. Terrific Broth (TB) has been used previously to produce prodigiosin in *P. putida* cultures [165]. Alternatively, Riesenberg-Korz (RK) medium is a minimal medium developed for high-cell density cultures of *E. coli*, and variations of this medium have been used for the production of other products from *P. putida* [391, 381]. After determining the optimum L-rhamnose concentration for induction (0.2% w/v), we compared prodigiosin production in rich media and minimal media (Figure 4.2b). In non-baffled shake flask cultures of *P. putida* PIG01, prodigiosin titers were less than 100 mg/L in TB media (Figure 4.3a). Prodigiosin is insoluble in water, so we added an organic overlay to provide a product sink, which greatly improved prodigiosin production in TB media to over 500 mg/L. Prodigiosin production in RK media with 2.5% glycerol reached approximately 500 mg/L without an organic overlay and over 600 mg/L with an overlay (Fig. 4.3a). Interestingly, using glucose as the carbon source in minimal media led to much lower prodigiosin titers of approximately 200 mg/L

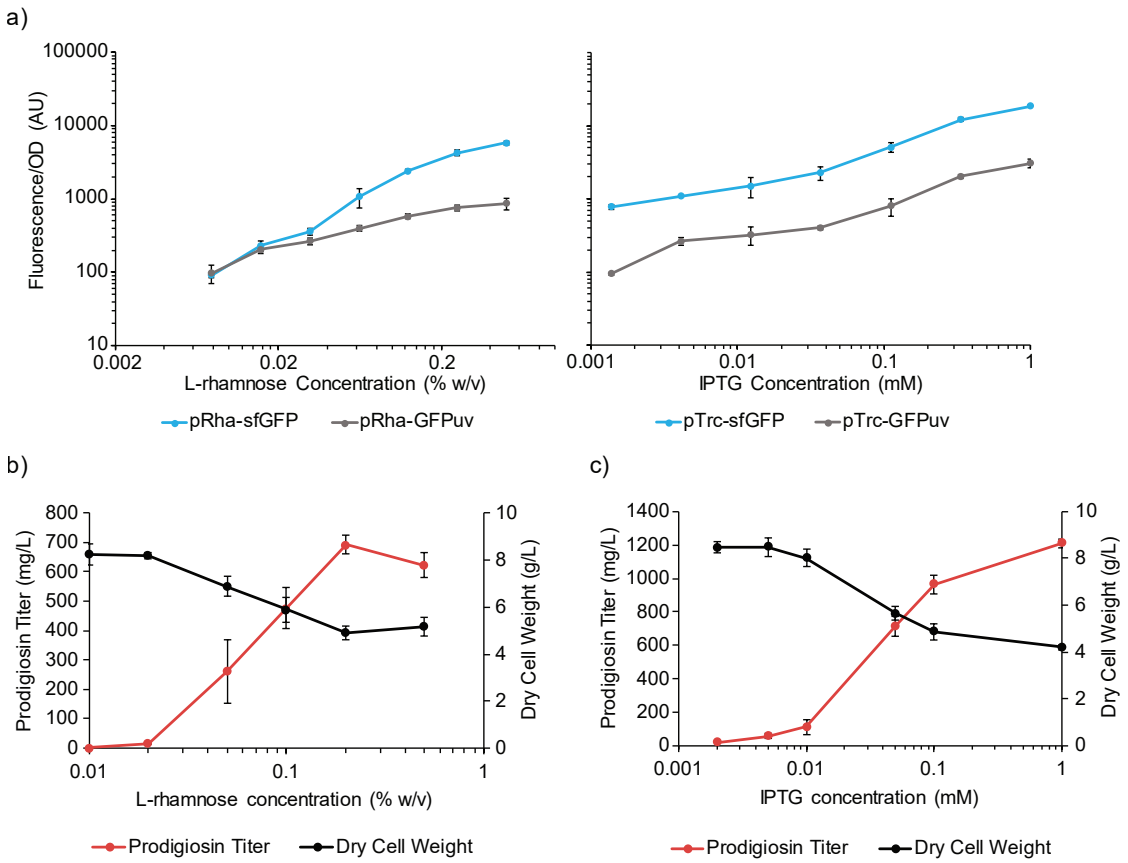


Figure 4.2: **Inducible promoters and prodigiosin production in minimal media. a)** Induction curve for sfGFP and GFPuv fluorescence under control of *rha* and *trc* promoters. Reporter constructs containing combinations of the two inducible promoters and a gene encoding sfGFP or GFPuv were introduced to pNVLTv2-pvdIJD. They were integrated into the Δ*pvdIJD* locus of *P. putida* pvd-, resulting in strains GFP01-04. **b)** Induction curve for prodigiosin production in PIG01 in 25 mL RK glycerol with 10-mL dodecane overlay. **c)** Induction curve of PIG02 in the same conditions.

(Figures 4.1d, 4.3b).

After establishing optimal media and cultivation conditions for strain PIG01, we sought to further improve prodigiosin production through transcriptional control. In lieu of cloning a pig integration construct with P_{trc}, we designed a chromosomal integration vector for replacing P_{rha} with P_{trc} in PIG01, pNVLTv2-pvdIJD-Ptrc-pigA. Integration of this construct resulted in strain PIG02, which produced a strong red pigment in cultures supplemented with IPTG, and 1 mM was the optimal inducer concentration (Figure 4.2c). Introducing a stronger promoter for pig expression improved prodigiosin production, as *P. putida* PIG02 generated higher prodigiosin titers than PIG01 (Figure 4.1d). This strain also produced more prodigiosin on minimal media with glycerol than with glucose. PIG02 cultures grown in RK media with glycerol generated prodigiosin titers around 1.1

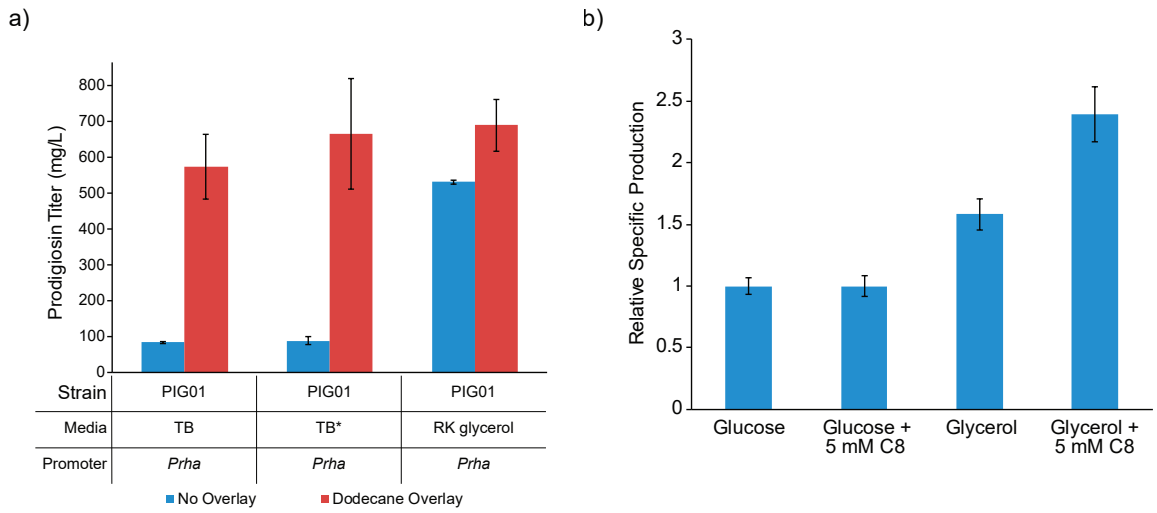


Figure 4.3: **Prodigiosin production in various media formulations. a)** Comparing prodigiosin production on rich vs. minimal media. Cultures were grown in 25 mL media in 250-mL non-baffled flasks, with and without a dodecane overlay. TB = terrific broth with 0.5% glycerol, TB* = TB media with 2.5% glycerol. **b)** Relative prodigiosin titers from PIG01 in RK glucose and RK glycerol with and without octanoic acid added to media. Cultures were grown in 2 mL media in 12-well plates and prodigiosin production was evaluated at 24 hr.

g/L, whereas production on glucose was approximately 900 mg/L.

4.4.3 Screening for mutants with improved prodigiosin biosynthesis

To identify strategies for improving prodigiosin production, we devised a screen based on the appearance of *P. putida* colonies producing prodigiosin. Tn5 mutagenesis was used to create mutant libraries of *P. putida* PIG01 and PIG02, which normally have a light-pink phenotype on solid media with a low concentration of inducer (Figure 4.4a). Libraries of these strains were plated on LB agar or minimal agar with glycerol, and individual colonies with a strong red color, indicating increased prodigiosin production, were isolated and sequenced to determine the location of the Tn5 insertion. Some isolated mutants included those involved in lipid A biosynthesis (*yijP*), an ABC transporter (*ttgB*), and an aconitase (*acnA-I*) (Table 4.5). Similar targets were also identified in another study screening for mutants with improved fitness during indigoidine production [392]. Surprisingly, the most common type of mutants were disruptions in components of the electron transport chain.

The two operons containing Tn5 insertions, *cyo* and *ccm*, encode production of the bo_3 type oxidase (Cyo) and cytochrome c maturation, respectively (Figure 4.4b). Cyo is the primary terminal oxidase used during exponential growth and has been shown to be involved in gene regulation in *P. putida* [395]. Mutants with disruptions in the *ccm* operon had a drastic growth defect on minimal media and did not grow on rich media. Disruptions in the *cyo* operon led to only a slight growth defect, so we investigated this operon's effect on prodigiosin production (Figure 4.5). The cyo operon contains five genes, *cyoABCDE*, and we deleted the region containing the RBS and all five genes in this operon in both the L-rhamnose- and IPTG-inducible prodigiosin producing strains (resulting in PIG07 and PIG08, respectively). Transcriptomic data available in the literature suggested that the *cyo* operon is upregulated during growth on glucose compared to glycerol [394], so we compared prodigiosin production in RK media with glucose in addition to glycerol. No significant changes in production were observed from strains grown on glycerol, but PIG08 had improved production on glucose, making its performance equivalent to PIG02 on glycerol (Figure 4.4c).

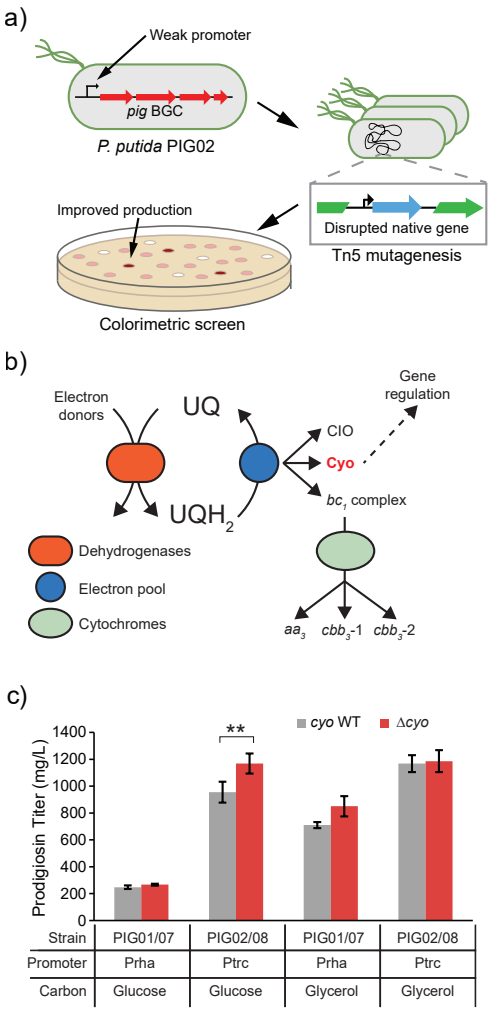


Figure 4.4: **Tn5 libraries highlight disruptions in electron transport chain. a)** Generalized workflow for generating and screening mutant libraries for improved prodigiosin production. **b)** Components of electron transport chain, highlighting role of the bo3 oxidase, Cyo. UQ – ubiquinones, UQH2 – ubiquinols, CIO – cyanide insensitive oxidase. Adapted from Ugidos *et al.* (2008) and Nikel *et al.* (2014) [393, 394]. **c)** Effect of deleting *cyo* operon on prodigiosin production on glucose and glycerol. Error bars represent standard deviation, n = 3 biological replicates. Differences between samples marked with asterisks were found to be statistically significant by the Student's *t*-test (** = P<0.01).

Table 4.5: Gene disruptions sequenced in "red" mutants from Tn5 libraries in prodigiosin producing strains. Mutants in Library A were found in *P. putida* PIG01 on LB agar. Mutants in Library B were found in PIG02 on RK agar with glycerol.

Gene	**Annotation**
Library A	
acnA-I	Aconitase
phoP	Two-component system DNA-binding regulator
oprH	Outer membrane protein H1; regulated by PhoP
sotB	Sugar efflux transporter
yijP	Phosphoethanolamine transferase; lipid A biosynthesis
cyoC	Cytochrome bo termainl oxidase subunit III
ptsP	Phosphoenolpyruvate-dependent regulator
Library B	
ppx	Exopolyphosphatase
cyoE	Protoheme IX farnesyltransferase
ccmA	Cytochrome c biogenesis ABC transporter ATP-binding protein
*ccmD**	Heme exporter protein D
*ccmF**	Holocytochrome c synthetase
tolA	Colicin S4/ filamentous phage transport protein
PP_5608	Valyl-tRNA synthetase
PP_5348	LysR family regulator
ttgB	Efflux pump membrane protein TtgB
rnr	Exoribonuclease R

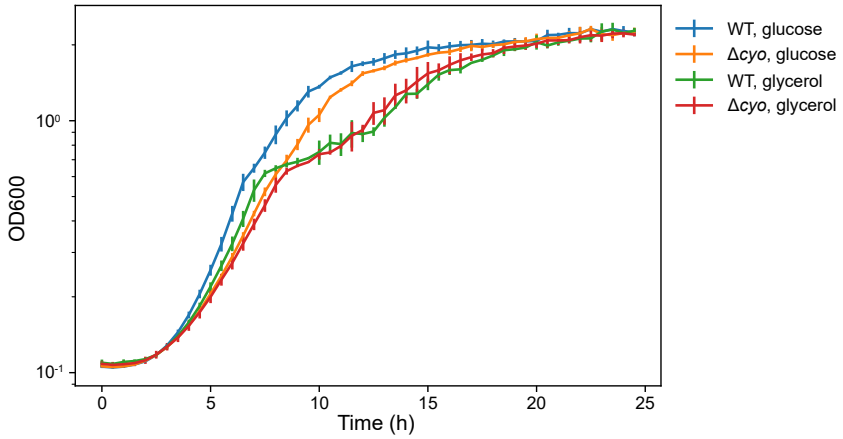


Figure 4.5: **Growth of *P. putida* with and without deletion of cyoABCDE.** Strains were grown in 100-μL volumes in a 96-well plate in RK media with the stated carbon source. Plate was incubated in a Tecan M1000 plate reader at 30°C with linear shaking with a radius of 3 mm. Growth was monitored by absorbance at 600 nm.

4.4.4 Heterologous expression of two glidobactin A BGCs

The high titers of prodigiosin achieved from heterologous expression in *P. putida* were encouraging, but we also wanted to interrogate variables affecting the expression of PKSs and NRPSs with a canonical modular structure. Glidobactin A provided this opportunity, as its biosynthesis involves a hybrid PKS/NRPS that has three modules in total and is encoded by genes approximately 12.4 kb in length (Figure 4.6a). The two homologous BGCs in *S. brevitalea* (*glb*) and *P. luminescens* (*lum*) both produce glidobactin A as the primary product [396, 379]. Glidobactin A biosynthesis starts with the acylation of an L-threonine residue by GlbF/PLU1878p [397]. The hybrid PKS/NRPS, GlbC/PLU1880p, then incorporates L-lysine, L-alanine, and malonyl-CoA before performing the final cyclization reaction [398]. The most apparent difference between the two BGCs is that the *glb* and *lum* clusters have a GC content of 70% and 47%, respectively. The *glb* cluster also has three genes that are not present in the *lum* cluster: *glbE*, which encodes an MbtH-like protein (MLP); *glbA*, a transcriptional regulator; and *glbH*, which has been shown to be involved in the hydroxylation of a lysine residue that is incorporated into glidobactin A (Table B.1) [84, 379]. Unlike the prodigiosin BGC, neither glidobactin BGC contains a gene encoding a dedicated PPTase. Expressing each BGC afforded us the opportunity to assess several factors in heterologous production of polyketides and nonribosomal peptides: the effect of MLP expression on the activity of a heterologous NRPS, interactions between the host's PPTase and heterologous PKSs and NRPSs, and the significance of similarities in phylogeny and GC content between *P. putida* and heterologous BGCs.

Using the same strategy applied to prodigiosin, both BGCs were integrated into the *pvdIJD* locus of *P. putida* pvd- with P_{rha} directly upstream of the first gene (*glbB*, *plu1881*), resulting in the strains *P. putida* GLB01 and LUM01. Glidobactin A production from these two strains was initially assessed by HPLC; based on peak area, LUM01 produced 15-fold more product than GLB01 (Figure 4.6b). We verified that the primary compound was glidobactin A using LC-MS and ^{1}H-NMR (Figure 4.6c, Figure B.5). Similar to prodigiosin, glidobactin A production was greater on glycerol compared to glucose (Figure B.3a). Before performing further modifications to strains GLB01 and

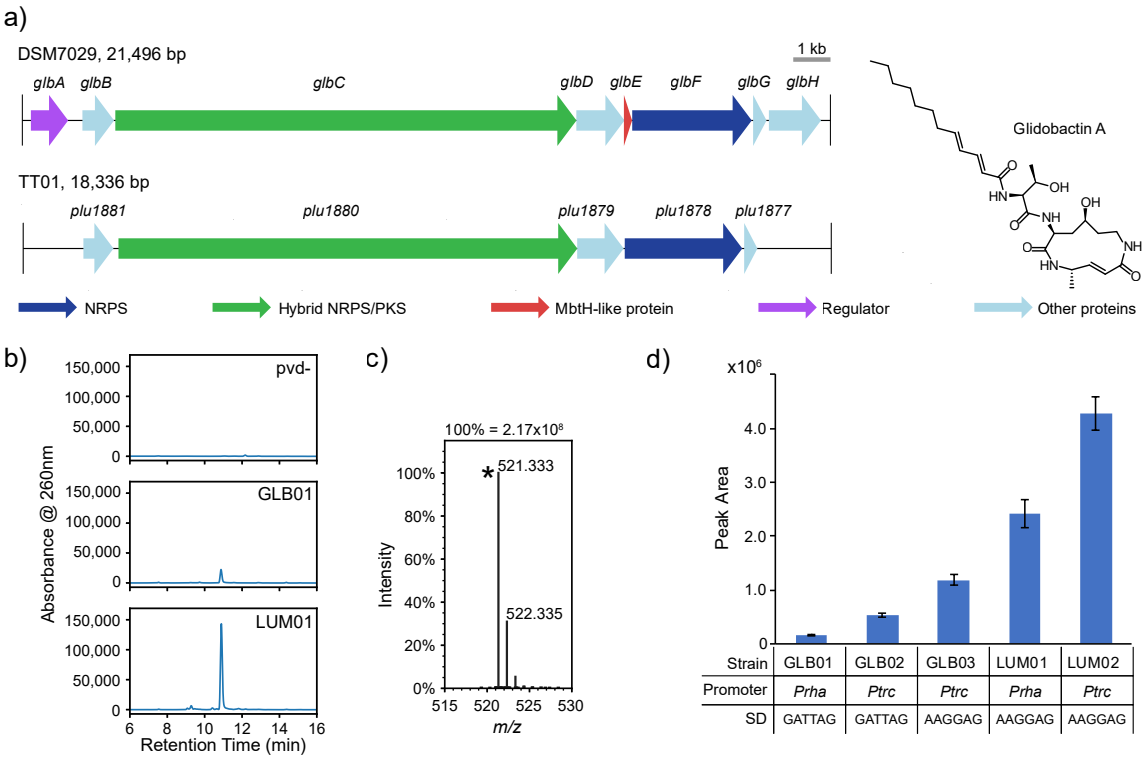


Figure 4.6: **Identifying glidobactin A as the primary product from expression of *glb* and *lum* clusters. a)** Gene maps of glidobactin BGCs from *S. brevitalea* DSM7029 and *P. luminescens* TT01. Colored arrows indicate the enzyme class according to the legend. The structure of glidobactin A is shown to the right. **b)** Unique peak identified in HPLC analysis of extracts from glidobactin production strains. **c)** LC-MS shows that the major peak has m/z value corresponding to glidobactin A + H^+ (calculated m/z = 521.3334). **d)** Effects of promoter and RBS strength on glidobactin production. All strains were grown in 25 mL of glycerol-based media in 250-mL non-baffled shake flasks. SD=Shine-Dalgarno sequence. Error bars represent standard deviation, n = 3 biological replicates. Differences in values between all samples were found to be statistically significant (P<0.01) by the Student's t-test.

LUM01, we confirmed by RT-PCR that both BGCs were fully transcribed in *P. putida* (Figure B.4). It has been shown that deleting the last gene in the *glb* cluster, *glbH*, does not abolish glidobactin A production but significantly reduces it [379]. The RT-PCR results demonstrated that *glbH* was transcribed, so we did not investigate GlbH activity as a limiting step in glidobactin production in strain GLB01.

After determining the relative production titers between GLB01 and LUM01, we sought to improve production by modifying transcription and translation of the BGCs with the primary goal of determining why the *glb* cluster yielded less product than the *lum* cluster in *P. putida*. We continued to mirror our strategy with the *pig* cluster and swapped to the stronger promoter, P_{trc} in both strains, resulting in strains GLB02 and LUM02. As expected, production of glidobactin A increased, with both strains producing

approximately 2-fold more product (Figure 4.6d). Production from GLB02 was still much lower than LUM02, so we investigated translation initiation as a potential factor. Using the RBS calculator [399], we had designed the RBS sequences for the first gene in each BGC to have similar translation initiation rates (TIR) (*glb*: 13,000 au; *lum*: 9,000 au). Coincidentally, the Shine-Dalgarno sequences upstream of the BGCs in strains PIG01 and LUM01 were the same (AAGGAG), whereas GLB01 had a different sequence (GATTAG). When we changed the Shine-Dalgarno sequence for *glb* to AAGGAG in the RBS calculator, the predicted TIR increased to 43,000 au. Therefore, we constructed strain GLB03 by simultaneously inserting P_{trc} and the AAGGAG Shine-Dalgarno sequence upstream of the glb cluster. This modification increased glidobactin A production by 2-fold compared to GLB02, which had P_{trc} and the original RBS sequence, consistent with the increased RBS strength predicted in silico (Figure 4.6d).

4.4.5 Improving glidobactin A production through MLP and PPTase overexpression

Functional expression of large, multi-modular enzymes is often the greatest challenge in engineering natural product synthesis, so we identified the steps governed by PKSs and NRPSs as a possible bottleneck in glidobactin production. The presence of a cognate MLP in the *glb* cluster but not the *lum* cluster raised the question of whether MLP availability is affecting heterologous NRPS activity. NRPSs often require a complementary MLP for optimal solubility and activity, and homologs from different BGCs or strains are usually not interchangeable [253]. Indeed, it has been demonstrated that soluble expression of the NRPS module, GlbF, in *E. coli* requires co-expression of its cognate MLP, GlbE [397]. The genome of *P. putida* KT2440 contains one gene (PP_3808) encoding an MLP specific for the NRPSs responsible for pyoverdine biosynthesis. Deleting the gene encoding *P. putida*'s native MLP in strain GLB03 (resulting in GLB04) did not significantly change glidobactin A production (Figure 4.7c). To increase MLP availability, we inserted a second copy of glbE with a constitutive promoter in place of *P. putida*'s native MLP, resulting in strain GLB05 (Figure 4.7a). Overexpression of *glbE* resulted in greater than

2-fold increase in glidobactin A production compared to GLB03 and GLB04. In contrast, deleting PP_3808 or overexpressing glbE on a plasmid did not have a noticeable effect on glidobactin production for strain LUM01 (Figure B.3b,c).

After demonstrating that expression of *glbE* was non-optimal for glidobactin A production using only the *glb* cluster, we next investigated whether PPTase activity in *P. putida* could be limiting production. We initially modified the strains LUM01 and LUM02 by inserting a strong constitutive promoter upstream of PP_1183, resulting in the strains LUM03 and LUM04, respectively (Figure 4.7a). Both strains saw an increase in production, and specifically, strain LUM04 had about 50% higher glidobactin A production compared to the parent strain, LUM02 (Figure 4.7d). However, these strains had a slower growth rate compared to the parent strains, and when we measured glidobactin production from these strains in 3-mL cultures in tubes instead of 25-mL cultures in flasks, we found that they produced less glidobactin compared to the parent strains (Figure B.3d).

To avoid any deleterious effects on native metabolism from over-producing *P. putida*'s native PPTase, we introduced a heterologous PPTase as an orthogonal alternative. Sfp is a promiscuous PPTase from *Bacillus subtilis* that has enabled the production of functional PKS and NRPS enzymes in *E. coli* [248]. We designed an expression construct for integrating *sfp* into the *pvdL* locus with the *tac* promoter, enabling induction by the addition of IPTG in strains expressing LacI (Figure 4.7a). Introducing this construct to the IPTG-inducible production strains, LUM02 and GLB05, resulted in LUM05 and GLB06. Co-expression of *sfp* improved glidobactin production by 30% in both strains (Figure 4.7c,d). These strains also did not have a noticeable growth defect and did not have a significant drop in production in smaller cultures (Figure B.3d).

4.4.6 Modifying translation of hybrid PKS/NRPS via Cas9-assisted mutagenesis

While analyzing GC content and codon usage of the BGCs described in this work, we noticed that the genes encoding the hybrid PKS/NRPS (*glbC*, *plu1880*) in both glidobactin clusters each used the alternative start codon, GTG, which lowers translation initiation

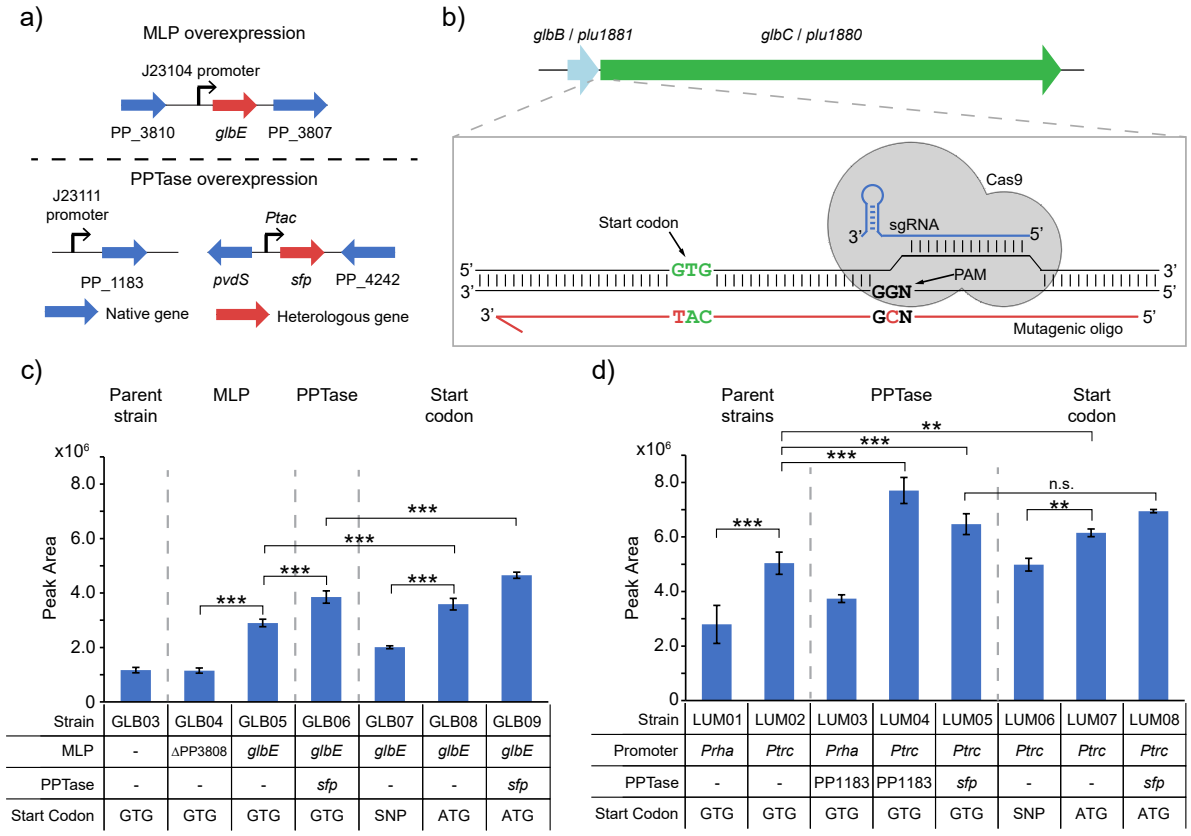


Figure 4.7: **Improving functional expression of PKS/NRPS enzymes in glidobactin pathway** **a)** Strategies for overexpressing MLPs and PPTases in production strains. The MLP gene, *glbE*, was integrated in place of PP_3808, the gene encoding *P. putida*'s native MLP, along with a constitutive promoter from the Anderson promoter library. A constitutive promoter from the Anderson library was integrated upstream of PP_1183, the gene encoding *P. putida*'s native PPTase. The inducible promoter, P_{tac}, and *sfp* were integrated in place of *pvdL*. **b)** Cas9-assisted ssDNA oligo recombination scheme for introducing ATG start codons in *glbC/plu1880*. To select for start codon mutations, an sgRNA targets the PAM sequence immediately downstream of the start codon. A mutagenic oligo contains mutations in the PAM and the start codon, and Cas9 activity selects against cells that don't incorporate mutations from this oligo. PAM=protospacer adjacent motif. **c)** and **d)** Effects of MLP overexpression, PPtase overexpression, and start codon mutagenesis on glidobactin production strains. Charts are labeled and split into sections to emphasize genetic changes made (e.g. "PPTase" samples have modifications to PPTase expression). All cultures were in 25-mL of glycerol-based media in 250-mL non-baffled flasks. Error bars represent standard deviation, n $\geq$ 3 biological replicates. Differences between samples marked with asterisks were found to be statistically significant by the Student's t-test (n.s. = not significant, * = P<0.05, ** = P<0.01, *** = P<0.001).

rates compared to the canonical ATG start codon in prokaryotes, including *P. putida* [400, 401]. This realization provided the opportunity to further engineer glidobactin production through the simple strategy of generating a point mutation in the start codon of both genes. We hypothesized that this change would improve translation initiation for these genes and increase glidobactin A production.

To generate strains with these start codon mutations, we adapted our genome editing vectors for introducing point mutations via oligo recombineering. Wu and colleagues

recently reported the generation of knockouts in *P. putida* using only one λ phage recombinase, Beta [197]. We constructed a derivative of pCas9, pCas9-beta, and were able to construct small chromosomal deletions using 80-nt oligos (Table B.3). However, it has been shown in the literature that *P. putida* can efficiently repair mismatches in its chromosome, and G to A changes are much more sensitive to mismatch repair compared to other mutations [200]. We accounted for this potential issue by incorporating a strategy developed by Aparicio *et al.* to temporally inhibit mismatch repair in *P. putida*. We constructed a second plasmid for oligo recombineering, pCas9-beta-mmr, which co-expresses *beta* and a defective mutant of *mutL* from *P. putida*, $mutL_{\mathrm{E36K}}$. Expression of the *mutL* mutant inhibits mismatch repair in *P. putida*, allowing for the generation of point mutations that would normally be repaired by the host.

We identified sgRNA targets 18-24 bp downstream of the start codon in *glbC* and *plu1880* (Figure 4.7b). The single-stranded DNA oligos (125 nt) designed for each BGC contained two point mutations: the desired G to A start codon mutation and a synonymous point mutation in the PAM site of the sgRNA target. Attempts to introduce these mutations with pCas9-beta resulted in most mutants having only the PAM mutation (Table 4.6). Repeating the recombineering protocols with pCas9-beta-mmr resolved this problem, and most mutants sequenced contained both the start codon and PAM mutations. We initially introduced these mutations into IPTG-inducible strains without *sfp* overexpression (LUM02 and GLB05), resulting in strains LUM06 and GLB07, which have the PAM mutation only (designated as SNP), and strains LUM07 and GLB08, which have both mutations (designated as ATG). The ATG derivatives of both strains produced about 20% more glidobactin A (Figure 4.7c). Notably, GLB07 produced 20% less glidobactin A than GLB05, suggesting that the synonymous PAM mutation in *glbC* had a negative effect on expression, even though the mutated codon, TCG, is more common in *P. putida* than the native one, TCC. Next, we introduced the ATG mutation to strains containing *sfp* overexpression, resulting in strains LUM08 and GLB09. LUM08 did not have a significant increase in glidobactin A production compared to its parent strain, LUM05, but GLB09 produced 20% more glidobactin A than GLB06 (Figure 4.7c,d).

Table 4.6: Fraction of colonies with mutations in introduction of ATG start codon in *glbC* and *plu1880*. Mutants denoted "*beta* only" were created with pCas9-beta and mutants denoted "*beta* + $mutL_{E36K}{}^{PP}$" were created with pCas9-beta-mmr. "SNP only" denotes sequences with only PAM mutation. "ATG + SNP" denotes sequences with both PAM mutation and ATG mutation. "Other" refers to sequences that do not contain the PAM mutation and/or have other mutations.

BGC	***beta* only**			***beta* + $mutL_{E36K}{}^{PP}$**		
	SNP only	**ATG + SNP**	**Other**	**SNP only**	**ATG + SNP**	**Other**
lum	19/23	4/23	0/23	3/24	17/24	4/24
glb	21/24	2/24	1/24	2/24	20/24	2/24

The best glidobactin A production strain resulting from modifications mentioned above was LUM08, which had an approximately 3-fold increase in production compared to the initial production strain constructed with the *lum* cluster. GLB09 produced 30 times more glidobactin A than GLB01. After purifying a glidobactin A standard that we verified by ^{1}H-NMR, we used LC-MS to determine the titer from LUM08 and GLB09, which we found to produce approximately 390 mg/L and 270 mg/L glidobactin A, respectively.

4.4.7 Engineering metabolism and gene regulation for improved natural product titers

After establishing robust production of both prodigiosin and glidobactin A, we set out to identify metabolic engineering strategies that would improve production of both compounds. Despite differences in the PKS and NRPS components in prodigiosin and glidobactin biosynthesis, both compounds are synthesized from similar primary metabolites. Aided by the visible phenotype of prodigiosin production, we screened for metabolites that could be limiting prodigiosin production in *P. putida*. We hypothesized that results from these experiments could apply to improving glidobactin A production as well.

We grew cultures of *P. putida* PIG01 in RK media supplemented with compounds related to the metabolites directly incorporated into prodigiosin, including octanoic acid and amino acids (Figure 4.8a). Octanoic acid was the only supplement that improved specific prodigiosin production, suggesting that the availability of 2-octenoyl-ACP/CoA could be limiting prodigiosin production in *P. putida*. This effect was only apparent when

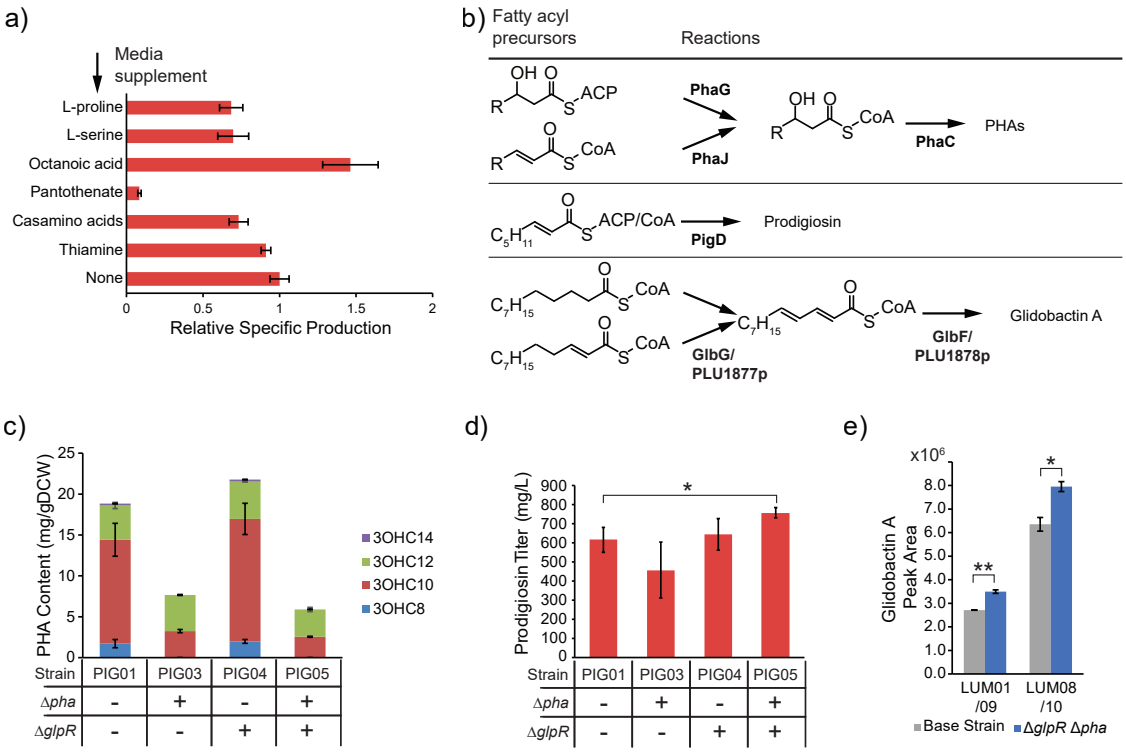


Figure 4.8: **Identification of fatty acyl precursors as a metabolic engineering target a)** Supplementing various precursors in minimal media (RK 2.5% glycerol) affects prodigiosin production. Cultures were grown in 3 mL media without dodecane overlay. **b)** Metabolites from fatty acid metabolism that are incorporated into PHA, prodigiosin, and glidobactin A biosynthesis. *P. putida* incorporates mostly C8-C12 3-hydroxyalkanoates into PHAs. 2-octenoyl-ACP or CoA is incorporated into prodigiosin, and dodecanoyl-CoA or 2-dodecenoyl-CoA is incorporated into glidobactin A. **c)** PHA composition from *P. putida* strains with and without Δ*glpR* and Δ*phaC1ZC2* genotypes. Cultures were grown in RK media with glycerol and extractions were completed at 24h of growth. **d)** Prodigiosin production from *P. putida* strains with and without Δ*glpR* and Δ*phaC1ZC2* genotypes. Cultures were grown in 25 mL RK glycerol with dodecane overlay for 48h. **e)** Effect of Δ*glpR* Δ*phaC1ZC2* genotype on glidobactin A production. Cultures were grown in 25 mL RK glycerol for 48h. Error bars represent standard deviation, n = 3 biological replicates. Differences between samples marked with asterisks were found to be statistically significant by the Student's t-test (n.s. = not significant, * = P<0.05, ** = P<0.01, *** = P<0.001).

glycerol was the carbon source, and not on glucose (Figure 4.3b). These results suggested to us that polyhydroxyalkanoate (PHA) biosynthesis was a potential carbon sink competing with the heterologous pathways for metabolites. Early reactions in both heterologous pathways incorporate an acyl-ACP or acyl-CoA into the natural product. *P. putida* synthesizes medium-chain-length PHAs (C8-C12) in order to store carbon and energy when an excess of carbon source is available [402]. Two PHA polymerases, PhaC-I and PhaC-II, are encoded in a single operon, *phaC1ZC2*, and incorporate 3-hydroxyalkanoates directly from fatty acid metabolism (Figure 4.8b).

We initially deleted *phaC1ZC2* in strain PIG01, resulting in PIG03, which success-

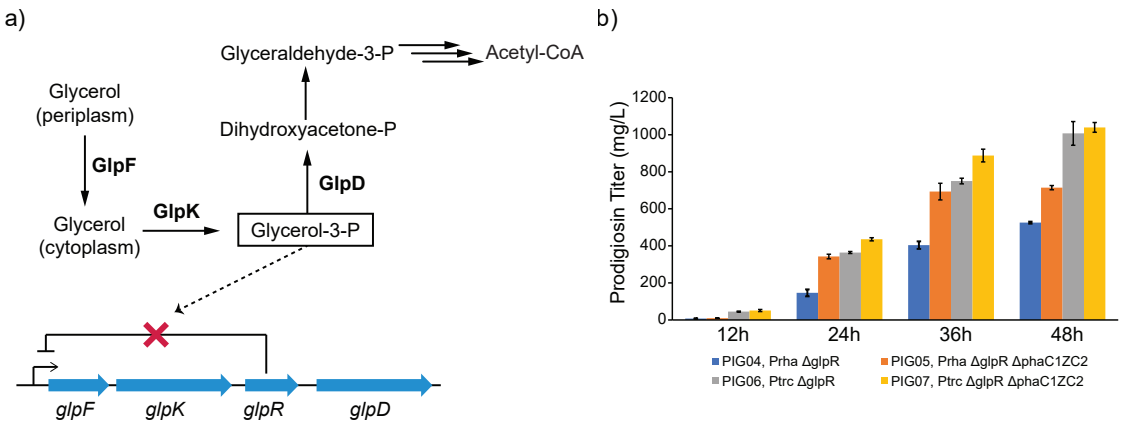


Figure 4.9: **Time course of strains with *glpR* and *phaC1ZC2* knockouts.** **a)** Regulation of glycerol catabolism by GlpR. GlpR represses expression of its own operon, including genes encoding the first three steps in glycerol uptake and catabolism. Glycerol-3-P, an intermediate in this pathway, inhibits repression by GlpR. **b)** Time course of prodigiosin titers from strains containing either Δ*glpR* only or Δ*glpR* Δ*phaC1ZC2* genotypes. Production media was RK glycerol, and strains were induced with 0.5% (w/v) L-rhamnose or 1 mM IPTG.

fully abolished PHA biosynthesis from *P. putida* (Figure 4.8c). However, PIG03 had lower prodigiosin production than the parent strain on RK media with glycerol (Figure 4.8d). Based on reports in the literature, we had also identified *glpR*, which encodes the transcriptional regulator that represses the expression of genes involved in glycerol uptake and catabolism in *P. putida* (Figure 4.9a) [403]. Escapa *et al.* found that deleting *glpR* decreased the length of lag phase during growth on glycerol as well as increased PHA production, so we hypothesized that this mutation could be beneficial for prodigiosin production as well. Deleting *glpR* alone in *P. putida* PIG01, resulting in PIG04, did not increase prodigiosin production, but introducing this mutation to PIG03, resulting in PIG05, slightly improved the prodigiosin titer to approximately 750 mg/L (Figure 4.8d). In contrast, introducing the Δ*glpR* Δ*phaC1ZC2* genotype to *P. putida* PIG02 did not improve production (Figure 4.9b). However, measuring prodigiosin titers at earlier time points did reveal that strains with both deletions had increased initial productivity of prodigiosin compared to strains without (Figure 4.9b).

To test our hypothesis that metabolic modifications improving prodigiosin production can be applied to other natural products, we deleted *glpR* and *phaC1ZC2* in LUM01, creating strain LUM09. This mutation had a more noticeable positive effect on glidobactin A production, resulting in a 30% increase in titer after 48h of growth as determined by

HPLC (Figure 4.8e). There was a similar benefit to production in the best glidobactin production strain, LUM08, resulting in a strain capable of producing 470 mg/L glidobactin A (LUM10).

4.4.8 Heterologous expression of other bacterial BGCs

P. putida strains chromosomally expressing another BGC from *P. luminescens* TT01 and two BGCs from Actinobacteria were also constructed. *P. luminescens* has the potential to produce anthraquinones, which are antitumor compounds synthesized by a type II PKS [404, 405]. The *ant* BGC encoding this pathway is 9,817 bp and has a GC content of 36%. Integrating this cluster with P_{rha} resulted in a production strain that produced a faint yellow pigment, which is indicative of anthraquinone production. Further analysis of this strain was not performed, such as UV absorbance measurements or LC-MS analysis, which is needed to confirm the identity of the pigment.

Attempts to produce natural products from actinobacterial BGCs in *P. putida* were not successful. The first compound, arylomycin, is produced by several *Streptomyces* strains and has antibiotic activity against Gram-negative bacteria [406, 407]. The BGC responsible for arylomycin biosynthesis in *Streptomyces filamentosus* NRRL 11379 is 32,353 bp and is 72% GC. However, the BGC contains two operons that are expressed in opposite directions. To refactor this BGC for expression in *P. putida*, one operon was integrated in the *pvdIJD* locus with P_{rha} and the other was integrated in the *pvdL* locus with P_{tac}. No product was detected by HPLC in acetone/ethyl acetate extractions from production cultures induced with IPTG and L-rhamnose.

The final product, capreomycin, is synthesized by *Saccharothrix mutabilis* subsp. *capreolus* ATCC 23892 and is used to treat multi-drug resistantant tuberculosis [35]. The BGC is structured as a single operon of 27,694 bp in length and has 76% GC content. After confirming that *P. putida* can grow in up to 4 mg/mL capreomycin in the RK production media, we integrated the *cmn* cluster into the *pvdIJD* locus with P_{rha}. We initially did not detect capreomycin in culture supernatants by HPLC, and created derivatives of this strain by replacing P_{rha} with P_{trc}, overexpressing the cognate MLP CmnN,

and overexpressing *sfp*. After performing a purification protocol specific for capreomycin on culture supernatants from all of these strains [408], capreomycin was still not detected by HPLC or disk diffusion assays.

4.5 Discussion

This work establishes several tools for rationally engineering natural product biosynthesis using *P. putida* as a heterologous host. After optimizing production through various genetic modifications, we achieved production titers greater than any previously reported values for heterologous production for both prodigiosin (1.1 g/L) and glidobactin A (470 mg/L) [387, 276, 309]. A major strength of our expression and production strategies is that most initial strains (i.e. strains with rha promoter and zero modifications to the BGC or host) were able to produce significant amounts of prodigiosin or glidobactin A, with titers that were on par or greater than values reported in the literature. Applying this approach to the heterologous expression of uncharacterized BGCs in *P. putida* could be a viable strategy for drug discovery. Combined with the simple metabolic background of *P. putida*, achieving functional expression of heterologous BGCs using a single integration provides a rapid workflow for building strains to analyze for novel compounds.

In cases where the initial heterologous production titers are lower than desired, there are multiple options for well-characterized promoters in *P. putida* [176]. We have shown that in the case of chromosomal expression, increasing the promoter strength provides an expected increase in heterologous production (Figure 4.1d, Figure 4.2). In contrast, a recent paper that attempted to express the *lum* BGC in *P. putida* using T7 RNA polymerase (T7 RNAP) did not detect production of glidobactin A [309]. Even though the T7 promoter is stronger than other bacterial promoters, transcription of long transcripts with T7 RNAP in *P. putida* may lead to poor mRNA stability because T7 RNAP, unlike bacterial RNAPs, is uncoupled from translation in bacteria [409, 410, 411]. T7 RNAP could also be detrimental to protein solubility, as it has been demonstrated in *E. coli* that switching from the T7 promoter to a bacterial promoter improved solubility and overall

expression of a PKS [412].

In contrast to transcription, translational control is rarely investigated in articles describing heterologous BGCs in bacteria. Here, we only tested two RBS sequences guided by *in silico* predictions for expressing the *glb* cluster (Figure 4.6d). A systematic analysis of RBS libraries in future studies would reveal the extent to which translational control can affect heterologous expression of BGCs [413]. Translation initiation of internal genes in heterologous BGCs also needs to be optimized, as demonstrated by the introduction of ATG start codons into the *lum* and *glb* clusters (Figure 4.7b-d). Mutating the start codons in heterologous genes expressed in *P. putida* could be a general strategy for optimizing BGCs from actinomycetes, which have a higher rate of GTG start codons compared to other bacterial phyla [414]. With the availability of mutagenic recombineering techniques for *P. putida*, it would be feasible to optimize pathway expression by mutating the 5' untranslated regions (UTRs) for individual genes [200]. Indeed, this strategy has been used recently to improve production of the natural product, violacein, in *E. coli* [415].

GC content is often cited as a concern when designing heterologous BGCs, but the GC content of the two glidobactin BGCs did not appear to be a major factor in heterologous production. After accounting for issues in translation initiation and MLP expression in the *glb* cluster, glidobactin A production in *P. putida* was still higher with the low-GC *lum* cluster (47% GC) compared to the *glb* cluster (70% GC) (Figure 4.7c,d). For comparison, the native BGC for pyoverdine biosynthesis in *P. putida* is 66% GC [73]. It is worth noting that *P. putida* and *P. luminescens* are both γ-proteobacteria while *S. brevitalea* belongs to the class of β-proteobacteria, so the phylogenetic relationship between the native host and the heterologous host may have a greater influence on heterologous production than GC content and codon usage. Furthermore, codon optimization has been shown in the literature to be detrimental to heterologous gene expression in *P. putida* [311]. For now, there is not enough knowledge on codon optimization for *P. putida* and other heterologous hosts to consistently rely on it for improving PKS and NRPS expression. Successful production with the *lum* and *ant* clusters suggests that *P. putida* could be an alternative host for expressing BGCs from lower GC bacteria, such as cyanobacteria,

which are becoming a more prevalent source for novel natural products [416]. Our failure to functionally express pathways from Actinobacteria does suggest that *P. putida* will not be a universal heterologous host for natural products. More work is needed to determine why expressing the arylomycin and capreomycin BGCs did not result in production. Most of the genes in these BGCs contain GTG start codons, so poor translation could be preventing production. Several *Streptomyces* bacteria are being developed as heterologous hosts [217, 268, 417], so future development of *P. putida* as a heterologous host should focus on BGCs from proteobacteria and other groups that do not yet have an established heterologous host.

The results presented here also demonstrate how *P. putida*'s native metabolism affects heterologous production. For instance, media that elicited a strong carbon catabolite repression (CCR) response in *P. putida*, such as glucose-based or rich media [418], resulted in lower production titers compared to glycerol-based media (Figure 4.1d, Figure B.3a). Bacteria use CCR to alter their metabolism to maximize growth rate and consumption of a preferred carbon source, which could prevent the diversion of metabolites needed for the production of a heterologous product [419]. An alternative hypothesis is that the carbons in glycerol are more reduced than those in glucose and increase the availability of NAD(P)H for biosynthesis of the acyl-ACP/acyl-CoA substrates incorporated into the heterologous products [420]. For example, *P. putida* has a higher composition of PHAs when grown on glycerol compared to glucose [421], suggesting that *P. putida* has higher fluxes through fatty acid metabolism during growth on glycerol. Furthermore, the absence of the *cyo* operon in PIG08 improved prodigiosin production on glucose to be on par with production from PIG02 on glycerol (Figure 4.4c). Inactivation of the *cyo* operon has been shown to alleviate CCR in *P. putida* [422, 423], and deleting a major component of oxidative phosphorylation could heavily impact the availability of reducing equivalents [424]. Improvements in prodigiosin production observed from Δcyo strains could be attributed to a partial disruption of CCR or an increased availability of NAD(P)H. Considering these results and other findings in the literature, glycerol is a promising carbon source for bioprocesses involving *P. putida*. In addition to being a

relatively inexpensive carbon source, glycerol catabolism causes a weaker stress response and is converted into biomass more efficiently in *P. putida* [394, 425].

Addition of a dodecane overlay improved prodigiosin production, especially in rich media (Figure 4.3a). We observed that prodigiosin cultures in TB media without the organic overlay appeared dark blue or black instead of red. This indicated an accumulation of the prodigiosin precursor, MBC, possibly resulting from poor production of MAP. Dodecane can act as a vector for dissolved oxygen in production cultures [426], and the last step in MAP biosynthesis, catalyzed by PigB, requires molecular oxygen. The reduction in prodigiosin titers in rich media without an organic overlay could be due to low oxygen availability impairing this reaction (Figure 4.3a). Production on glycerol was less dependent on the dodecane overlay, potentially making it an optimal carbon source for scaling up production cultures, as oxygen transfer is a common engineering problem in large-scale cultivations.

Another benefit to using *P. putida* as a heterologous host is that it is amenable to metabolic engineering for improved production of targeted molecules [15]. Deleting PHA production in *P. putida* demonstrated that fatty acid metabolism can affect heterologous product titers and potentially established a general chassis strain for producing natural products containing fatty acid precursors (Figure 4.8d,e). Deleting *phaC1ZC2* only increased natural product titers on glycerol when it was accompanied with a *glpR* deletion (Figure 4.8d). Removing PHA biosynthesis likely caused a shift in *P. putida*'s central metabolism that had a negative effect on heterologous production. It has been shown that deleting *phaC1* increases flux through the TCA cycle during growth on fatty acids, causing more carbon equivalents to be lost to CO_2 formation [427]. The increased TCA flux could have also reduced availability of amino acid precursors for prodigiosin biosynthesis. Both PHA synthases were targeted for gene repression in a recent study that optimized *P. putida* for heterologous production of indigoidine, a non-ribosomal peptide synthesized directly from glutamine [15]. This strategy demonstrates the utility of removing PHA biosynthesis from production strains to improve carbon yields, even for products that do not directly incorporate fatty acid precursors. The authors downregulated 12 other

native genes in the engineered strain, which likely prevented any deleterious effects from limiting PHA production. Our findings demonstrate that deleting pathways under strict native regulation could lead to adverse consequences, and strategies for re-shaping cellular metabolism, either through modifying the expression of regulatory genes or directly targeting the enzymes themselves, will be necessary to successfully improve production in the engineered strain.

Strategies for overproducing fatty acids in *P. putida* have recently been reported in the literature and could be applied towards metabolic engineering efforts for natural products [428, 205]. In addition, combinatorial engineering of enzymes in glidobactin A biosynthesis enabled the targeted production of derivatives with different chain lengths for the lipid tail, which can modulate the potency of glidobactin derivatives as an anticancer drug [74]. *P. putida* would be an ideal platform for furthering this work and engineering strains to produce novel derivatives of glidobactin A and other compounds.

4.6 Conclusions

The wealth of synthetic biology tools available to *P. putida* is its primary strength as a heterologous host. We generated multiple chromosomal deletions, insertions and point mutations in *P. putida* using a generalized genome editing toolkit. These methods enabled rational engineering of *P. putida* production strains to improve heterologous titers for both products, primarily through improving transcription and translation of the heterologous BGCs and overexpressing auxiliary proteins (MLPs and PPTases). Our efforts highlight simple and rational steps to improve heterologous production that could be applied to other heterologous hosts for improving expression and activity of PKSs and NRPSs. We also identified a carbon sink negatively affecting heterologous production and discovered a regulatory change required for improving productivity of prodigiosin and glidobactin A. As knowledge of metabolism and regulatory networks in *P. putida* continues to improve, increasing heterologous product titers from *P. putida* by engineering its native metabolism will be more common. In future studies, our methodology

described here could be applied towards uncharacterized BGCs or enzymes engineered to synthesize natural product derivatives, establishing *P. putida* as a platform for producing novel drug candidates.

4.7 Acknowledgments

This study was supported by research grants from the National Science Foundation (MCB-1716594). TBC and MVV are recipients of the NIH Biotechnology Training Program Fellowships (NIGMS 5 T32 GM08349). MVV is the recipient of a NSF Graduate Research Fellowship (DGE- 1256259). The Bruker Avance III HD 600 NMR spectrometer was supported by grant NIH S10 OD012245. The authors would like to thank Víctor de Lorenzo for providing the plasmid pSEVA2514-rec2-mut$L_{E36K}{}^{PP}$ and Francis A. Stewart for providing the strains *E. coli* GB2005 and *E. coli* GB05-dir. The authors would also like to thank William Cordell for his assistance in designing the glidobactin extraction protocol.

Chapter 5

Characterization and Manipulation of Siderophore Systems

Authors and Contributors:

- **Taylor B. Cook** Performed EntF library construction and selection. Characterized NRPS variants *in vitro*. Performed pyoverdine-related experiments.
- **Dr. Vladimir Vinnik** Performed EntF library construction and selection. Characterized EntF variants *in vitro*. Wrote the enterobactin and fabrubactin manuscripts.
- **Dr. Kurt Throckmorton** Performed EntF library construction and selection. Wrote the enterobactin manuscript.
- **Sophia A. Liu** Purified *Pf* Ago and performed *in vitro* assays.
- **Elizabeth Ciborowski** Constructed and analyzed pyoverdine efflux pump knockout strains.
- **Dr. Michael G. Thomas** Provided guidance, edited and reviewed the manuscript.
- **Dr. Brian F. Pfleger** Provided guidance, edited and reviewed the manuscript.

Portions of this chapter were originally published in two articles in *ACS Chemical Biology* [429, 430].

5.1 Abstract

High-throughput methods for analyzing NRPSs and PKSs are rare or require automation capabilities that are not accessible to most research groups. Several bacteria use these classes of enzymes to produce siderophores, which are iron-chelating compounds secreted by microorganisms to transport iron across the cell membrane. Selections can be designed around siderophore systems so that cell growth is dependent on siderophore production, providing an opportunity to perform enzyme directed evolution. Based on this motivation, we manipulated the biosynthesis of several bacterial siderophores to investigate substrate specificity of the "gatekeeper" domains in NRPS and PKS biosynthesis. First, we performed directed evolution on the serine-specific adenylation domain in EntF, an enzyme required for enterobactin biosynthesis in *E. coli*. This approach revealed the functional sequence space of the binding-site residues impacting L-serine activation, and we characterized variants of other adenylation domains based on these results. Second, we performed subdomain swaps in adenylation domains involved in pyoverdine synthesis in *P. putida*. Chimeric enzymes with varying activity levels highlighted important residues within the adenylation domain and suggested that a subdomain-swap strategy may be a viable route for NRPS engineering in pyoverdine biosynthesis. Lastly, we characterized a heme peroxidase involved in the biosynthesis of fabrubactin, a siderophore produced by *Agrobacterium fabrum* using a hybrid PKS/NRPS. Fabrubactin provides an avenue for directed evolution studies on acyltransferases, the gatekeeper domain in polyketide biosynthesis.

5.2 Introduction

Many bacteria secrete siderophores to capture iron in the environment and transport it across the cell membrane. Iron is essential for several cellular processes, such as oxidative phosphorylation. These compounds are secreted when bacteria are in iron-deficient conditions, and each species has evolved siderophores with unique structures to prevent competing species from uptaking siderophore-iron complexes. When siderophores are synthesized by pathogenic bacteria, siderophore production is often required for virulence [431]. In the laboratory, researchers can devise growth conditions where siderophore production is required for cell growth. Growth-based selections are an especially powerful method for processing large libraries of genetic variants, enabling directed evolution studies of proteins and pathways linked to cell growth.

NRPSs are commonly involved in the biosynthesis of bacterial siderophores. These are modular enzymes with multiple catalytic domains, and nonribosomal peptides are often synthesized in an assembly-line fashion where the sequence of modules determines the amino acids incorporated into the nascent peptide. The modular structure of NRPSs has invited many engineering attempts with the goal of changing substrate specificity and synthesizing natural product derivatives. The adenylation (A) domain is commonly referred to as the "gatekeeper" in NRPS enzymology. Structural information has revealed the binding site residues responsible for substrate recognition [66, 48]. The binding site consists of 10 residues that are situated in a compact subdomain within the A domain, and 8 residues vary depending on the amino acid substrate (Figure 5.1c-e). Bacterial A domains have evolved so that binding sites with the same specificity have a similar "specificity code". However, attempts to modify A domain specificity by mutating only the binding site residues has been limited to specificity changes between structurally similar amino acids [133]. More knowledge on how A domains recognize their preferred substrate is needed to enable future engineering strategies that reliably switch A domain specificity.

Directed evolution has been used recently to isolate functional NRPS variants. Random mutagenesis of chimeric NRPSs can yield variants with improved activity because

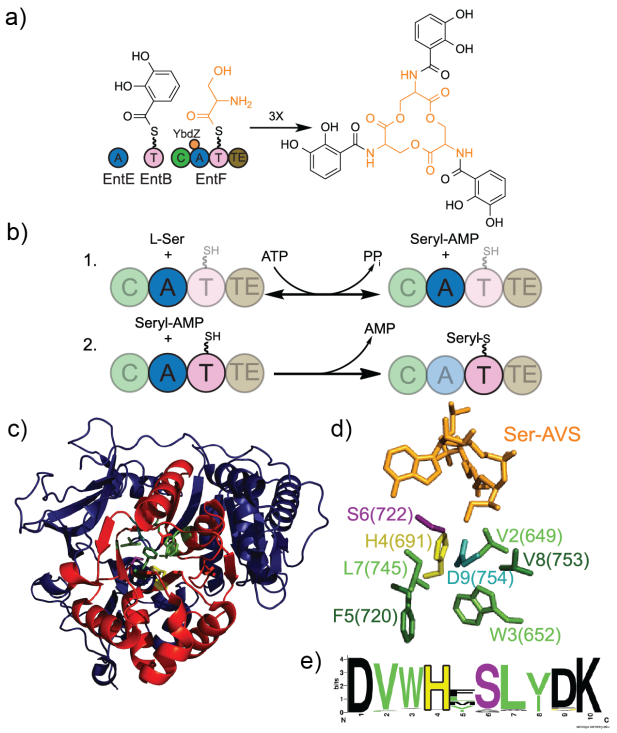


Figure 5.1: **Enterobactin (ENT) formation and the structure, specificity code, and function of the A domain of EntF. a)** Diagram of the ENT biosynthetic pathway. EntE tethers 2,3-dihydroxybenzoic acid (DHB) to EntB. The condensation (C) domain of EntF condenses DHB and L-Ser, previously activated and bound to the thiolation (T) domain of EntF. After one turnover, the DHB-L-Ser monomer is stored on the thioesterase (TE) domain. After three iterations of this process, the final product is cyclized and released. **b)** Two half-reactions of the EntF A domain. In the first, the A domain activates L-Ser as Seryl-AMP (1) and transfers the seryl group to the T domain (2) in the second. **c)** Ribbon representation of the EntF (Protein Data Bank entry 5JA1) A domain with the recognition subdomain (red) and specificity code residues (highlighted in colors matching those of panels D and E). **d)** Binding pocket residues that form the specificity code and the reaction intermediate mimic inhibitor, seryl-AVS [432], used for crystallization. **e)** WebLogo [433] of specificity code sites 1-10 for 82 characterized L-Ser codes [48].

domain swaps often disrupt important but poorly understood protein-protein interactions within the enzyme [123]. Libraries created by saturation mutagenesis of the specificity code residues have enabled the activation of non-cognate substrates, but activity on the original substrate is often retained [20, 134]. Methods based on yeast surface display have enabled rapid screening of variant libraries and the isolation of A domain variants with specificity switched to β-amino acids over α-amino acids [130]. An alternative strategy for modifying A domains is through subdomain swaps [128, 78]. These examples define a compact region containing 9 of the 10 active site residues and replace it with a homologous region from A domains with different specificity. *In vitro* assays demonstrated that chimeric A domains have the same specificity profile as the donor A domain.

Enterobactin is a siderophore that is secreted by *E. coli* and synthesized by an NRPS [434]. Biosynthesis involves the condensation of 2,3-dihydroxybenzoic acid with L-serine, followed by cyclization of three dipeptides to form a triketide lactone ring (Figure 5.1a).

The enzyme responsible for L-serine activation, EntF, is a complete NRPS module containing a condensation (C), adenylation (A), thiolation (T), and thioesterase (TE) domain. The A domain of EntF is responsible for L-serine activation and incorporation into enterobactin (Figure 5.1b). The residues that directly interact with L-serine during substrate recognition are known because a crystal structure of EntF with the A domain in complex with a transition state mimic, serine adenosine vinylsulfonamide (Ser-AVS), has been solved (Figure 5.1c,d) [140]. The EntF A domain has a specificity code of DVWHF-SLVDK, which is equivalent to the most common residues found in Ser-activating A domains found in Nature (Figure 5.1e).

To investiate L-serine-activation in enterobactin biosynthesis, we generated libraries of EntF variants by saturation mutagenesis of the active site residues. These efforts resulted in 157 amino acid-unique EntF variants with serine-specificity. The sequences of these variants expand on the sequences of A domains in Nature known to activate L-serine. Based on these findings, we designed A domain variants in other NRPSs and characterized their specificity *in vitro*, establishing that A domains are likely to share this extended sequence space. We also performed complementary experiments that interrogated subdomain swaps within the pyoverdine NRPS from *P. putida*. These experiments demonstrated that A domains in the pyoverdine NRPS are amenable to subdomain swaps and highlighted important residues for enzyme activity. Lastly, we identified a heme peroxidase responsible for L-DOPA formation in the biosynthesis of fabrubactin in *Agrobacterium fabrum*. Elucidation of this pathway will enable future directed evolution studies on PKSs in addition to NRPSs.

5.3 Materials & Methods

5.3.1 Plasmids and bacterial strains

Bacterial strains and plasmids are listed in Table 5.1. *E. coli* strains were grown in LB at 37°C and *P. putida* strains were grown in LB at 30°C. All pyoverdine production and

complementation experiments were performed with King's B broth or agar supplemented with Gent30 and various concentrations of EDDHA. All cultures and petri dishes were incubated at 30°C. *P. putida* growth was monitored by OD600 and pyoverdine production was monitored by fluorescence intensity (excitation: 398 nm; emission: 455 nm). Plasmids were constructed via Gibson cloning, restriction ligation cloning, polymerase incomplete primer extension (PIPE), or RecET direct cloning [382, 435].

5.3.2 EntF Library Creation

The RS-encoding fragments of *entF* were amplified using primers containing NNK codons corresponding to the residues targeted for mutagenesis. The mutagenized RS-encoding fragments were combined by overlap extension PCR and ligated into pACYC184entF-ES-RS_{sub}. Ligations were electroporated into NEB DH10β cells. Transformants were pooled, and pACYC184entF-ES-L1-4 plasmids, consisting of 2.5, 3.5, 2.5, and 7.5 million transformants, respectively, were recovered.

5.3.3 Selection and isolation of enterobactin producers

Plasmids pACYC184entF-ES-L1-4 were electroporated into BW27749 Δ*entF* cells and incubated on M9 minimal medium noble agar plates with 0.4% (v/v) glycerol, 1 μM EDDHA (Complete Green Company), and an antibiotic [chloramphenicol (34 μg mL^{-1}) or streptomycin (100 μg mL^{-1})]. Chloramphenicol was used for pACYC184entF-ES-L1 and -2, and pACYC184entF-ES-L3- and -4 were switched to streptomycin to eliminate a chloramphenicol-resistant pACYC184entF-wildtype contaminant. For technical reasons, libraries were not saturated. Most notably, the frequency of observation of wild-type contamination was much higher than that of successful transformants for library 3, which had the most restrictive combination of sites targeted for mutagenesis. This wild-type contamination was traced to the reversion of the *recA1* allele in commercially purchased competent cells, followed by recombination with wild-type *entF* in the chromosome of these cells during library construction. Enterobactin producer colonies were streaked for

Table 5.1: Strains and plasmids used in this work.

Strain or plasmid	Genotype and relevant characteristics	Source
Strains		
E. coli DH10β	Δ(*ara-leu*) *7697 araD139 fhuA* Δ*lacX74 galK16 galE15 e14- ϕ80dlacZΔM15 recA1 relA1 endA1 nupG rpsL* (Str^R) *rph spoT1* Δ(*mrr-hsdRMSmcrBC*)	New England Biolabs
E. coli DH5α	F- Φ80*lacZ*ΔM15 Δ*(lacZYA-argF)* U169 *reacA1 endA1 hsdR17* (r_{k-} m_{k+}) *phoA supE44 thi*-1 *gyrA*96 *relA*1 λ-	Invitrogen
E. coli BW27749	F-, Δ(*araD-araB*)*567*, Δ*lacZ4787*(::rrnB-3), λ-, Δ(*araH-araF*)*570*(::*FRT*), Δ*araEp531*::kan, ϕP_{cp8}*araE535*, *rph-1*, Δ(*rhaD-rhaB*)*568*, *hsdR514*	[351]
E. coli BW27749 Δ*entF*	BW27749 with in-frame deletion of *entF*	This study
E. coli BL21 (DE3)	F-, *ompT, gal, dcm, lon, hsdSB*(*rB- mB-*), λ(DE3[*lacI*, *lac*UV5-T7 gene 1, ind1, *sam7, nin*5])	[68]
E. coli BL21 (DE3) Δ*ybdZ*	BL21 (DE3) *ybdZ::aac(3)IV*	[68]
E. coli MG1655	K-12 F- λ- *ilvG- rfb-50 rph-1*	ECGSC
E. coli GB2005	DH10β Δ*fhuA* Δ*ybcC* Δ*recET*	[84]
E. coli GB05-dir	GB2005 Δ*ybcC*::P_{BAD}*ETgA*	[84]
E. coli GB05-red	GB2005 Δ*ybcC*::P_{BAD}*abgA*	[84]
P. putida KT2440	Wild-type	ATCC 47054
P. putida KTU	KT2440 Δ*upp*	[158]
P. putida KTU Δ*pvdD*	KT2440 Δ*upp* Δ*pvdD*	This study
Plasmids		
pBBR1k-GFPuv	IPTG-inducible P_{trc} expressing *gfpuv* on BBR1 origin	This study
pACYCduet-ybdZ	*ybdZ* MLP co-expression vector	[68]
pACYCduet-PA2412	PA_2412 MLP co-expression vector	[253]
pET28b	T7 expression vector	[351]
pET28bentF-ES-wildtype	T7 expression vector with entF gene containin EagI/SacI restriction sites	This study
pET28bdltA-wildtype	Expression of dltA wild-type	This study
pET28b-entF-ES-variant	Expression of EntF variants described in this study	This study
pACYC184	Plasmid backbone used for all selections	ATCC
pACYC184entF-ES	Complementation vector containing entF with EagI/SacI restriction sites	This study
pACYC184entF-ES-RS_{sub}	EntF complementation vector containing random stuffer region in place of EntF subdomain. Used for cloning *entF* libraries.	This study
pET28b-dhbF1	T7 expression of first module of DhbF	This study
pET28b-dhbF2	T7 expression of second module of DhbF	This study
pMP11	Cas9 counterselection and λRed recombineering in *E. coli*	[19]
pgRNAcm-loxP-RFPtemp	Guide RNA plasmid towards *loxP* containing RFP repair template	This study
p15A-RFPtarget	Target plasmid for plasmid recombination studies with *loxP* scar and incomplete RFP, medium-copy origin	This study
pRK2-RFPtarget	Target plasmid for plasmid recombination studies with *loxP* scar and incomplete RFP, low-copy origin	This study
pgRNAcm-pvdDshell2	Guide RNA plasmid towards subdomain of A domain in second module of *pvdD* from *P. putida*	This study
pgRNAcm-pvdDshell2-EntF500	pgRNAcm-pvdDshell2 with repair template for inserting EntF subdomain into *pvdD*	This study
pRK2T-Ptrc-PP_4219-gen	*pvdD* complementation vector	This study
pRK2T-Ptrc-PP_4219shell1-TcR-gen	*pvdD* complementation vector with TcR stuffer inserted into A domain of first module	This study
pRK2T-Ptrc-PP_4219shell2-TcR-gen	*pvdD* complementation vector with TcR stuffer inserted into A domain of second module	This study
pRK2T-Ptrc-PP_4219shell2-EntF-gen	*pvdD* complementation vector with EntF subdomain inserted into A domain of second module	This study
pRK2T-Ptrc-PP_4219shell2-EntFmt-gen	*pvdD* complementation vector with EntF subdomain inserted into A domain of second module, EagI/SacI boundaries	This study
pRK2T-Ptrc-PP_4219shell2-PvdJA1tc-gen	*pvdD* complementation vector with PvdJA1 subdomain inserted into A domain of second module	This study
pRK2T-Ptrc-PP_4219shell2-PvdJA1mt-gen	*pvdD* complementation vector with PvdJA1 subdomain inserted into A domain of second module, EagI/SacI boundaries	This study
pRK2T-Ptrc-PP_4219shell1-DhbFA1tc-gen	*pvdD* complementation vector with DhbFA1 subdomain inserted into A domain of second module	This study
pRK2T-Ptrc-PP_4219shell1-DhbFA1mt-gen	*pvdD* complementation vector with DhbFA1 subdomain inserted into A domain of second module, EagI/SacI boundaries	This study
pET28a-PfAgo	PfAgo overexpression vector	[81]
pTEV5	Empty T7 expression vector	[436]
pTEV-pvdD_A2	Expression vector of *pvdD*, A domain from module 2	This study
pTEV-pvdD_A2-EntF	Expression vector of *pvdD*, A domain from module 2 with EntF subdomain	This study
pJOE_pvdRTopmQ	Suicide vector for deleting *pvdRTopmQ* in *P. putida*	This study
pRK2T-pvdRTopmQ-gen	Expression vector for *pvdRTopmQ*	This study

Table 5.2: Oligos used in this work. /5Phos/ indicates that the oligo is 5' phosphorylated.

Oligo	Sequence	Description
rTBC0072	CACCCTTAGAggtgatgtcgacgcaaccttcctggttagaCTCGAGTAAGGATCTCC	pRK2T-pBAD forward primer for *pvdD*
rTBC0073	CACCCTTAGAtgactccacggaatcgagcaatgcttgcacatgtatatctccttcttaaa	pRK2T-pBAD reverse primer for *pvdD*
rTBC0074	cgataccgtgggtgcgtaccttgcccgaggcttcgtagcaggtgatgtcgacgcaacctt	Forward extension primer for *pvdD*
rTBC0075	tgagcaaggcggccagggccttgcgctcccgggacgacagtgactccacggaatcgagca	Reverse extension primer for *pvdD*
rTBC0471	/5Phos/TGAAGGAGATATACAT	PfAgo pBAD 5' guide 1
rTBC0472	/5Phos/TTGTATATCTCCTTCT	PfAgo pBAD 5' guide 2
rTBC0473	/5Phos/TAACTCGAGTAAGGAT	PfAgo pBAD 3' guide 1
rTBC0474	/5Phos/TTCCTTACTCGAGTTT	PfAgo pBAD 3' guide 2
rTBC0460	/5Phos/TAATTCAAAAGATCTT	PfAgo nicking guide

isolation on M9 plates with 0.4% (v/v) glycerol, 1 μM EDDHA, and an antibiotic. From each, four colonies were incubated in a 96-well plate containing M9 with 0.4% (v/v) glycerol, 50 μM EDDHA, and an antibiotic. Plasmid preps were performed from colonies that grew. Plasmids were screened by digestion to eliminate pACYC184entF-wildtype (non-EagI or SacI). Each correct construct was re-transformed into BW27749 Δ*entF* and selected in liquid M9 with 50 μM EDDHA. Plasmid DNA was recovered, and the RS-encoding region was sequenced.

5.3.4 Phenotypic characterization of EntF variants

MIC_{EDDHA} assays were performed in technical triplicates. BW27749 Δ*entF* strains were grown overnight in LB, subcultured into LB, grown until an OD600 of approximately 0.5 was reached, and normalized to an OD600 of 0.5, and 0.5 μL was inoculated into 200 μL of M9 medium with 0.4% (v/v) glycerol, an antibiotic, and 250, 350, 400, 450, 500, or 600 μM EDDHA. Cultures were inoculated through an Excel Scientific AeraSeal sterile membrane, covered with a second membrane, and incubated at 37°C and 250 rpm for 45 h. If the OD600 increased to $\geq$0.2, from a calculated starting OD600 of 0.01, the strain was considered to have grown. Strains that grew at 600 μM EDDHA were tested at 600, 650, 700, 750, 800, and 850 μM EDDHA. Strains that did not grow at 250 μM EDDHA were tested at 50, 100, 150, and 200 μM EDDHA.

5.3.5 Overproduction and purification of EntF, DltA, DhbF, and PvdD variants

EntF variants were co-overproduced in *E. coli* BL21(DE3) with a C-terminal hexahistidine tag with the MLP, YbdZ. Other NRPS proteins were overproduced in E. coli BL21(DE3) *ybdZ::acc(3)IV* with a C-terminal hexahistidine tag with the corresponding MLP where applicable. To overproduce proteins from each of these constructs, 2L LB (supplemented with antibiotic) cultures of each overexpression strain were grown at 25°C until an OD600 of 0.5 was reached. The cells were shifted to 4°C for 1 hr, isopropyl β-D-1-thiogalactopyranoside (IPTG; 60 μM) was added, and cells were shifted to 15°C and grown for 15 hr. Cells were harvested by centrifugation (10 min at 10,000 rpm) and resuspended in buffer A (Tris-HCl pH8.0 [20 mM], NaCl [300 mM], glycerol [10% v/v]) containing imidazole (5 mM). The cells were broken by sonication (Fisher 550 Sonic Dismembrator, power = 5, 15 min sonication with 1 sec on, 1 sec off). Cell debris were removed by centrifugation (30 min at 15,000 rpm). Clarified cell extract was incubated with 1 ml of Ni-NTA resin (Qiagen) for 2 hr on ice with gentle rocking. The resin was recovered and washed with buffer A containing imidazole (20 mM). His-tagged proteins were eluted with a step gradient of buffer A plus varying concentrations of imidazole (40, 60, 80. 100, 250 mM). Fractions containing His-tagged protein, based on SDS-PAGE/Coomassie Blue staining, were pooled and dialyzed at 4°C overnight against buffer B (Tris HCl, pH 8.0 [50 mM], NaCl [100 mM], glycerol [10% v/v]) using 3500 MWCO Pleated Dialysis Tubing (Thermo Scientific). The protein was flash frozen in liquid nitrogen. Protein concentrations were determined by the BCA assay (Pierce).

5.3.6 Radiolabeled ATP/PP_i Assays of NRPS Variants

Amino acid-dependent ATP-PP_i exchange assays were performed as previously described [437]. Briefly, each 100-μL reaction mixture contained 75 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 5mM dithiothreitol, 3.5 mM ATP(pH 7.0), 1 mM [32P]PP_i (0.9 Ci/mol), and various concentrations of amino acid and enzyme. For amino acid specificity studies all amino acids were present at 1 mM. Enzyme concentrations were at 50 μM, and each

reaction lasted 30 minutes.

5.3.7 RecET direct cloning of *pvdD* complementation vector

The plasmid pRK2T-Ptrc-PP_4219-gen was constructed using a protocol based on RecET direct cloning as described previously [84, 80]. PP_4219 (*pvdD*) was isolated from purified genomic DNA from *P. putida* KT2440 by restriction digestion with NotI-HF and HindIII-HF. Digested genomic DNA was purified by ethanol precipitation. Linearized pRK2T-Ptrc-gen with homology arms to *pvdD* was prepared with a two-step PCR protocol. First, primers annealing to the promoter and terminator regions of pRK2T-Ptrc-RFP-gen were used to amplify the plasmid backbone. These primers contained 40-bp overhangs with homology to the 5' and 3' ends of *pvdD* plus 10 random bp. A second pair of primers annealing 10 bp from the ends of the PCR product was used to amplify the backbone again, adding another 40 bp of homology to the PCR product. The final PCR product had 80 bp of homology to either end of the *pvdD* sequence released from *P. putida* genomic DNA by digestion.

To prepare competent cells for RecET direct cloning, 5 mL of fresh LB were inoculated with 150-μL of overnight culture of *E. coli* GB05-dir. Cultures were incubated at 37°C for about 2 hours, and then induced with 100μL 20% (w/v) L-arabinose. Growth continued for another 45 minutes and then the cultures were placed on ice. For every aliquot of competent cells needed, 1mL of culture was centrifuged for 30 seconds at 11,000g and washed with 1mL 10% (v/v) glycerol. The samples were washed with 1mL 10% glycerol two more times, and then resuspended in 20μL 10% glycerol. At least 500ng of linearized vector and 5μg digested genomic DNA were added to each aliquot of competent cells to a total approximate volume of 50 - 60μL. The competent cells were then added to chilled 1-mm electroporation cuvettes and electroporated with a voltage of 1.8kV, yielding time constants of 3 - 4 ms. After electroporation, cells were immediately mixed with 1mL SOC media and incubated at 37°C for 90 minutes. All of the recovered cells were plated on LB Gent30 and incubated overnight at 37°C. Plasmid DNA was prepped from colonies and screened by digest analysis.

5.3.8 Cas9-assisted plasmid recombination

Plasmid recombination experiments were performed with *E. coli* GB2005 containing pMP11 and pRK2T-Ptrc-PP_4219-gen. To prepare electrocompetent cells, 5 mL of fresh LB were inoculated with 150-μL of overnight culture of *E. coli* GB2005. Cultures were incubated at 30°C for about 2 hours, and then induced with 100μL 20% (w/v) L-arabinose. Growth continued for another 45 minutes and then the cultures were placed on ice. For every aliquot of competent cells needed, 1mL of culture was centrifuged for 30 seconds at 11,000g and washed with 1mL 10% (v/v) glycerol. The samples were washed with 1mL 10% glycerol two more times, and then resuspended in 20μL 10% glycerol. Five hundred nanograms of pgRNAcm-pvdDshell2-EntF500 was transformed by electroporation in 1-mm cuvettes with a voltage of 1.8kV. Cells were allowed to recover for 3 hours and plated on LB Carb100, Gent30 and incubated overnight at 30°C. Colonies were initially screened by colony PCR. To purify isogenic plasmid DNA, colonies were cultured overnight in LB media and plasmid DNA was prepped. The DNA was re-transformed into chemically competent *E. coli* GB2005. Cell-culturing steps were performed at 37°C to prevent replication of pMP11, which is temperature sensitive. Individual colonies were screened again by colony PCR and later digest analysis after purifying plasmid DNA.

5.3.9 *Pf* Ago production, purification, and *in vitro* assays

Pf Ago was overproduced from 1-L LB cultures (containing Kan50) of *E. coli* BL21 (DE3) with pET28a-PfAgo. Cultures grown at 37°C were induced with 100 μM IPTG when they reached an OD600 = 1.0, cooled down to 30°C, and incubated overnight at the same temperature. Cells were harvested by centrifugation (10 min at 10,000 rpm) and resuspended in binding buffer (100 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8.0, 10% (v/v) glycerol). The cells were broken by sonication (Fisher 550 Sonic Dismembrator, power = 5, 15 min sonication with 1 sec on, 1 sec off). Cell debris were removed by centrifugation (30 min at 15,000 rpm). Clarified cell extract was filtered with a 0.4 μm filter and passed through a 1-mL StrepTrap HP column using an Akta Start. Gradient

elution from 0 to 100% elution buffer (binding buffer containing 2.5 mM desthiobiotin) was used to elute Strep-tagged *Pf*Ago from the column. Fractions containing *Pf*Ago were pooled and buffer exchanged by dialysis into storage buffer (20 mM Tris- HCl, pH 8.0, 300 mM NaCl, 0.5 mM $MnCl_2$, 15% (v/v) glycerol). Aliquots were stored at -80°C.

In vitro reactions with linear DNA were completed in 25 μL *Pf*Ago reaction buffer (20 mM HEPES pH 7.50, 250 mM NaCl, 0.5 mM $MnCl_2$) containing 5% (v/v) DMSO. The reactions contained 2.5 μM of each phosphorylated oligo, 0.5 μg plasmid DNA, and 0.06 - 0.24 μM *Pf*Ago. Reactions were incubated at 95°C for 10 minutes and slowly cooled down to 10°C at 0.1°C/s. Reactions with circular DNA were completed with 0.24 μM *Pf*Ago and 5 μM nicking oligo (5'-phosphorylated) in addition to the targeting oligos.

5.3.10 L-DOPA Dioxygenase Assay

Supernatants from cultures of *E. coli* BL21(DE3) overexpressing *fbnLM* were analyzed with a colorimetric and fluorometric L-DOPA dioxygenase (DOD) assay. Pre-cultures were prepared by growing *E. coli* BL21(DE3) containing pTEV-fbnLfbnM or pTEV5 overnight at 37°C. Fresh culture tubes containing 5 mL LB with carbenicillin (100 μg mL^{-1}) were inoculated to OD600nm=0.01. Cultures were incubated at 30°C while shaking. Once the OD600 reached approximately 0.5, FbnLM overexpression was induced with 100 μM IPTG. The cultures grew for another 5 hr, and at the end of incubation cells were removed from the cultures by centrifugation and the supernatants were sterilized by filtration.

The 5X DOD reaction buffer was prepared by mixing 5 mL 10X DOD buffer, 3 mL water, 0.088 g L-ascorbic acid, and 50 μL 100 mM $FeCl_2$. The pH was adjusted to 6.0 using 5 M NaOH, and water was added to a final volume of 9.5 mL. 10X DOD buffer contains 500 mM morpholinoethanesulfonic acid (MES) monohydrate, 1 M NaCl, 1 M tricine, and 100 mM glycine. The final 5X DOD reaction master mix was prepared by mixing 237.5 μL of the 5X DOD reaction buffer with 12.5 μL of 2.5 mg mL^{-1} 6xHis-MBP-L-DOPA dioxygenase in 50% (v/v) glycerol with storage buffer. The storage buffer contains 250 mM NaCl and 10 mM Tric-Cl.

Every sample was assayed by mixing 80 μL with 20 μL 5X DOD reaction master mix and incubating at 30°C for up to 16 hr. Reactions were monitored by measuring absorbance at 420 nm and 470 nm and fluorescence (excitation: 470 nm, emission: 515 nm). A standard curve was prepared by adding 1000, 500, 250, 100, 50, and 0 μM L-DOPA to supernatant from *E. coli* BL21(DE3) containing pTEV5.

5.3.11 *In vitro* L-DOPA formation assays

Fifty-microliter reactions contained 100 mM Tris-HCl (pH 7.5 at 25°C), 100 mM NaCl, 20 mM H_2O_2, 500 μM L-Tyr, 2 mM dithiothreitol, 5 μg FbnL/M. Reactions were run for 30 min and then stopped by the addition of 15 μL of 20% (w/v) trichloroacetic acid. *In vitro* assays were analyzed by HPLC with a Shimadzu HPLC system (Shimadzu Co., Columbia, MD, USA) equipped with a fluorescence detector. HPLC separations were performed using a Luna C18 column (250 x 4.6 mm; 5 μm particle size). The mobile phase was 2.4% acetonitrile and 0.08% formic acid in water, the flow rate was 1.0 mL min^{-1}, and the injection volume was 2 μL. L-DOPA and L-tyrosine were detected using the fluorescence detector (excitation: 281 nm, emission: 314 nm).

5.4 Results & Discussion

5.4.1 Novel L-serine-specific codes in EntF libraries

We chose to mutagenize 8 of the 10 active site residues among 4 EntF libraries. The first and last residues, Asp649 and Lys952, were kept constant because they are highly conserved in all A domains and interact directly with the carboxy and amino groups of the amino acid substrate (Figure 5.1d). Each library had five mutagenized positions, and in Library 1 (L1), residues 2-4, 6, and 7 were mutagenized. Residues 3, 4, 6, 7, and 9 were mutagenized in Library 2 (L2). These residues were targeted first because positions 5 and 8 in L-serine-specific A domains had more variability in Nature, and focusing on the positions with consensus residues increased the likelihood of isolating variants with

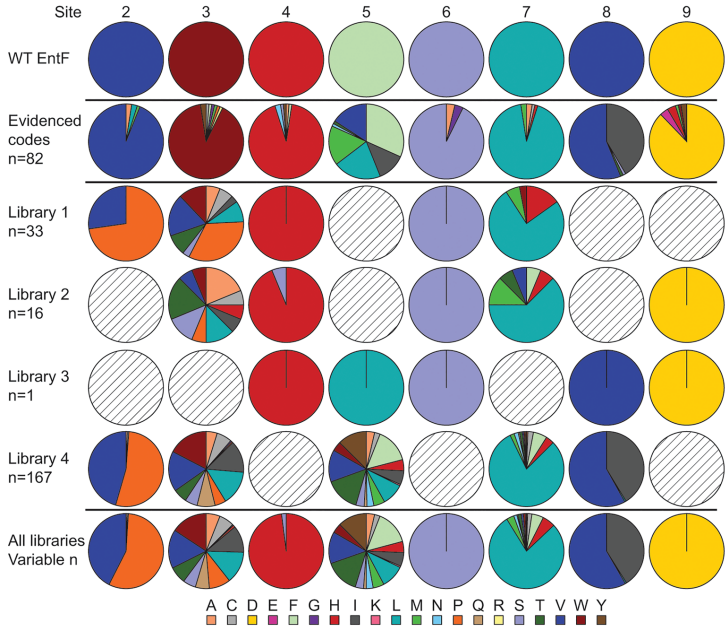


Figure 5.2: **Residue usage by code site in the EntF variants and characterized L-Ser-specific A domains.** Pie charts describing, by site in the specificity code, the amino acid usage across all libraries. The wild-type EntF residues are shown in the top row followed by the 82 characterized L-Ser-specific A domain residues and, subsequently, each of the four libraries. Circles with diagonal lines designate nonmutagenized sites. The color of each sector corresponds to the key at the bottom; both are organized alphabetically, clockwise, and from left to right. For each library, n designates the number of isolated strains containing DNA-unique entF mutants, provided they had no non-code residue substitutions. For all libraries, n = 200, 216, 50, 168, 50, 216, 168, and 17 for sites 2-9, respectively.

unique specificity codes [48].

We transformed these libraries into an *E. coli* $\Delta entF$ strain and plated the transformants on minimal agar with 1 μM ethylenediamine-di(o-hydroxyphenylacetic acid) (EDDHA), an iron-chelating compound that requires *E. coli* to produce enterobactin in order to grow. Strikingly, isolated variants from these libraries revealed that positions 2, 3, and 7 were more amenable to variation than one would expect from known L-serine-specific A domains in Nature (Figure 5.2). Between L1 and L2, 9 additional residues were observed at position 3. Some variants contained L-proline at positions 2 and/or 3, another surprising result because L-proline is one of the most rare residues among all A domain specificity codes [48].

Positions 4, 6, and 9 were found to be mostly invariant, suggesting that these residues were more important for L-serine activation in EntF. These results prompted us to construct two more libraries. Library 3 (L3) had positions 4-6, 8, and 9 mutagenized. In

contrast, Library 4 (L4) focused on the residues expected to have more variation based on characterized A domains and the library results so far: positions 2, 3, 5, 7, and 8. Only one functional DNA-unique mutant was isolated from L3, whereas L4 resulted in 167 DNA-unique mutants, supporting the hypothesis that the residues mutagenized in L3 are mostly invariable (Figure 5.2). The results from L4, like those from L1 and L2, showed increased variability in positions 2, 3, and 7 compared to natural specificity codes. The residue variability in positions 5 and 8 more closely matched that of evidenced codes, although site 5 is more variable in the EntF libraries than natural codes.

It is clear from the sequences of functional EntF variants that the three-residue combination of His4-Ser6-Asp9 is required for L-serine activation, as only one variant out of 157 broke this "rule". A recent structure of EntF bound to an intermediate substrate mimic shows that these three residues are closest to the hydroxyl group of the amino acid side chain, underscoring the importance of these residues in L-serine recognition [140].

5.4.2 *In vivo* enterobactin production vs. *in vitro* activity of A domain variants

Strains isolated from the selection described above were screened using minimum inhibitory concentration (MIC) assays with liquid cultures supplemented with EDDHA. All mutants except one had a lower MIC value than a strain containing wild-type EntF (Table 5.3). We surmised that the decreased activity could be due to impaired L-serine activation or promiscuous recognition of other amino acids. To determine the exact cause, we quantified amino acid activation for 18 variants *in vitro*. The selected variants were cloned into a T7 expression vector with a C-terminal His-tag and overproduced and purified from *E. coli* BL21 (DE3).

All tested variants only activated L-serine, so the reduced *in vivo* production was due to impaired substrate activation (Figure 5.3). However, kinetic parameters did not have a strong correlation with MIC_{EDDHA} values. For example, variants 4-136 and 3-58 both activated more substrate than wild-type EntF *in vitro*, but strains expressing these variants had the same or lower MIC_{EDDHA} values. This discrepancy suggests that some

Table 5.3: Characteristics of EntF variants characterized *in vitro*. The units for V_{max}/K_m values are nmol product/min/mg protein/mM substrate. Catalytic parameters were experimentally determined by Vladimir Vinnik. The residues differing from the wild-type EntF code are highlighted in bold.

Strain	**Code**	**Residues different from wild-type**	**MIC_{EDDHA} (μM)**	V_{max}/K_m
WT	DVWHFSLVDK	N/A	>850	6312
4-136	DVWH**Y**SLVDK	1	>850	8958
3-58	DVWH**L**SLVDK	1	600	6828
4-19	D**P**WH**V**S**FI**DK	4	600	317
2.9A	DVWHFS**M**VDK	1	500	355
K16A	DV**I**HFSLVDK	1	500	480
2.16A	DVWHFS**H**VDK	1	400	789
10A	D**PV**HFSLVDK	2	400	931
26A	D**PC**HFSLVDK	2	350	672
106	DV**P**HFSLVDK	1	350	359
9F	D**PP**HFSLVDK	2	250	260
7B	D**PV**HFS**H**VDK	3	250	645
44.2	DV**C**HFS**F**VDK	2	250	399
4-16	D**PH**H**L**S**CI**DK	5	250	115
2-10	DVW**S**FSLVDK	1	250	1330
34.78	DV**T**HFS**M**VDK	2	150	749
4-176	D**A**WH**M**SL**I**DK	3	150	408
4-71	DVWH**I**SL**L**DK	2	100	394

mutations in the EntF A domain affected later steps in enterobactin biosynthesis, such as T domain acylation or condensation of 2,3-dihydroxybenzoic acid with L-serine.

Unsurprisingly, most of the highly active EntF variants only have one residue change compared to wild-type EntF, but there are some exceptions (Table 5.3). For example, variant 4-19 has four mutations but has a MIC_{EDDHA} of 600 μM, the second-highest concentration tested. Its mutations are located in positions 2, 5, 7, and 8, all positions that have been shown to be relatively variable in our libraries. The two most active variants, 4-136 and 3-58, have mutations in the fifth position of the specificity code, exemplifying the variability observed at this position in Nature and in the libraries described here.

5.4.3 Amino acid activation of variants in a non-EntF context

The discovery of unique specificity codes that function in EntF raised the question of whether other L-serine-activating A domains shared this sequence space. Searching known A domains in Genbank, we found five unique specificity codes from A domains predicted to activate L-serine that matched specificity codes from the EntF libraries.

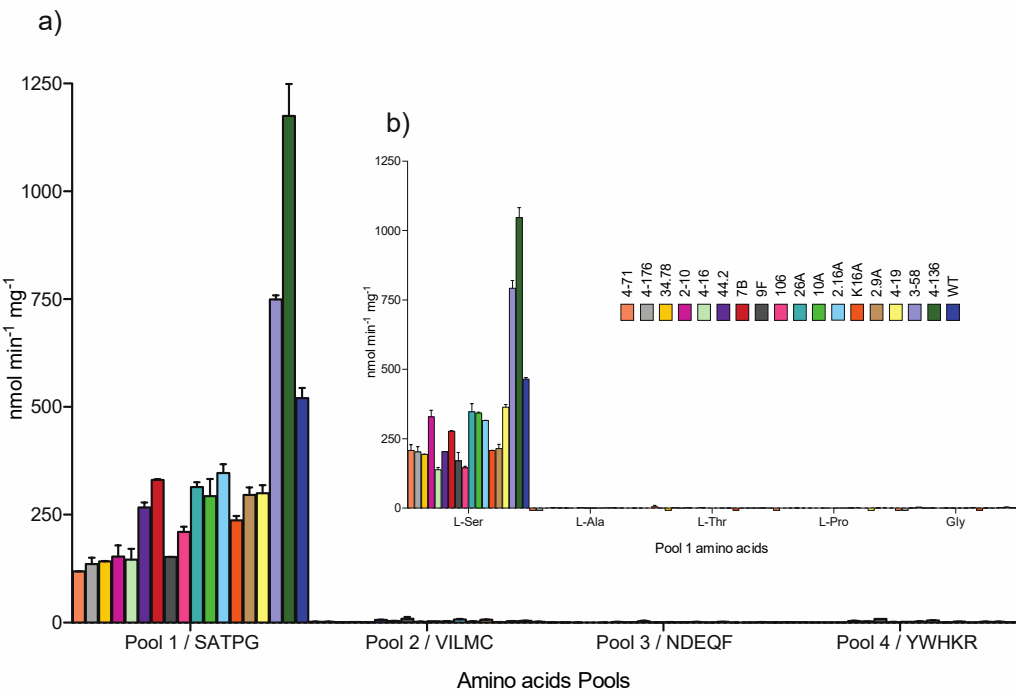


Figure 5.3: **ATP/PP$_i$ exchange assays for the select EntF variants presented in Table 5.3 with pooled (A) or individual (B) amino acid substrates.** In both panels the variants are order from left to right by increasing associated MIC$_{EDDHA}$ (as in Table 5.3) and both panels are color coded according to the key in panel B. Amino acid Pool 1: SATPG, 2: VILMC, 3: NDEQF, 4: YWHKR.

Two of these matched variants 3-58 (DVWHLSLVDK) and 4-213 (DVWHLSLIDK, not listed in Table 5.3), and several A domains with these specificity codes have been demonstrated experimentally to activate L-serine [438, 439, 440]. The three other specificity codes matching 4-136, K16A, and 4-54 (DVLHSSLVDK, not listed in Table 5.3), had not been observed in an A domain that was experimentally confirmed to activate L-serine.

To determine if an A domain with a previously uncharacterized specificity code did infact activate L-serine, we identified and characterized an A domain with a specificity code matching that of variant 4-136. The A domain in DltA from *Paenibacillus dong-haensis* is 42.4% identical to EntF overall and has the specificity code DVWHYSLVDK, one residue different from the specificity code of EntF. We purified the A-T didomain region of DltA and characterized it using ATP/PP$_i$ exchange assays. The DltA A domain is specific for L-serine, similar to EntF and the variants described here (Figure 5.4).

To further demonstrate that the sequence space revealed in our directed evolution studies is shared with other L-serine-specific A domains, we sought to characterize another

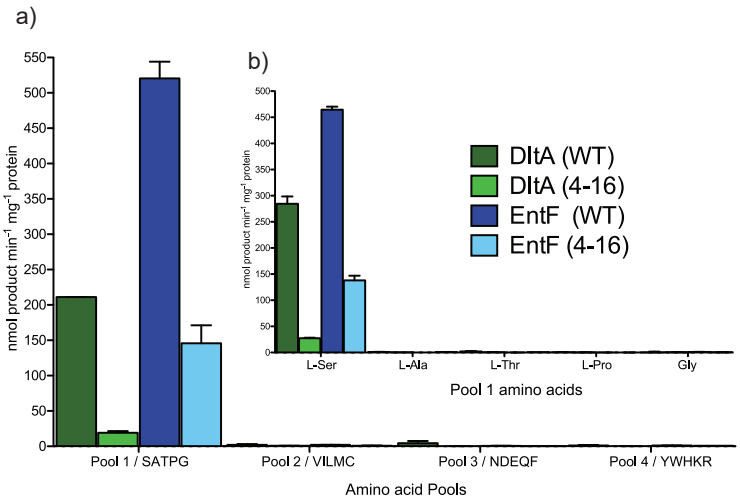


Figure 5.4: **Variant EntF specificity code, differing at all five mutagenized sites, that functions in a non-EntF context.** Comparison of the in vitro activity of purified proteins EntF wild-type (WT), EntF variant 4-16, DltA WT, and DltA variant 4-16 with pooled amino acid substrates (A) and the individual amino acids contained in pool 1 (B) as determined by the ATP/PP_i exchange assay. The color key in panel A applies to both panels: pool 1, SATPG; pool 2, VILMC; pool 3, NDEQF; pool 4, YWHKR.

specificity code in the context of DltA. Variant 4-16 still functions in EntF despite having five mutations in the active site residues. We introduced this code to DltA and determined that this variant activates only L-serine. The 4-16 code is the most distant from wild-type EntF, and its function in DltA suggests that the other variant specificity codes characterized here would also function in non-EntF A domains.

While the results so far have established that the functional sequence space of EntF is shared with other L-serine-specific A domains, we were curious if the novel specificity codes could activate L-serine in A domains that natively activate a different amino acid. The majority of specificity codes identified in this study have not been found in Nature, and efforts modifying A domain specificity through code mutations could have been limited by relying on naturally occurring sequences. We identified the glycine and L-threonine-activating A domains from DhbF as a new context to test various specificity codes (Figure 5.5). DhbF is involved in the biosynthesis of bacillibactin, a siderophore produced by *Bacillus subtilis* [441]. DhbF is a two-module NRPS, and the A domain in the first module (F1) activates glycine and the A domain in the second module (F2) activates L-threonine. F1 has 43.3% homology to the A domain of EntF and has a specificity code of DILQLGLIWK, and F2 has 40.1% homology with a code of DFWNIGMVHK.

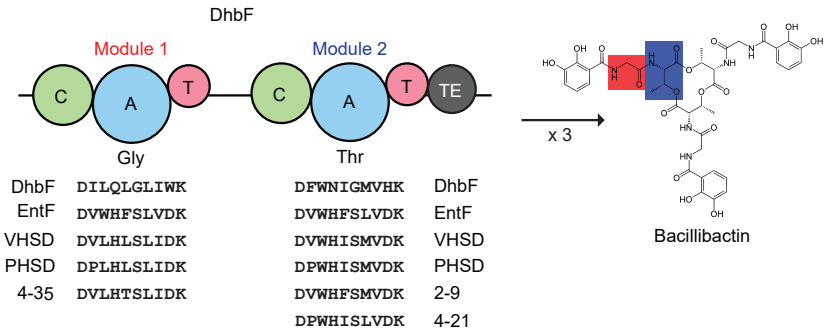


Figure 5.5: **Biosynthesis of bacillibactin through DbhF in *B. subtilis*.** Bacillibactin biosynthesis is similar to that of enterobactin in that it begins with the activation of 2,3-dihydroxybenzoic acid. Glycine and L-threonine are then incorporated via a two-module NRPS, DhbF. After three cycles, peptides are cyclized by the TE domain of DhbF to form a triketide lactone ring. The portions of the ring that contain the glycine and L-threonine residuse are highlighted in red and blue, respectively. Variants of the first A domain in DhbF were designed based on EntF variant isolated in this study.

We designed three specificity codes to characterize in F1 and four to characterize in F2, all of which contain the His4-Ser6-Asp9 residues established to be essential for L-serine activation in EntF. The first two specificity codes had the residues Val2-His4-Ser6-Asp9 (VHSD) and Pro2-His4-Ser6-Asp9 (PHSD), respectively, while the rest of the specificity codes of F1 and F2 were unchanged (Figure 5.5). A third variant of F1 based on variant 4-35 was also constructed. Two more variants of F2 had specificity codes matching variants 2-9 (DVWHFSMVDK) and 4-21 (DPWHISLVDK). Variants based on specific variants from the EntF libraries were chosen because they required the fewest number of mutations, five, to each A domain. We established conditions for characterizing the wild-type DhbF A domains, which required the presence of cognate MbtH-like protein (MLP) for detectable activity (Figure C.1). However, when we tested the F2 variants with ATP/PP_i exchange assays, activation of any amino acid was abolished (data not shown). These results highlight the bias that A domains have for their native substrate. In future experiments focused on mutagenizing A domains, changes to residues outside of the specificity code will likely be required to fully change substrate specificity.

5.4.4 Establishing a second siderophore system for directed evolution studies

Fluorescent Pseudomonads all produce pyoverdines, a class of fluorescent siderophores that have a yellow-green color [442]. Species that fall under this group include *Pseudomonas aeruginosa*, *P. fluorescens*, *P. syringae*, and *P. putida*. These bacteria synthesize pyoverdine through a large NRPS encoded by four or more genes. The pyoverdine molecule synthesized by each species has a unique structure due to the evolution of these NRPSs. In fact, it is possible to identify *Pseudomonas* isolates based on the structure of their native pyoverdine, a process known as "siderotyping" [443]. Pyoverdines are efficient at binding iron(III) and are the primary route that fluorescent Pseudomonads use to acquire iron in deficient conditions. Like *E. coli* and enterobactin, pyoverdine production is required for growth on iron-deficient media, such as media containing a strong iron-chelating compound.

The *P. putida* genome encodes four NRPS genes responsible for synthesizing the peptide backbone for its native pyoverdine [73]. Synthesis begins with the NRPS, PvdL, which is heavily conserved in all pyoverdine-producing strains (Figure 5.6a). PvdL catalyzes the condensation of myristyl-CoA to glutamic acid, followed by incorporation of L-tyrosine and L-2,4-diaminobutyric acid (Dab). The remainder of the NRPS is encoded by three genes, *pvdI*, *pvdJ*, and *pvdD*, which incorporate eight more amino acids. The peptide product is called acyl-ferribactin, which is transported into the periplasm for maturation into pyoverdine. De-acylation of ferribactin by PvdQ is the first periplasmic step [444]. The first three residues of the ferribactin peptide, Glu-Tyr-Dab, are then converted into the aromatic chromophore that is characteristic of all pyoverdines (Figure 5.6a) [445]. Periplasmic efflux systems, such as PvdRT-OpmQ and MdtABC-OpmB, then transport the mature pyoverdine into the extracellular space [446].

The final peptide in the pyoverdine NRPS, PvdD, has three full modules that activate glycine, L-serine, and N(5)-hydroxy-L-ornithine. To complement our experiments with EntF, we intended to explore various A domain engineering strategies with the L-serine-activating A domain of PvdD. Initially, we needed to construct a functional *pvdD*

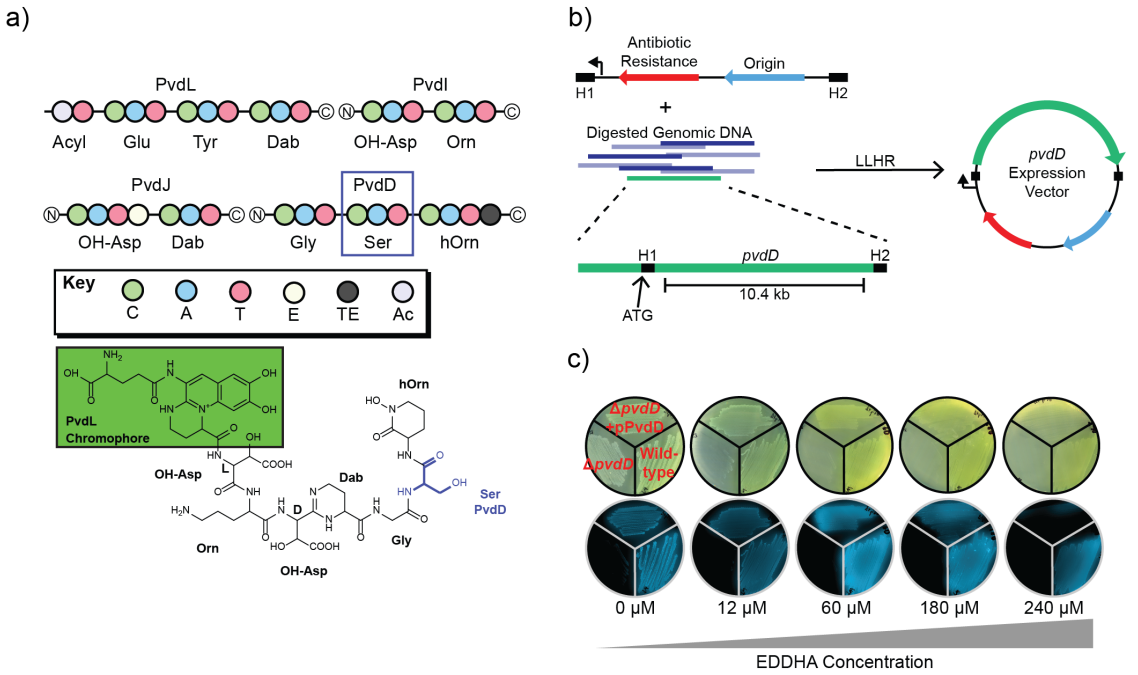


Figure 5.6: **Pyoverdine biosynthesis in *P. putida* and complementation of a *pvdD* knockout a)** Four proteins make up the pyoverdine NRPS in *P. putida*. Eleven amino acids are incorporated into the peptide backbone, six of which are non-proteinogenic. **b)** RecET direct cloning was used to clone *pvdD* directly from *P. putida* genomic DNA. The final expression vector contained the IPTG-inducible P_{trc} and the RK2 origin. **c)** The *pvdD* expression vector complements a *pvdD* knockout strain. Fluorescence and growth in the presence of EDDHA is restored. Top row was imaged in visible light and the bottom row was imaged in UV light.

expression vector that complemented a knockout strain. We successfully deleted the 10.4-kb region containing *pvdD* from the *P. putida* chromosome using a counterselection based on 5-fluorouracil [158]. We then sought to clone *pvdD* into a broad-host range vector capable of replicating in *P. putida*. Expression vectors were constructed from RecET direct cloning (Figure 5.6b), a protocol that enables the cloning of large genes and gene clusters directly from genomic DNA [84]. *PvdD* was originally cloned into a vector with the BBR1-UP origin, a mutant origin selected to have a higher copy number [329, 176]. However, complementation with this vector was never achieved as this construct was unstable in *P. putida*. We hypothesized that plasmid instability was due to high copy number, so we cloned *pvdD* into another vector with the RK2 origin, resulting in the plasmid pRK2T-Ptrc-PP_4219-gen. This plasmid was able to complement the *pvdD* knockout when induced with IPTG (Figure 5.6c). *P. putida* Δ*pvdD* was unable to grow in as little as 12 μM EDDHA, and introducing the complementation vector restored growth in as much as 180 μM EDDHA.

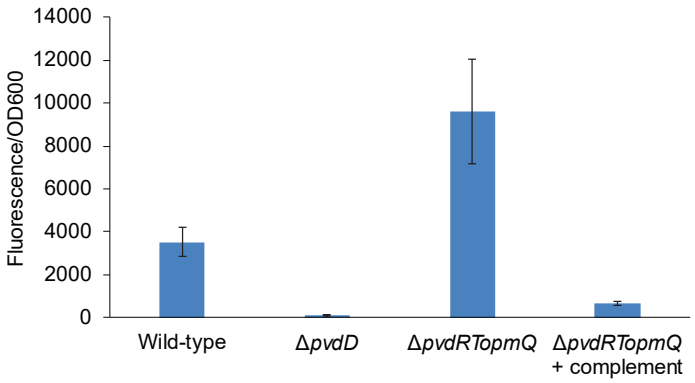


Figure 5.7: **Role of PvdRTOpmQ in secreting pyoverdine from *P. putida* KT2440.** *P. putida* with deletion of *pvdRTopmQ* operon accumulated more pyoverdine in cells than wild-type cells. Pyoverdine production decreased when *pvdRTopmQ* was overexpressed on a complementation vector. Cells were grown in King's B broth supplemented with Gent30 overnight and resuspended in PBS. The fluorescence intensity and OD600 of the cell suspension were determined in a M1000 plate reader.

It was previously established that the efflux pump PvdRTOpmQ is responsible for secreting pyoverdine from *P. aeruginosa* [447]. The *P. putida* genome encodes a homologous efflux pump, so we hypothesized that constructing a deletion strain for this secretion complex would increase the amount of pyoverdine accumulation in *P. putida* cells, potentially enabling fluorescence-based screens for pyoverdine production. We deleted the *pvdRTopmQ* operon encoding this efflux pump and compared pyoverdine production of this strain to that of wild-type *P. putida*. After production cultures were washed one time with PBS, cell suspensions of *P. putida* Δ*pvdRTopmQ* had greater fluorescence intensity due to pyoverdine accumulation (Figure 5.7). Interestingly, overexpressing *pvdRTopmQ* in the deletion strain reduced pyoverdine production overall. These results demonstrated that PvdRTOpmQ is responsible for pyoverdine secretion from *P. putida*, but there was still much more pyoverdine in the supernatant from all strains analyzed, indicating that another pathway for pyoverdine secretion exists, potentially a homologue for MdtABC-OpmB [446].

Complementing pyoverdine production established a system for investigating NRPS enzymology using a growth-based selection. There is considerable diversity in pyoverdines downstream of the chromophore residues among different species, so there is a wealth of phylogenetically-related NRPS modules that could serve as donors for domain substitutions. At the time of this work, the most successful attempt at generating pyoverdine

derivatives through module swaps used C-A didomain swaps to change the amino acid specificity of the second module in PvdD from *P. aeruginosa* [121]. A-domain swaps using the same donor enzymes were unsuccessful unless the donor A domain had the same specificity as the cognate A domain. One possible explanation for these results is that the C domains in pyoverdine NRPSs are specific for the donor amino acid. Alternatively, the failed A domain swaps could be due to disrupted interactions between the C and A domains.

An alternative strategy to A domain swaps that avoids issues with C and A domain interactions is subdomain swaps, which is based on the knowledge that the specificity-conferring region of A domains is a compact subdomain surrounding the active site. Residues near the N- and C-termini of the A domain are not mutated when performing these substitutions, so theoretically they should be less likely to disrupt interactions of the A domain with the C and T domains. When we began this work, chimeric enzymes from subdomain swaps had only been characterized *in vitro* [78]. We were interested in whether subdomain swaps could be a more successful strategy for synthesizing pyoverdine derivatives, so we set out to generate chimeric variants of PvdD focused on the first two A domains, which are specific for glycine and L-serine, respectively. However, the length of *pvdD* considerably increased the difficulty in constructing chimeras through standard cloning procedures, so we had to explore several cloning methods to establish a cloning pipeline.

5.4.5 Cloning pipeline for constructing module and domain swaps in assembly-line enzymes

Many metabolic engineering projects, especially in bacterial hosts, require cloning one or a few genes into an expression vector that is usually no more than 10 kb in size. Cloning mutants of large genes, such as type I PKS and NRPS genes, increases DNA construct size, therefore decreasing the efficiency of several cloning steps. In addition, multiple cloning techniques are sequence-dependent, which limits their utility when one desires to make modifications to DNA sequences that fall within a CDS, e.g. domain

substitutions. The prevalence of repetitive DNA sequences within modular assembly-line genes also contributes to these technical challenges [448].

Multiple cloning methods involve *in vitro* DNA assembly, including traditional restriction and ligation cloning and sequence-independent methods like Gibson assembly [382]. Once the assembly is complete, these reactions are transformed into competent *E. coli* cells. The efficiency of this step (i.e. the number of transformants per unit mass of DNA) is much lower for large plasmids or plasmids with low-copy number origins. An alternative method that bypasses this limitation uses homologous recombination to modify large plasmids, specifically linear-circular homologous recombination (LCHR) [449]. DNA modifications in this protocol occur *in vivo*, where the target plasmid is introduced to the cloning strain beforehand, and a linear DNA construct required for modification is introduced via electroporation and recombines with the circular plasmid (Figure 5.8). This modification requires a selectable marker, often for antibiotic resistance, as well as a counterselection gene for future cloning steps. Using this method, we constructed mutants of the *pvdD* expression vector with a tetracycline resistance (TcR) cassette in place of the subdomain of the first and second A domains in *pvdD* (pRK2T-Ptrc-PP_4219shell1-TcR-gen and pRK2T-Ptrc-PP_4219shell2-TcR-gen). These two plasmids served as the "capture" vectors for various cloning protocols in order to insert alternative subdomains from *E. coli*, *B. subtilis*, and *P. fluorescens*. We investigated three different cloning methods: type IIS restriction and ligation cloning, programmable argonaute restriction and ligation cloning, and Cas9-assisted homologous recombination.

The TcR cassette was flanked by recognition sites for BsaI, a type IIS restriction enzyme used commonly in Golden Gate assembly. These sequences were designed so that restriction digestion cleaves the TcR region from the capture vector, resulting in a linearized vector with sticky ends that are homologous to the 5' and 3' ends of the wild-type subdomain. PCR products containing complementary BsaI sequences were designed so that they could be ligated with the linearized capture vector, resulting in a chimeric *pvdD* expression vector. We constructed several variant plasmids with a low incidence of undesired plasmids, but the transformation efficiency was low due to the size of the

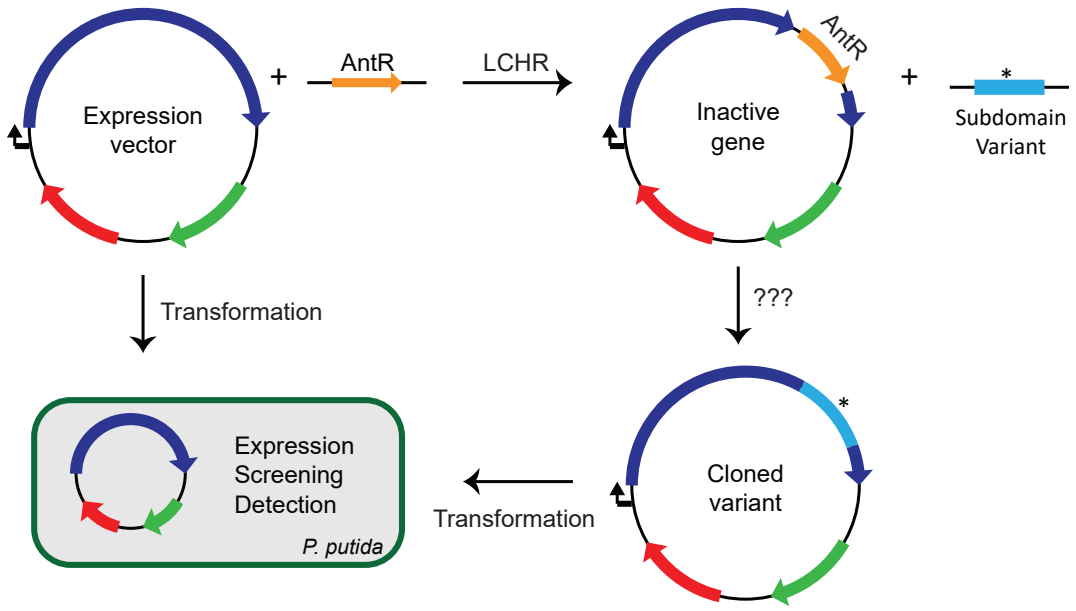


Figure 5.8: **Cloning NRPS and PKS variants requires a specialized workflow with multiple steps.** A typical expression vector containing a NRPS or PKS gene is over 10 kb. A "stuffer" is inserted into the target region of the CDS via linear plus circular homologous recombination (LCHR). The stuffer region contains an antibiotic resistance (AntR) marker and/or a counterselection marker. These markers may also be flanked by sequences that enable a second cloning step where the stuffer is replaced with a variant of interest. Variant constructs are then introduced into the expression host for analysis.

plasmids. In addition to constructing domain substitutions, we were also interested in generating libraries of the *pvdD* expression vector, so we explored alternative cloning methods to be able to generate larger libraries than what was possible with restriction and ligation cloning.

We explored using a thermostable argonaute as a programmable restriction enzyme to avoid the initial insertion of the stuffer region via LCHR. An argonaute from *Pyrococcus furiosus* (*Pf*Ago) was characterized previously by the Zhao group to develop a programmable restriction enzyme that cuts any DNA sequence of interest [81]. *Pf*Ago was able to digest plasmid DNA using 5'-phosphorylated DNA oligomers that were complementary to the target sequence. We intended to use this enzyme to linearize the *pvdD* expression vector directly with oligomers that targeted the 5' and 3' ends of the subdomains. This would allow module and domain substitutions in a single cloning step. We purified *Pf*Ago via affinity chromatography (Strep tag) and demonstrated endonuclease activity using 5'-phosphorylated oligos (Figure C.2). To test its ability to generate constructs for restriction and ligation cloning, we designed oligos to linearize an expression

vector and attempted to ligate a RFP cassette downstream of the promoter. We were able to construct this vector using plasmid DNA that was linearized with *Pf*Ago, but the cloning efficiency was incredibly low (data not shown).

Lastly, we sought to expand on the LCHR method with a Cas9-based counterselection. This method was conceptually the same as Cas9-assisted genome editing methods, except the target DNA was a plasmid replicating in *E. coli*. The cloning strain also contains a plasmid expressing Cas9 and the λRed genes to facilitate homologous recombination (Figure 5.9a). The major difference between this protocol and LCHR is that the repair template is a circular plasmid instead of a linear PCR product. The donor plasmid contains the gene/domain/construct to be inserted along with homology arms flanking the target region. To apply the Cas9 counterselection, the donor plasmid also contains a single guide RNA CDS targeting the region to be replaced. This protocol was designed to have the following advantages to the methods mentioned above: the Cas9 counterselection enabled modifications to large plasmids in a single step (LCHR beforehand is not necessary), double-stranded breaks caused by Cas nucleases increases recombination efficiency with the λRed enzymes (Figure C.3), and the transformation efficiency of the high-copy donor plasmid is higher than that of the large target plasmid with a low-copy origin.

To demonstrate the feasibility of the cloning protocol, we designed a high-throughput cloning screen based on expression of RFP, which is visible when expressed in *E. coli*. The target plasmids in these experiments contained the 5' and 3' ends of the RFP gene downstream a constitutive promoter and had either the *E. coli*-specific p15A origin or the broad-host range RK2 origin. The donor/sgRNA plasmid contained the remainder of the RFP gene plus homology arms to the target plasmid. These constructs were designed so that transformants in which both plasmids hybridized would form red colonies on solid media. For both target plasmid origins, we found that transformation of the donor plasmid led to at least 80% of transformants appearing red (Figure 5.9b). These results were encouraging because this recombination efficiency was high enough for constructing plasmid libraries.

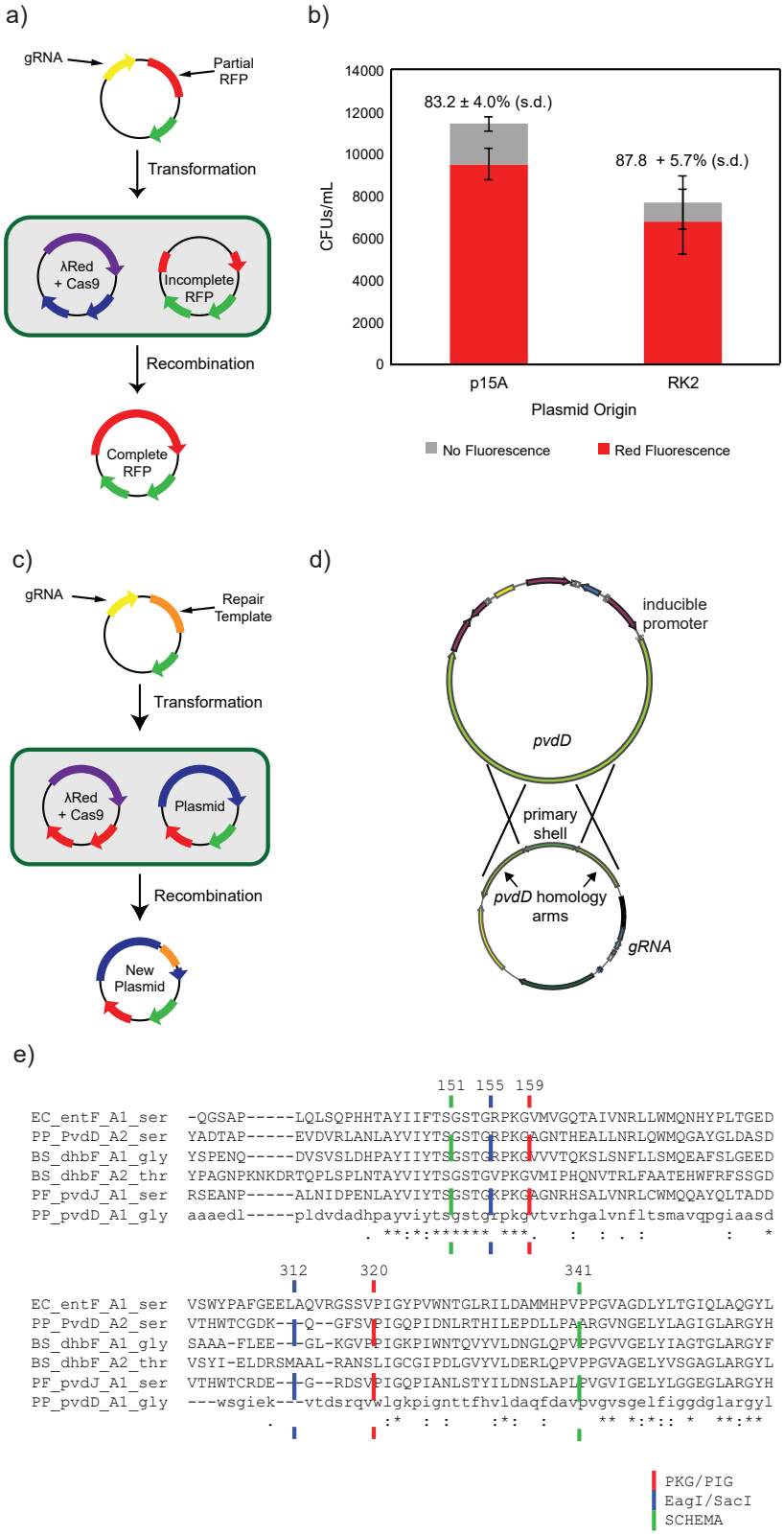


Figure 5.9: **Cas9-assisted homologous recombination of plasmid DNA *in vivo*** **a)** Plasmids used to clone RFP expression vector. The target plasmid contained 5' and 3' ends of RFP gene, and the donor plasmid contained the remaind of RFP with 200-bp homology arms. Either plasmid alone does not express full RFP, so fluorescence indicates recombination of the two plasmids. **b)** Cas9-assisted recombination efficiency based on transformation efficiency and rate of "red" colonies. Percentage values indicate the percentage of colonies that appeared red. Error bars represent standard deviation from three independent transformations. **c)** A general schematic for Cas9-assisted plasmid recombination. **d)** Detailed view of plasmids used to insert alternative subdomains into *pvdD*. The donor plasmid contains a guide RNA targeting a *pvdD* subdomain and an alternative subdomain with 500-bp homology arms. **e)** The 5' and 3' boundaries for subdomain substitutions in *pvdD*. The PKG/PIG boundaries are chosen manually based on location of PKG and PIG amino acid motifs in various A domains. The EagI/SacI boundaries are the same boundaries used when cloning the EntF libraries. For comparison, these A domains were analyzed by SCHEMA [450], an algorithm for determining optimal boundaries for chimeragenesis.

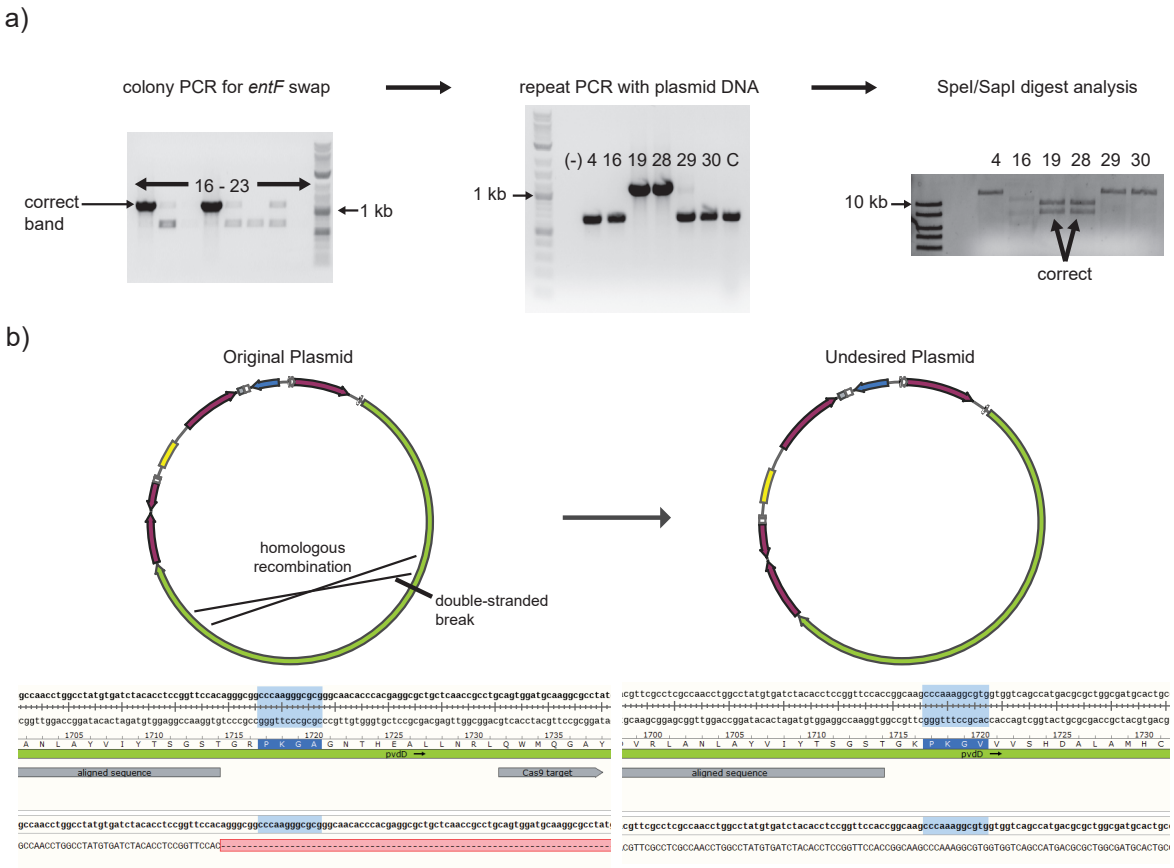


Figure 5.10: **Mode of failure for Cas9-assisted plasmid recombination. a)** After screening via colony PCR, not all hits have desired plasmid DNA. An alternative recombinant becomes the predominant plasmid in the overnight culture and therefore the purified plasmid prep. **b)** Undesired recombinant arises from intramolecular recombination that results in a module deletion. The site of recombination is the PKG motif (highlighted in Snapgene screenshots) on the N-terminus of the A subdomain targeted by Cas9.

After demonstrating cloning of an RFP construct, we then designed constructs to substitute various subdomains into the first and second A domains of PvdD (Figure 5.9c,d). We chose A domains from *E. coli*, *B. subtilis*, and *P. fluorescens* that activated glycine or L-serine, and we initially planned substitutions that did not change A domain specificity in PvdD. We tested two different subdomain boundaries based on the location of PKG and PIG motifs on the N- and C-termini of the subdomains or based on the location of the EagI and SacI sites used to generate EntF libraries, as described earlier in this chapter (Figure 5.9e).

When we attempted to construct these plasmids, we initially screened transformants by colony PCR, which resulted in several hits for most constructs (Figure 5.10a). However, when plasmid DNA was extracted from overnight cultures, most of the samples

did not have the desired plasmid. Digest analyses and sequencing revealed that these recombinants resulted from intramolecular recombination of the target plasmid (Figure 5.10b). In rare cases, the purified plasmid prep retained the colony PCR product but did not have the expected band pattern in a digest analysis, indicating that the desired plasmid was present in a small molar fraction compared to the undesired plasmid. In these cases, re-transforming the plasmid DNA into competent *E. coli* followed by screening transformants allowed us to isolate pure DNA preps of the desired plasmid.

The intramolecular recombination occurred between the sections encoding the PKG motifs near the N-termini of the subdomains in the second and third A domain in PvdD. This recombination functionally deleted the L-serine-activating portions of PvdD. This recombination occurred in several transformants, suggesting that this location is a recombination hot-spot within the pyoverdine NRPS genes and could have played a role in pyoverdine evolution in Pseudomonads. Another mode of failure not shown in Figure 5.10 is that some transformants contained plasmid hybrids of the target and donor plasmids, indicating that the recombination had not completely resolved. This could not be detected with the colony PCR screen and also reveals a major limitation for this protocol. The recombination efficiencies for the RFP constructs in Figure 5.9b are likely much lower than the reported values because the RFP phenotype is still possible with the unresolved target/donor plasmid hybrid.

These technical issues with plasmid recombination arise from the multi-copy nature of plasmids. Having multiple copies of the target DNA increases the likelihood that an unwanted recombination event occurs. This problem could be fixed by replacing the plasmid origins with a single-copy bacterial artificial chromosome (BAC), which is commonly used to clone large gene clusters. In fact, a protocol for modifying NRPS and PKS genes via oligo recombineering was developed recently, and the authors reported making edits to BACs with this method [308, 89].

5.4.6 Pyoverdine production with PvdD chimeras

Despite the reported technical issues with cloning *pvdD* variants, we were able to clone several subdomain substitutions in the first and second A domain of PvdD. Each substitution was designed with the PKG/PIG or the EagI/SacI boundaries. Two chimeras had the glycine-specific subdomain from DhbF (*B. subtilis*) inserted into the first A domain of PvdD. The other chimeras had the L-serine-specific subdomains from EntF (*E. coli*) or from PvdJ (*P. fluorescens*) inserted into the second A domain of PvdD. None of the constructs were intended to change A-domain specificity because we were primarily interested in comparing the different boundaries of the chimeras.

To test the PvdD chimeras, we introduced the expression vectors into *P. putida* Δ*pvdD* and grew the transformants in liquid cultures containing King's B broth and 60 μM EDDHA. The only two constructs that conferred growth were the L-serine-specific subdomains with the PKG/PIG boundaries (Figure 5.11a). While the growth defect compared to wild-type was not significant, these chimeras produced significantly less pyoverdine based on fluorescence measurements (Figure 5.11b). This indicates that growth selection can be more sensitive for detecting pyoverdine production than by fluorescence. Wild-type *P. putida* produced significantly more pyoverdine than was necessary for growth in these media conditions. None of the chimeras with the EagI/SacI boundaries or the glycine-specific chimeras produced detectable amounts of pyoverdine. It is unsurprising that the EagI/SacI chimeras functioned poorly because the SacI location appears in a region with little homology and a significant amount of gaps in the sequence alignment for these A domains (Figure 5.9e). Interestingly, the PF_pvdJA1-PKG/PIG and PF_pvdJA1-EagI/SacI chimeras only differ in three amino acids (**GRD**SVPIG vs. **QGF**SVPIG, respectively), indicating that these residues are important for subdomain function. The performance of the EntF-PKG/PIG chimera suggests that subdomain swaps can be functional with phylogenetically distant A domains, but more variants need to be investigated to confirm this. A recent publication reported little to no activity from chimeras designed by subdomain swaps, but the constructs did not use the same combination of the PKG/PIG boundaries described here [126].

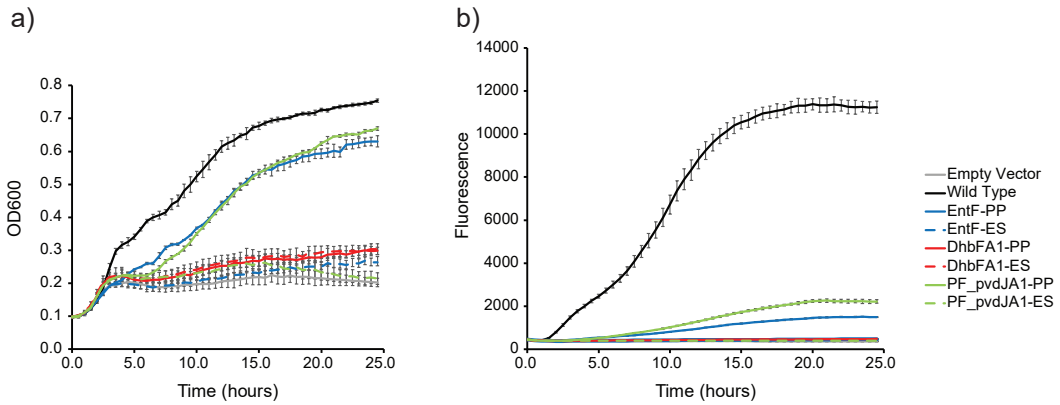


Figure 5.11: **Complementing *pvdD* knockout with chimeric enzymes a)** *P. putida* growth over 24 hours in King's B broth with 60 μM EDDHA. Growth was monitored by OD600. **b)** Pyoverdine production over 24 hours in same conditions. Fluorescence was monitored with an excitation wavelength of 398 nm and an emission wavelength of 455 nm. Growth media was supplemented with Gent30. Cultures were grown in a 96-well plate maintained at 30°C in a plate reader with shaking. EntF - EntF subdomain in second PvdD A domain; DhbFA1 - glycine-specific subdomain from DhbF in first PvdD A domain; PF_pvdJA1 - L-serine-specific subdomain from *P. fluorescens* PvdJ in second PvdD A domain; PP - PKG/PIG boundaries; ES - EagI/SacI boundaries. The legend applies to both panels.

To learn more about how these chimeras differed from wild-type PvdD, we designed expression vectors for overproducing the L-serine-specific A domain of PvdD and the EntF-PKG/PIG chimeric A domain in *E. coli*. Both A domains did not appear to be soluble based on SDS-PAGE of protein extracts from production cultures (Figure C.1b,c). Co-expressing the MLP from *P. aeruginosa* (PA_2412) did not improve solubility in the protein extracts, but it did increase the amount of protein detected in the inclusion body fractions, especially for the EntF chimera, which was not detected at all unless MLP was co-expressed (Figure C.1c). This indicates that co-expressing MLP improved A domain expression, and this was true for the wild-type A domain as well as the EntF chimera, demonstrating that subdomain substitutions do not interfere with A domain-MLP interactions. These results are consistent with structures of A domains bound to their cognate MLP showing that the MLP-binding site is not near the subdomain [451]. Module or full A domain substitutions require knowledge of MLP-dependency of the donor A domain, so these results reveal an advantage to the subdomain swap strategy in that the cognate MLP of the donor subdomain does not need to be co-expressed for chimeras to function.

5.4.7 Heme peroxidase activity in fabrubactin biosynthesis

The third and final siderophore system that we investigated is fabrubactin from *A. fabrum* strain C58. Fabrubactin is synthesized by PKS and NRPS components, and when it was first discovered, the authors noticed that the biosynthetic enzymes are more closely related to pathways in filamentous cyanobacteria rather than other known enzymes [452]. Fabrubactins produced by *A. fabrum* strain C58 have an incredibly unique structure, including a rare 1,1-dimethyl-3-amino-1,2,3,4-tetrahydro-7,8-dihydroxy-quinolin (Dmaq) moiety that was only previously seen in anachelin-type siderophores from cyanobacteria [453].

The Dmaq moiety is proposed to be synthesized originally from levodopa (L-DOPA). L-DOPA is incorporated into the fabrubactin backbone via an NRPS, FbnH. The L-DOPA portion of the nascent polyketide/peptide is converted to Dmaq via several enzymatic reactions. We initially proposed that L-DOPA is synthesized from L-tyrosine by FbnM because it had weak homology to SfmD, a heme peroxidase that was shown to hydroxylate 3-methyltyrosine to synthesize 3-hydroxyl-5-methyltyrosine [454]. FbnM also contains a HxxxC motif that is crucial for heme binding, further supporting its status as a heme peroxidase. Attempts to overproduce and purify FbnM were initially unsuccessful. FbnM only showed homology to the C-terminal side of SfmD, so we hypothesized that the enzyme encoded upstream, FbnL, formed a complex with FbnM that is necessary for enzyme solubility and function. We found that both enzymes were soluble and co-purified on a Ni-NTA column when FbnL was cloned with an N-terminal His-tag.

E. coli production cultures containing both FbnL and FbnM produced a brown pigment, which is indicative of the production of melanin-like compounds and can occur from the oxidation and polymerization of L-DOPA [455]. To generate more direct evidence of L-DOPA synthesis in these cultures, we utilized a colorimetric assay based on an L-DOPA dioxygenase from *Mirabilis jalapa* (*Mj*DOD) [456]. *Mj*DOD catalyzes the conversion of L-DOPA to betalamic acid, which spontaneously reacts with various amino acids to form the yellow pigments, betaxanthins. To determine if *E. coli* cultures expressing FbnLM were secreting L-DOPA, we collected supernatants from expression cultures,

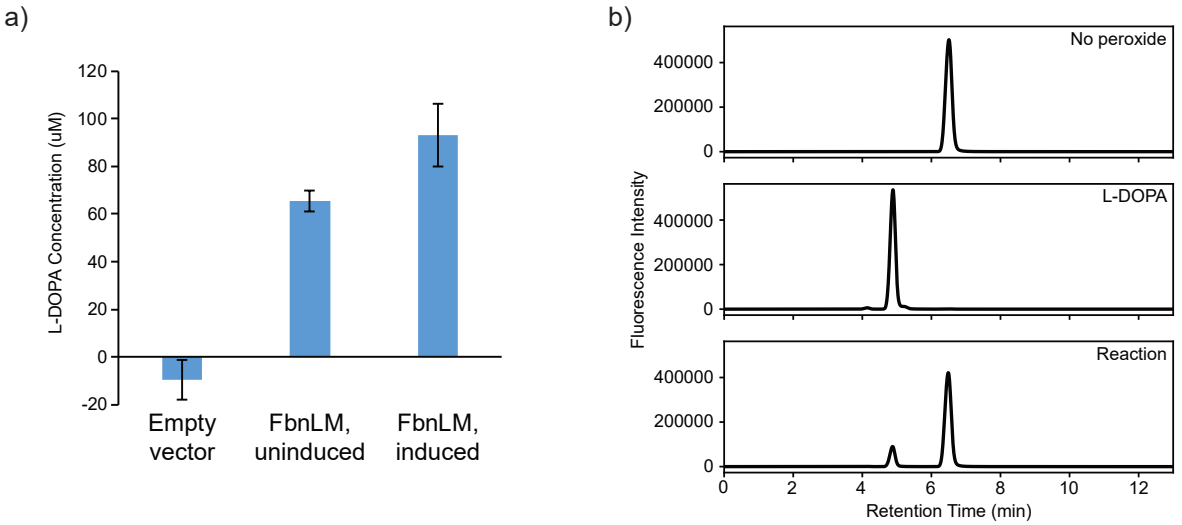


Figure 5.12: **FbnLM is a heme peroxidase that converts L-tyrosine to L-DOPA a)** Supernatant from *E. coli* cultures co-expressing FbnLM have an increased fluorescent signal in *Mj*DOD assay. L-DOPA concentration was calculated based on a standard curve of L-DOPA added to supernatant from *E. coli* with an empty vector control. **b)** FbnLM catalyzes conversion of L-tyrosine to L-DOPA *in vitro*. Hydrogen peroxide is required for enzymatic activity, and the total turnover is relatively low.

introduced them to *Mj*DOD reactions, and monitored them for an increase in yellow pigmentation (Figure 5.12a). Supernatants from *E. coli* cultures expressing FbnLM had an increase in fluorescence, with and without induction. The specificity of *Mj*DOD for L-DOPA demonstrates that the accumulation of brown pigment is due to the cells exporting L-DOPA.

Purified FnbLM was also able to catalyze L-DOPA formation from L-tyrosine *in vitro*, albeit with a low turnover (Figure 5.12b). Initial formulations of the reaction buffer for these reactions contained only 1-2 mM H_2O_2 and did not show L-DOPA formation, but L-DOPA was detected by HPLC from reactions containing 20 mM H_2O_2. Taken together, these results confirm that FbnLM is responsible for L-DOPA formation in fabrubactin biosynthesis. The elucidation of this pathway will enable future directed evolution studies on its biosynthetic enzymes, particularly the PKS components.

5.5 Conclusions

The work discussed in this chapter demonstrates the many challenges of investigating and engineering secondary metabolism. In the case of fabrubactin biosynthesis, we char-

acterized enzymes with few known homologues, and biochemical evidence is still needed to elucidate other steps in this pathway. Efforts on establishing the pyoverdine system as a platform for directed evolution highlighted the technical challenges of constructing libraries of modular enzymes. Lastly, the deep investigation into the active site residues of A domains, based on EntF, reinforced that there are many aspects to substrate specificity in NRPSs that are not fully understood.

The most striking feature of variants from the EntF libraries is that they are all still specific for L-serine. Other attempts to mutagenize the specificity code in A domains usually results in promiscuous variants [20], although these results could be an artefact of how the targeted A domains have evolved. The TE domain of EntF does not accept other amino acids, so there is selective pressure against the A domain of EntF accumulating mutations that broaden its substrate specificity. In contrast, a class of nonribosomal peptides from cyanobacteria, the microcystins, exhibit several multi-specific A domains, and bioinformatic evidence suggests that there has been positive selection for the synthesis of new microcystin derivatives [457]. A domains from different genera or biosynthetic contexts will likely have different propensities for promiscuous activity, which would significantly affect attempts to alter their substrate specificity. Similar directed evolution studies with promiscuous A domains may reveal mechanisms for how A domains obtain substrate specificity, and future work will likely need to interrogate the residues that are not part of the specificity code. However, the heavy bias of some A domains for their native substrate suggests that substitution-based engineering strategies will be more successful in the short term for generating NRPS variants.

Subdomain swaps remain a viable strategy for engineering the pyoverdine NRPS domains, although a considerable amount of work is needed to determine optimal substitution boundaries. The PKG motif appears to be the optimal boundary for the N-terminus of the subdomain, supported by SCHEMA calculations and inadvertently by the intramolecular recombination of *pvdD* near this sequence motif. Unintended recombinations have been reported previously for PKSs, and the authors took advantage of this phenomenon to generate polyketide derivatives [99, 458]. These strategies could be

leveraged towards engineering the pyoverdine NRPSs, perhaps with a Cas9-based counter-selection to accelerate laboratory evolution. Alternatively, a recent publication elucidated the optimal boundaries for constructing A domain swaps in the pyoverdine NRPS from *P. aeruginosa*, and reported little to no activity from chimeras designed by subdomain swaps [126]. Phylogenetic analysis suggested that these domain boundaries are relevant in non-*Pseudomonas* NRPSs, so this strategy will likely be more feasible for generating NRPS chimeras compared to subdomain swaps and may enable new directed evolution studies in these siderophore systems (see Chapter 8).

5.6 Acknowledgments

This work was supported in part by the National Science Foundation (1716594 to M.G.T. and B.P.), the National Institutes of Health (GM100346 and AI065850 to M.G.T.), and the Great Lakes Bioenergy Research Center, U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research (DOE BER Office of Sciences DESC0018409). V.V. was supported in part by the Jerome J. Stefaniak Predoctoral Fellowship. K.T. was supported in part by funds supplied by the E. B. Fred Professorship (M.G.T.). T.C. is the recipient of a National Institutes of Health Biotechnology Training Program (NIGMS 5 T32 GM08349). S.A.L. is the recipient of a fellowship from the Promega Corporation through the Dane County Youth Apprenticeship Program in Biotechnology. The authors thank C. Ané for statistical consultation. We thank P. Romero and M. Politz (UW-Madison) for discussions and conditions for the colorimetric L-DOPA assay.

Chapter 6

Production of a heterologous extracellular enzyme in Pseudomonads

Authors and Contributors:

- **Taylor B. Cook** Designed and performed experiments, wrote the manuscript.
- **Dr. Brian F. Pfleger** Provided guidance, edited, and reviewed the manuscript.

6.1 Abstract

This chapter describes our attempt to heterologously express and secrete an enzyme in a *Pseudomonas* host. Native secretion of this enzyme requires a chaperone that is bound to the inner membrane. The chaperone is often co-expressed with the enzyme in the same operon and facilitates proper folding and transport across the inner membrane of the enzyme via the Sec pathway. The enzyme is then secreted to the extracellular region via the Xcp secretion machinery. There are several *Pseudomonas* strains that have homologues of the Sec and Xcp pathways. At the onset of the project, we proposed that a non-pathogenic *Pseudomonas* host would be able to express and secrete this enzyme via the Sec and Xcp pathways similarly to the native host.

Before introducing an enzyme-chaperone operon in a heterologous host, we characterized different combinations of promoters and origins of replication and tested heterologous expression using a fluorescent reporter. We achieved the highest heterologous reporter production using the *trc* promoter on a plasmid with the RK2 origin in *Pseudomonas putida*. We used this promoter and origin combination on all subsequent experiments involving extracellular enzyme production.

We then investigated heterologous production and secretion of the enzyme in 7 heterologous hosts:

- *Pseudomonas putida* KT2440
- *Pseudomonas putida* EM383 - genome minimized derivative of *P. putida* KT2440
- Four *Pseudomonas fluorescens* strains obtained from NRRL
- *Pseudomonas acidophila* ATCC 31363

We only detected expression of soluble enzyme in *P. putida* (KT2440 and EM383) and *P. fluorescens* NRRL B-2550. In the cases of *P. putid*a EM383 and *P. fluorescens* B-2550, we determined that the enzyme primarily accumulated in the periplasm of the heterologous host and there was negligible secretion of the enzyme into the media.

6.2 Introduction

Extracellular enzymes are often important in biotechnology and in drug development. Several groups of bacteria, such as *Pseudomonas* species, produce and secrete enzymes into the extracellular region. Native enzyme producers are sometimes pathogenic, which can hinder production of these enzymes by requiring additional safety precautions in industrial processes. A potential strategy to circumvent the safety concerns of some native hosts is to produce enzymes of interest in heterologous Pseudomonads.

There are many bacteria that have the potential to secrete enzymes and other proteins heterologously, particularly in the genus *Pseudomonas*. *P. putida* KT2440 is a well-studied soil bacterium and has seen significant development as a synthetic biology and industrial chassis strain in recent years [168]. Several strains of *P. fluorescens* have been investigated for their ability to secrete enzymes, and some strains have been engineered for producing heterologous proteins [459]. Many genetic tools have been characterized for use in *P. putida* and other Pseudomonads, making them attractive hosts for expressing heterologous proteins [460, 176, 461]. While expressing recombinant proteins in these hosts has been made trivial through these developments, extracellular enzymes have an added challenge in requiring other host factors for secretion.

Enzymes that are natively secreted are often co-expressed with specific chaperones that are required for successful enzyme folding and secretion. For transport across the membrane, an enzyme must interact with this chaperone as well as secretory pathways native to the host. The enzyme is initially expressed with a peptide leader that serves as a signal sequence for the Sec translocase (Figure 6.1). The signal peptide is inserted into the inner membrane, and both the chaperone and the Sec machinery facilitate enzyme transport into the periplasm and cleaves the signal peptide from the mature enzyme. The chaperone and oxidoreductases in the periplasm promote enzyme folding during or after this step. The mature enzyme is then transported across the outer membrane via the Xcp pathway (also referred to as the Type II secretion system).

The host-specific pathways in this process, are conserved to varying degrees between bacterial hosts [462]. Interactions between these secretion pathways and extracellular

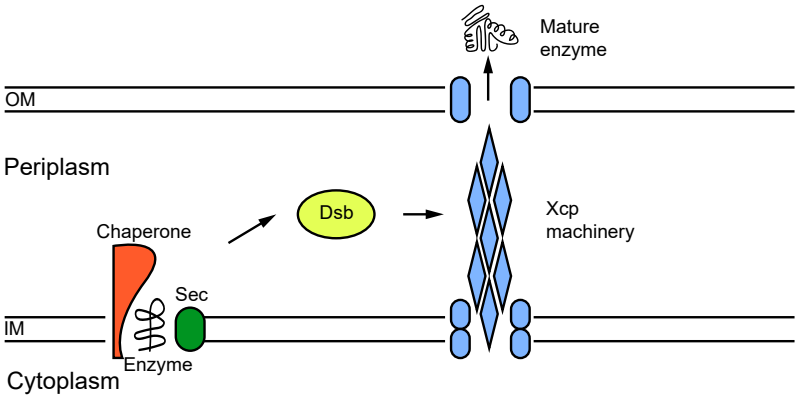


Figure 6.1: **Secretion pathways responsible for transport of extracellular enzymes in bacterial hosts.** Transport across the inner membrane (IM) of enzyme is facilitated by a chaperone and the Sec machinery. Disulfide oxidoreductases (Dsb) catalyze disulfide bond formation during enzyme folding. The enzyme then interacts with the Xcp pathway where it is transferred through the outer membrane (OM) as mature enzyme.

proteins, particularly in the periplasm with the Xcp machinery, are poorly understood. Substrates are recruited to the Xcp pathway through mechanisms that are not governed by a universal signal sequence [463]. It is not currently possible to predict how an extracellular protein will interact with the secretion systems of a heterologous host.

In this report, we describe our attempts to express a heterologous enzyme in several *Pseudomonas* species. Our objective was to develop a heterologous bacterial host to produce and secrete this enzyme at titers equivalent to or greater than what is possible by the native host. We initially investigated heterologous production in *P. putida*, a microorganism in which our lab has experience expressing heterologous genes. Before attempting to express heterologous enzyme, we characterized a set of inducible promoters and broad-host-range origins in two strains of *P. putida*, KT2440 and EM383. The RK2 origin and the *trc* promoter were the best combination for heterologous expression, and EM383 produced slightly more heterologous protein and secreted fewer native extracellular proteins. After demonstrating heterologous expression of enzyme, we found that the strength of the RBS did not have a noticeable effect on production. However, enzyme secretion in *P. putida* appeared to be limited. We used Western blots to determine that enzyme was accumulating primarily in the periplasm, suggesting that it was not being recruited efficiently through the Xcp secretory pathway. Five alternative *Pseudomonas* hosts were investigated for heterologous production, and we were only able to detect expression in one alternative strain, *P. fluorescens* B-2550. Like *P. putida*, the enzyme accumulated

primarily in the periplasm, and secreted enzyme was not detected in significant amounts in the supernatants of production cultures.

6.3 Materials & Methods

6.3.1 Strains, Plasmids, and Growth Conditions

All strains and plasmids used in this work are listed in Table 6.1. *E. coli* DH5α was grown in LB at 37°C and *Pseudomonas putida* (KT2440 and EM383) were grown in LB at 30°C. All strains of *Pseudomonas fluorescens* and *Pseudomonas acidophila* were grown in nutrient broth (3 g/L beef extract, 5 g/L peptone) at 30°C. Kanamycin was added to media at a final concentration of 50 μg/mL (Kan50) when applicable. LB powder and nutrient broth powder were purchased from Becton, Dickinson and Company. PCR products were generated using NEB Q5 High-fidelity 2X master mix. Plasmids were assembled using the NEBuilder HiFi DNA assembly kit.

Enzyme production cultures were performed in Reisenberg-Korz media or phosphate-limiting media containing 50 μg/mL kanamycin [464, 381]. RK medium was prepared by mixing 13.3 g potassium phosphate monobasic, 4.0g diammonium phosphate, 1.7 g citric acid, 0.1 g Fe(III) ammonium citrate, and 25 g glycerol or 25 g D-glucose to 800mL deionized/distilled water. To this solution, 10 mL of sterile 100X RK batch trace minerals and 10 mL of sterile 120 g/L magnesium sulfate heptahydrate were added. The pH was adjusted to 6.7 with 5 M NaOH and the volume was adjusted to 1 L. The 100X trace minerals solution was prepared by adding 0.42 g EDTA, 0.125 g cobalt(II) chloride hexahydrate, 0.75 g manganese(II) cloride tetrahydrate, 0.06 g copper(II) chloride, 0.15 g boric acid, 0.125 g sodium molybdate dihydrate, and 0.65 g zinc acetate dihydrate to 300 mL deionized/distilled water and adjusting to a final solution volume of 500 mL. All media components and the final media formulation were sterilized by filtration.

Phosphate-limiting media was prepared with 0.12 M Tris, 80 mM NaCl, 20 mM KCl, 20 mM NH_4Cl, 3 mM Na_2SO_4, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 2 μM $ZnCl_2$, 0.5% (w/v)

Table 6.1: Strains and plasmids used in this work.

Strain or plasmid	Genotype and relevant characteristics	Source
Strains		
E. coli DH5α	F- Φ80*lacZ*ΔM15 Δ*(lacZYA-argF)* U169 *reacAI endAI hsdR17* (r_{k-} m_{k+}) *phoA supE*44 *thi*-1 *gyrA*96 *relA*1 λ-	Invitrogen
P. putida KT2440	Wild-type	ATCC 47054
P. putida EM383	KT2440 ΔPP4329-PP4397 (flagellar operon), ΔPP3849-PP3920 (prophage 1), ΔPP3026-PP3066 (prophage 2), ΔPP2266-2297 (prophage 3), ΔPP1532-1586 (prophage 4), ΔTn7, Δ*endA*-1, *endA*-2, Δ*hsdRMS*, Δflagellum, ΔTn4652	[465]
Pseudomonas fluorescens B-10	Environmental isolate	NRRL
Pseudomonas fluorescens B-11	Environmental isolate	NRRL
Pseudomonas fluorescens B-1603	Environmental isolate	NRRL
Pseudomonas fluorescens B-2550	Environmental isolate	NRRL
Pseudomonas acidophila ATCC 31363	Wild-type	NRRL
Plasmids		
pBBR1-Prha-sfGFP	Broad-host range (BHR) sfGFP expression vector; *rha* promoter, BBR1 origin	This work
pBBR1-Ptrc-sfGFP	BHR sfGFP expression vector; *trc* promoter, BBR1 origin	This work
pBBR1-Pupx-sfGFP	BHR sfGFP expression vector; *upxB* promoter, BBR1 origin	This work
pRK2-Prha-sfGFP	BHR sfGFP expression vector; *rha* promoter, RK2 origin	This work
pRK2-Ptrc-sfGFP	BHR sfGFP expression vector; *trc* promoter, RK2 origin	This work
pRK2-Pupx-sfGFP	BHR sfGFP expression vector; *upxB* promoter, RK2 origin	This work
pRSF1010-Prha-sfGFP	BHR sfGFP expression vector; *rha* promoter, RSF1010 origin	This work
pRSF1010-Ptrc-sfGFP	BHR sfGFP expression vector; *trc* promoter, RSF1010 origin	This work
pRSF1010-Pupx-sfGFP	BHR sfGFP expression vector; *upxB* promoter, RSF1010 origin	This work
pRK2-Ptrc-RBS1-FL-enz-chp	BHR enzyme/chaperone expression vector; *trc* promoter, RK2 origin, full enzyme leader, RBS sequence 1	This work
pRK2-Ptrc-RBS1-SL-enz-chp	BHR enzyme/chaperone expression vector; *trc* promoter, RK2 origin, truncated (short) enzyme leader, RBS sequence 1	This work
pRK2-Ptrc-RBS2-FL-enz-chp	BHR enzyme/chaperone expression vector; *trc* promoter, RK2 origin, full enzyme leader, RBS sequence 2	This work
pRK2-Ptrc-RBS2-SL-enz-chp	BHR enzyme/chaperone expression vector; *trc* promoter, RK2 origin, truncated (short) enzyme leader, RBS sequence 2	This work
pRK2-Ptrc-RBS2-FL-enz(FLAG)-chp	BHR enzyme/chaperone expression vector; *trc* promoter, RK2 origin, full enzyme leader, RBS sequence 2, C-terminal FLAG-tag on enzyme	This work
pRK2-Ptrc-RBS2-FL-FLAG-enz-chp	BHR enzyme/chaperone expression vector; *trc* promoter, RK2 origin, full enzyme leader, RBS sequence 2, N-terminal FLAG-tag on enzyme	This work
pRK2-Ptrc-RBS2-FL-Eenz-chp	BHR enzyme/chaperone expression vector, codon optimized operon for *E. coli*; *trc* promoter, RK2 origin, full enzyme leader, RBS sequence 2	This work
pRK2-Ptrc-RBS2-FL-FLAG-Eenz-chp	BHR enzyme/chaperone expression vector, codon optimized operon for *E. coli*; *trc* promoter, RK2 origin, full enzyme leader, RBS sequence 2, N-terminal FLAG-tag on enzyme	This work

casein peptone (pancreatic digest of casein), and 0.5% (w/v) glucose or glycerol. The final pH was adjusted to 7.5 with HCl. To make high-phosphate conditions, 61 μL of phosphate solution (94 g/L KH_2PO_4, 22 g/L K_2HPO_4) was added for every 5 mL media.

6.3.2 Electroporation protocols

All plasmids were introduced into *Pseudomonas* strains via electroporation. For *P. putida* strains, 500 μL of overnight culture in LB were washed twice with 1 mL ice-cold 10% glycerol. The cells were resuspended in 50 μL 10% glycerol. Samples were kept on ice

at all times between centrifugation. Approximately 1 μL of plasmid DNA was added to each aliquot of competent cells. Samples were transferred to chilled 1-mm electroporation cuvettes and then electroporated at 1.8 kV with time constants around 5 ms. Cells were diluted in 1 mL LB and recovered for 1 hour at 30°C in shaking incubator.

For all other *Pseudomonas* species, an electroporation protocol was adapted from Cold Spring Harbor Protocols [466]. Twenty-five milliliters of fresh nutrient broth were inoculated with 500 μL culture. Cells grew for 4-6 hours until the cultures reached an OD of approximately 0.4. Flasks were transferred to ice-water for 15 minutes. Cultures were transferred to 50-mL centrifuge tubes and pelleted by centrifugation at 1,000g for 15 minutes at 4°C. After the supernatant was decanted, pellets were resuspended in 25 mL ice-cold sterile water. The pellets were washed several more times by centrifugation for 20 minutes and resuspended in 10 mL water, 5 mL water, and finally 500 μL GYT medium (10% (v/v) glycerol, 1.25 g/L yeast extract, 2.5 g/L tryptone). One hundred-microliter aliquots were prepared from the final cell suspension and 1 μL plasmid DNA was added to each one. Samples were transferred to 2-mm electroporation cuvettes and electroporated at 2.5 kV. Cells were diluted in 1 mL nutrient broth and recovered for 1 hour at 30°C in the shaking incubator.

6.3.3 Fluorescent reporter production assay

Fluorescent reporter measurements were completed in a Tecan M1000 plate reader in black 96-well plates with clear bottoms. Each well contained 100 μL RK media with glycerol with 50 μg/mL kanamycin. Induced wells contained 1 mM IPTG, 0.2% (w/v) L-rhamnose, or 4 g/L potassium phosphate (instead of 13.3 g/L). Each well was inoculated with 1 μL overnight culture in LB. The plate was incubated at 30°C for 48 hours with linear shaking (4 mm radius). Cell growth was monitored by absorbance at 600 nm and sfGFP production was monitored by fluorescence intensity (excitation: 485 nm, emission: 510 nm).

6.3.4 BCA protein assay

Bicinchoninic acid (BCA) assays were completed with the Pierce BCA protein assay kit from Thermo Fisher. Protein standards were prepared with provided bovine serum albumin (BSA) solutions by dilution in phosphate-buffered saline. To determine background signal in *P. putida* culture supernatants, BSA solutions were diluted in supernatants collected from cultures. Twenty-five microliters of each sample and standard were added to wells in a 96-well plate. Two hundred microliters of working reagent was added to each sample and the plate was incubated at 37°C for 30 minutes. At the end of the incubation period, the plate was cooled to room temperature and the absorbance at 562 nm for each sample was measured on a plate reader.

6.3.5 Enzyme production and protein extraction

Enzyme production cultures were started by inoculating 3 mL production media with overnight cultures to OD=0.1. Production media was RK media, nutrient broth, or phosphate-limiting media where specified. Cultures were incubated at 30°C until they reached an OD of approximately 1.0 (4-6 hours after inoculation) and then induced with 1 mM IPTG. Cultures continued to grow for up to 16-24 hours after inoculation.

Proteins were extracted for analysis from culture supernatant and total protein from cell-free extracts. Supernatants were collected by centrifugation and filtered with a 0.22 μm sterile filter. Supernatant proteins were then concentrated using a protein concentration filter with a 3 kDa cutoff. Total protein extracts were prepared from pellets using the BugBuster Master Mix, following the standard protocol. Protein from total soluble extracts and inclusion bodies were collected using this protocol. All fractions were analyzed by SDS-PAGE and western blot.

When stated in the text, periplasmic preps were also extracted from enzyme production cultures as reported previously [467]. After removing the supernatant fractions, cell pellets from 500 μL culture were incubated in 250 μL 30 mM Tris-HCl, pH 8.0, 20% sucrose, 4 mM EDTA, and 0.5 mg/mL lysozyme for 60 minutes at 30°C in the shaking in-

cubator. Two minutes after samples were removed from the incubator, $MgCl_2$ was added to a final concentration of 10 mM. The samples were then centrifuged at 4°C to collect the supernatant containing the periplasmic proteins. Total soluble protein and inclusion bodies were then extracted from the pellets resulting from this prep. All fractions were analyzed by SDS-PAGE and western blot.

6.3.6 SDS-PAGE and Western blot analysis

Samples were prepped for SDS-PAGE by mixing with 4X NuPAGE SDS loading buffer and denaturing protein by adding 2-mercaptoethanol (2.5% final concentration) and heating samples at 70°C for 10 minutes. Protein samples were analyzed by SDS-PAGE using pre-cast 4-15% polyacrylamide gradient gels running at 180V for 40-45 minutes. Proteins were visualized using the GelCode Blue Stain reagent from Thermo Fisher.

Protein samples with FLAG-tagged proteins were analyzed by western blot in addition to Coomassie staining. Proteins were transferred from the gel to a 0.2 μm nitrocellulose membrane. Transfer buffer contained 260 mL methanol, 9.3 g glycine, and 1.95 g Tris base in 1300 mL aqueous solution. The transfer occurred at 4°C for 1 hour with a voltage of 100V. Antibody incubations were performed with tris-buffered saline with 0.1% Tween-20 (TBST). TBST contains 0.8 g NaCl, 2.42 g Tris base, 1.0 mL Tween-20 in 1 L solution with a pH of 7.6. The membrane was incubated with 10 mL blocking buffer (0.5 g bovine serum albumin in 10 mL TBST) for 1 hour. The primary antibody, Anti-FLAG M2 antibody produced in mouse, was added to blocking buffer at a final concentration of 1-10 μg/mL, and the antibody solution was incubated with the membrane for 2-3 hours at room temperature or overnight at 4°C. Afterwards, the membrane was washed three times with TBST and then incubated with the secondary antibody solution (goat anti-mouse IgG HRP conjugate diluted 1:10,000 in blocking buffer) for 1 hour at room temperature or overnight at 4°C. The membrane was washed three times in TBST and then prepped for visualization. Antibody-bound proteins were visualized using the Pierce ECL western blot substrate.

6.3.7 Mass spectrometry of specific proteins bands in SDS-PAGE

Samples were analyzed at the Mass Spectrometry / Proteomics Facility at UW-Madison using their In-Gel Digest service. Bands of interest were cut from SDS-PAGE gels stained with Coomassie G250 and destained. A trypsin in-gel digest was performed and the resulting peptides were extracted from the gel. Peptides were purified with a C18 Zip-Tip cleanup and analyzed on an Orbitrap LC-MS/MS. Identified peptides were aligned with known proteins from *P. putida* KT2440 and the protein sequences of the enzyme and chaperone.

6.4 Results & Discussion

6.4.1 Analysis of secreted proteome in *P. putida* supernatants

Before attempting to produce and secrete enzyme, we investigated methods for characterizing heterologous protein production in supernatants from *P. putida* cultures. The BCA assay can be used to detect protein amounts in a 96-well plate format, allowing more rapid analysis of samples than what is normally possible with SDS-PAGE. We used a BCA assay to determine the background amount of protein in the supernatants of *P. putida* cultures in two media formulations, high and low phosphate. Low phosphate conditions have been shown to induce the secretion of enzymes from *P. putida*, so we hypothesized that phosphate availability would affect the proteins present in culture supernatants [464]. According to the BCA assay results, the supernatants with no added protein yielded an absorbance corresponding to approximately 1 mg/mL protein (Figure 6.2a). The lowest background signal was in RK media with glycerol after 48 hours of cultivation, and the low-phosphate formulations had the highest background. The linear range of detection in RK media after 48 hours spanned most concentrations of BSA, and added protein could be detected over 500 μg/mL. These results suggest that if we were to use BCA assays to identify cultures secreting enzyme, then production titers would

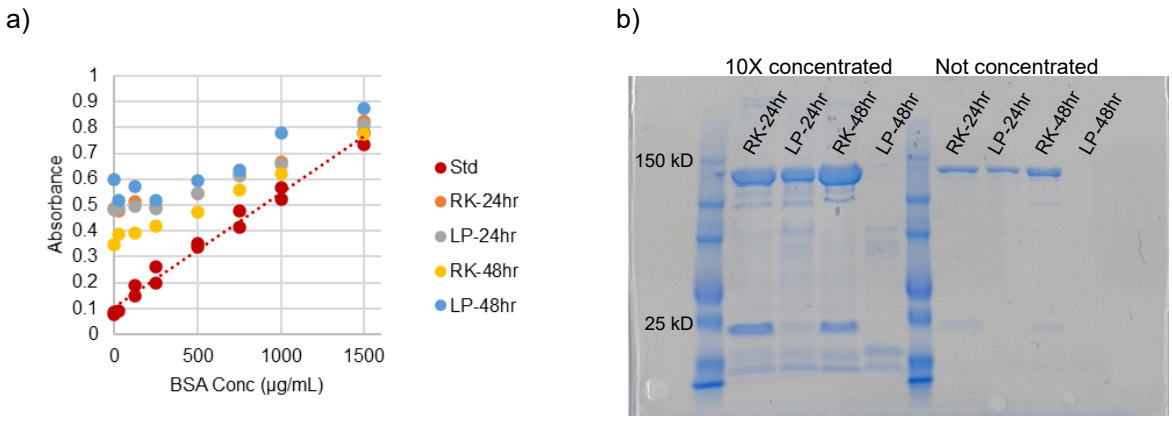


Figure 6.2: **Analysis of protein content in supernatant of *P. putida* KT2440 cultures. a)** Background signal in BCA assay during analysis of *P. putida* supernatants in different media. X-axis represents BCA concentration added to PBS or culture supernatant. **b)** Proteins in culture supernatants visualized by SDS-PAGE. "10X concentrated" samples were concentrated using a protein concentrator column with a 3 kD cutoff. RK-XXhr – culture supernatant from RK glycerol after XX hr, LP-XXhr – culture supernatant from low-phosphate RK glycerol after XX hr

need to be at least 750 μg/mL enzyme.

We also visualized the proteins located in the supernatant in *P. putida* cultures (Figure 6.2b). Major bands appear just under 150 kD in most samples. No major bands appear in culture supernatants from low-phosphate RK media after 48 hours, despite having the strongest background signal in a BCA assay. RK media uses phosphate as a buffer, and the low-phosphate media does not have an additional buffering system, so these results are likely due to the culture having a much lower pH than the other culture supernatants. These conditions could have affected the signal in BCA assays and reduced the stability of proteins secreted into the supernatant. Nevertheless, these results show that phosphate availability in cultures does affect what native proteins are secreted.

6.4.2 Comparing plasmid construction for heterologous expression

In past projects, our lab has had inconsistent results when expressing heterologous proteins from replicative vectors in *P. putida* [225]. We set out to verify a selection of plasmid origins and inducible promoters that would eventually drive enzyme expression. The BBR1, RK2, and RSF1010 origins are commonly used in replicative vectors in Pseudomonads [176]. The *rha* promoter (P_{rha}, induced by L-rhamnose) and the *trc* promoter

Figure 6.3: **Expression vectors tested in *P. putida* KT2440 and EM383.** Plasmid maps for expression vectors that had significant expression of sfGFP in *P. putida*. Expression vectors with P_{uxpB} are not pictured because no sfGFP expression was detected from these constructs.

(P_{trc}, induced by IPTG) have been used recently to chromosomally express heterologous pathways. We cloned all combinations of these origins and promoters along with the gene encoding the reporter protein sfGFP to compare their performance in *P. putida* (Figure 6.3). Expression experiments were completed with *P. putida* KT2440 and EM383. EM383 is a derivative of KT2440 that has been optimized for heterologous protein production [465].

To compare the performance of these expression vectors, we set up production cultures in 96-well plates and monitored growth and GFP fluorescence over a 48-hour growth period. Vectors with the BBR1 or RSF1010 origin had a negative effect on cell growth (Figure 6.4). *P. putida* KT2440 cultures with these origins did not reach a final OD600 above 0.5, whereas cultures with the RK2 origin reached at least an OD600 of 1. *P. putida* EM383 cultures with the BBR1 origin reached a final OD approaching 1, but still had lower values than equivalent cultures with the RK2 origin. Vectors with the RSF1010 origin had the greatest effect on growth, with growth appearing to stop after 10-20 hours of incubation. Interestingly, cultures that were induced with 1 mM IPTG reached higher final OD values than the corresponding uninduced strains.

We were able to detect GFP expression from vectors containing P_{rha} and P_{trc}. These promoters had similar levels of GFP production at max induction, but P_{rha} oddly had

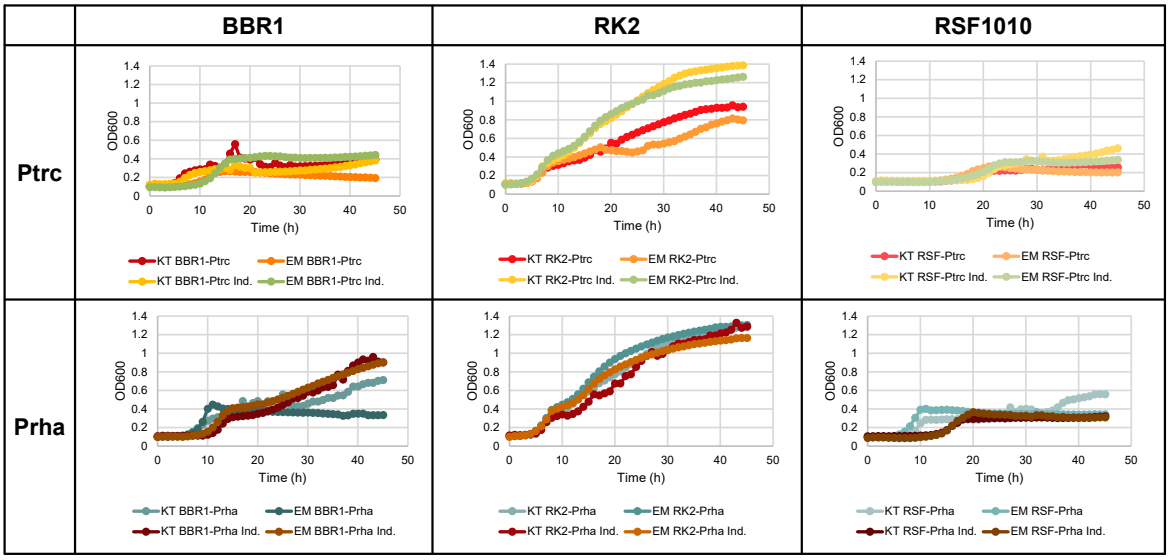


Figure 6.4: **Growth of *P. putida* KT2440 and EM383 with various origins and promoters.** KT – KT2440 cultures, EM – EM383 cultures. Induced cultures are designated by "Ind." P_{trc} cultures were induced with 1 mM IPTG and P_{rha} cultures were induced with 0.2% (w/v) L-rhamnose.

extremely high expression in the absence of the inducer in some cultures, to the point where no induction of gene expression was detected (Figure 6.5). *P. putida* EM383 cultures with P_{rha} and either the BBR1 or RSF1010 origin did have low basal expression and greater than 2-fold induction of gene expression with the addition of inducer. These results were surprising because expression constructs with P_{rha} typically have low basal expression when integrated onto the *P. putida* chromosome. It is unclear how EM383 allows for low basal expression from this promoter while KT2440 does not. We have shown previously that LacI expression needed to be reconfigured to improve inducible expression from P_{trc} (Chapter 3) [225]. Induction of P_{rha} requires autoinduction by RhaR of its own promoter, so changes in RhaR production due to an increased gene copy number could lead to the promoter behaving unexpectedly [468]. For example, it has been demonstrated that P_{rha} can be induced by other carbohydrates if RhaS expression is uncoupled from RhaR regulation [469]. It is possible that a native metabolite in *P. putida* is activating expression from the *rha* promoter in the absence of L-rhamnose, and differences in *rha* induction between KT2440 and EM383 could be due to metabolic changes between the two strains as a result of the genomic deletions in EM383.

The fold induction from cultures with P_{trc} vectors spanned 1-2 orders of magnitude,

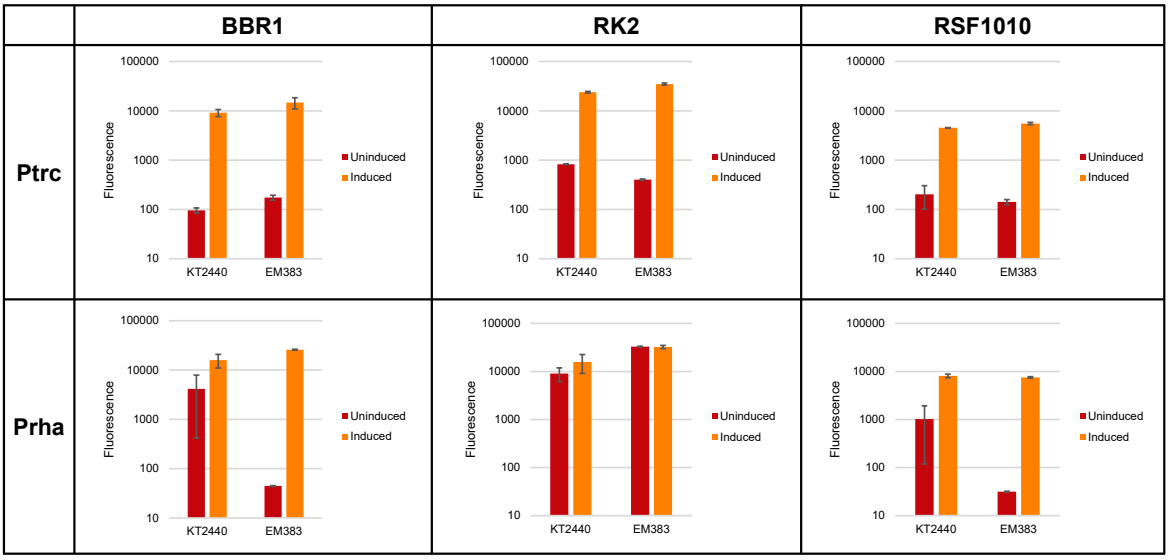


Figure 6.5: **Fold induction of sfGFP expression in *P. putida*.** Fluorescence values from *P. putida* KT2440 and EM383 cultures expressing sfGFP. Values reported were calculated from measurements at end of cultivation period, 48 hr. P_{trc} cultures were induced with 1 mM IPTG and P_{rha} cultures were induced with 0.2% (w/v) L-rhamnose.

regardless of the origin or strain used. The consistent performance of P_{trc} clearly demonstrated that it is the ideal promoter for heterologous gene expression, and the highest maximum expression from this promoter was achieved with the RK2 origin. Cultures reaching a higher OD with plasmid containing the RK2 origin suggests that cells were experiencing less stress due to plasmid maintenance or heterologous protein production. Expressing a heterologous protein that is intended to be secreted is likely to introduce other stresses, particularly in the periplasm and at the inner and outer membranes, so it is important to reduce any stress for the platform strain, if possible. For these reasons, we chose to use expression vectors containing the RK2 origin and *trc* promoter for all experiments involving the expression of extracellular enzyme.

6.4.3 Detecting enzyme production from *P. putida*

After determining the optimal expression vector for heterologous expression in *P. putida*, we designed several expression constructs for the enzyme/chaperone operon. This operon comes from a high-GC organism (approx. 70%), but codon optimizing genes with a high GC content for *P. putida* can sometimes be detrimental to heterologous expression [311].

In addition, most rare codons are shared between *Pseudomonas* species and the enzyme's native host. We opted for a minimal optimization strategy where only rare codons (usage less than 10% in *P. putida*) were replaced with more common codons. We also designed two ribosome binding site (RBS) sequences to test different translation initiation rates of the enzyme by varying the Shine-Dalgarno sequence [470]. The 5' untranslated region (UTR) with the stronger RBS was RBS1, "GAATTCAAAAGATCTTTTAAG***GAGGAA***ATATACAT", and the weaker sequence was RBS2, "GAATTCAAAAGATCTTTTAAG***GAGAAC***ATATACAT" (Shine-Dalgarno sequence in bold) (Figure D.1). Based on predictions by the RBS calculator, RBS1 has a four-fold greater translation initiation rate than RBS2 [399, 413]. Lastly, we designed a second version of the operon with a truncation in the leader sequences of the enzyme gene. A downstream ATG codon was predicted to have a higher translation initiation rate than the annotated start codon, and a second promoter was identified within the first few codons of the full-length leader sequence. The truncated leader sequence was also predicted to have a stronger secretion signal via the Sec pathway (Table D.1). Expression constructs with combinations of the RBS and leader sequences were tested in *P. putida* KT2440 and EM383 for enzyme production and secretion by SDS-PAGE (Figure 6.6).

In initial production cultures, enzyme was not immediately detected by SDS-PAGE in culture supernatants of *P. putida* KT2440 or EM383 (Figure 6.6a,b). In addition, EM383 supernatants had fewer visible protein bands compared to KT2440 supernatants. However, in samples where expression of enzyme with the full leader sequence was induced, the number of visible protein bands in the supernatant increased. This effect was most noticeable in EM383 supernatants (Figure 6.6c). Two possible explanations for this result could be that cells were lysing in the cultures due to enzyme production, or enzyme production is inducing a response by the cells leading to more protein secretion. Analysis of total soluble protein from cell pellets revealed that a prominent band around 25 kD is no longer present in intracellular extracts of these samples, suggesting that the increase in the number of protein bands in the supernatants is due to selective secretion by the cells. However, cultures overexpressing the full-leader enzyme often reached a lower cell

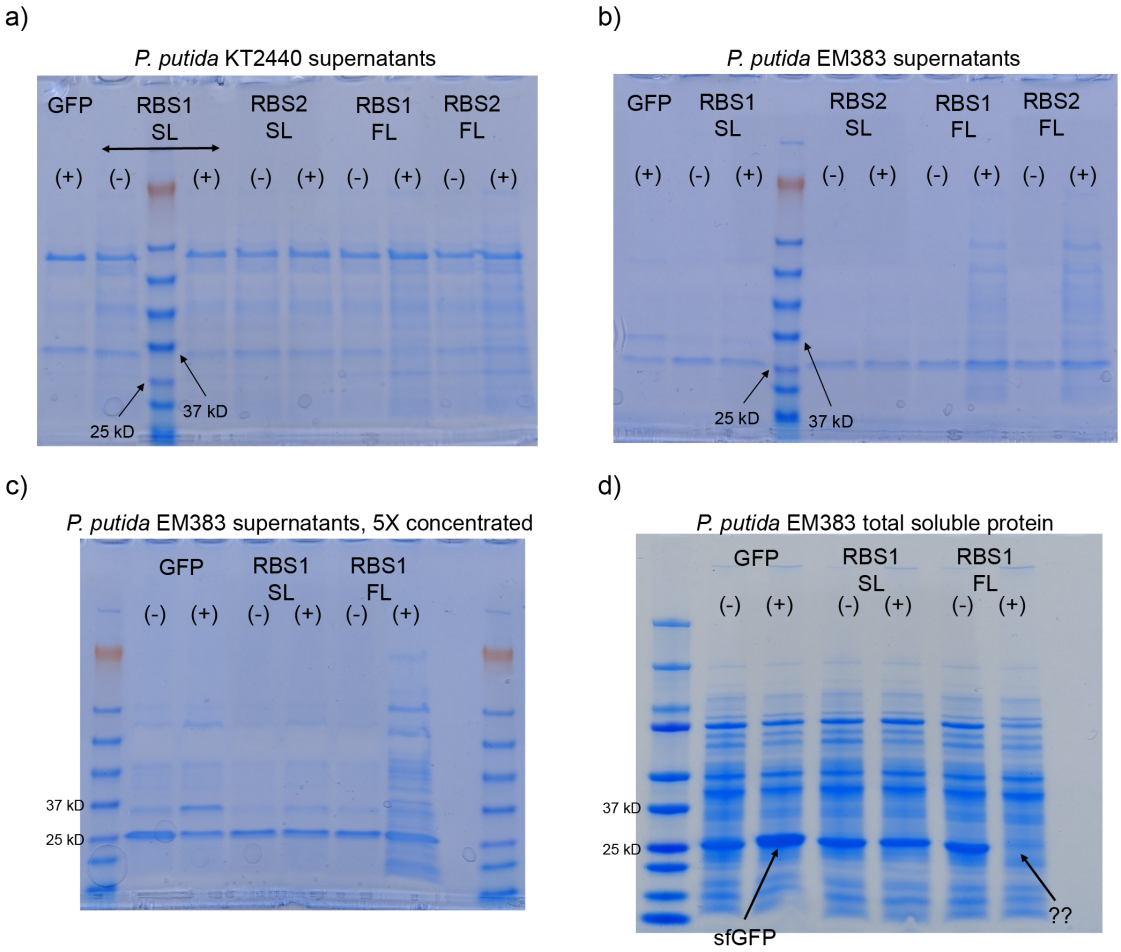


Figure 6.6: **Initial SDS-PAGE analysis of *P. putida* cultures expressing enzyme.** **a)** Supernatants (not concentrated) from *P. putida* KT2440 cultures with various enzyme expression constructs. **b)** Supernatants (not concentrated) from *P. putida* EM383 cultures with various enzyme expression constructs. **c)** Concentrated supernatants from *P. putida* EM383 cultures with enzyme expression constructs with RBS1 only. No significant levels of enzyme is detected in culture supernatants. It appears that overexpression of enzyme with the full leader sequence leads to more proteins in the culture supernatants. **d)** Total soluble protein extracts from *P. putida* EM383 cultures with enzyme expression constructs with RBS1 only. sfGFP can be detected in total soluble protein extracted from cell pellets. A native *P. putida* protein around 25 kD is not present in pellet extract when full-leader enzyme expression is induced (marked by ??). GFP – no enzyme control, pRK2-Ptrc-sfGFP; RBS1 – enzyme expression vector with RBS1; RBS2 – enzyme expression vector with RBS2; SL – enzyme expression vector with truncated enzyme leader; FL – enzyme expression vector with full enzyme leader; (-) – no IPTG added to culture; (+) – 1 mM IPTG added to culture. Production media was RK glycerol and protein extracts were collected after 48 hours of incubation.

density towards the end of cultivation, which could be a result of increased cell death in the later growth stages of these cultures.

At this point, we were concerned that overexpression of enzyme could be detrimental to cell growth, so we tested different levels of inducer concentration to determine if a lower IPTG concentration is optimal for expression. Cultures with lower inducer concentrations reached higher final OD values, and those induced with 0.1 mM IPTG had the most noticeable increase in proteins present in the supernatant. EM383 cultures overall had fewer protein bands appear in the supernatant extracts. No obvious bands corresponding to heterologous enzyme could be identified.

6.4.4 Identifying enzyme and chaperone by mass spectrometry

Results so far have been unable to prove that the enzyme was produced in *P. putida* cultures. To determine if enzyme production was too low to reliably detect by SDS-PAGE or if soluble enzyme was not being produced at all, we analyzed some protein bands appearing around the enzyme molecular weight by mass spectrometry. The supernatant extract from a EM383 culture overexpressing the full leader enzyme had several distinct bands around the expected protein size, and the corresponding total soluble protein extract also had a band of similar size. Lastly, the band appearing in culture supernatants at a smaller size is also analyzed. This band was unexpected to contain the enzyme target or its chaperone, but if its identity was determined, then we could have used this information to engineer a production strain in the future with its corresponding gene deleted from the chromosome.

As anticipated, the protein bands were confirmed to contain both the enzyme and chaperone. The peptides matching the enzyme sequence had a sequence coverage of 34%, and the sequence coverage for the chaperone was 51%. The smaller band surprisingly had hits corresponding to proteins involved in central metabolism of *P. putida* that are not expected to be targeted for secretion. The most abundant peptide group aligned with ProC, pyrroline-5-carboxylate reductase, which is involved in L-proline biosynthesis. A significant number of peptides aligned with Pgi, 6-phosphogluconolactonase, which

converts glucose-6-P to gluconate-6-P in the cytoplasm [471]. The presence of these cytoplasmic proteins suggests that cell lysis is contributing to the release of intracellular proteins into the supernatant. Nevertheless, MS analysis showed that production cultures were producing soluble enzyme, although it was unclear if it was mature enzyme with the leader sequence cleaved or if the leader peptide was still intact.

6.4.5 Determining compartmentalization of enzyme

Even though we were able to show that enzyme was being produced, it was clear that it was not being secreted efficiently by *P. putida*. Our next immediate goal was to determine at which step enzyme secretion was failing. We investigated this question experimentally by identifying the location of the enzyme in protein extracts. If the enzyme is found mostly in the cytoplasm, then interactions with the inner membrane, chaperone, and/or Sec secretion machinery are not properly functioning. On the other hand, if the enzyme is mostly in the periplasm, then interactions with the outer membrane and the Xcp secretion machinery are not functional. To detect heterologous enzyme more easily, we constructed expression vectors with the enzyme tagged with the FLAG peptide, either near the N-terminus immediately downstream of the enzyme leader peptide or at the C-terminus.

Before testing for enzyme localization, we compared the protein content of protein extracts of *P. putida* expressing untagged and tagged enzyme. The presence of a C-terminal FLAG tag on the enzyme appears to interfere with enzyme folding because *P. putida* cultures expressing this construct also had a large amount of proteins in the insoluble fractions. Despite this result, we continued working with FLAG-tagged enzyme constructs so that we could determine enzyme localization.

After characterizing the C-terminal FLAG-tagged enzyme construct, we devised a workflow for extracting protein from various compartments of *P. putida* (Figure 6.7). We had been following a simpler workflow using BugBuster master mix, where the supernatant from *P. putida* cultures were collected and the pellets were lysed following the BugBuster protocol. In this new workflow, we incorporated a periplasmic prep for ex-

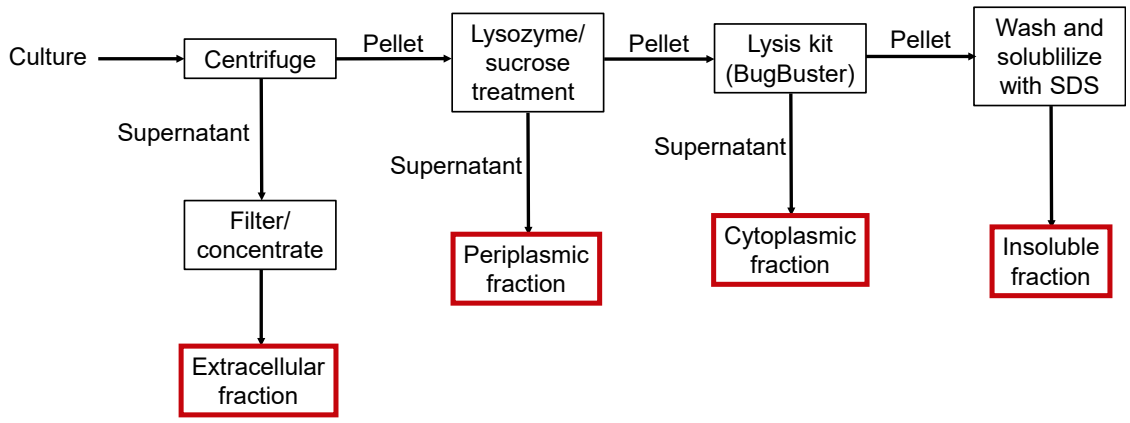


Figure 6.7: **Protein extraction workflow for separating protein fractions from *P. putida* cultures.** Fractions were prepared sequentially, with the extracellular fraction prepared first by collecting supernatant from live culture. Cells are then treated with lysozyme and sucrose to extract periplasmic proteins. The pellet from the periplasmic preps are then fully lysed using the BugBuster master mix protocol. The BugBuster protocol is followed further for preparing the insoluble (inclusion body) fraction.

tracting periplasmic proteins from *P. aeruginosa* [467]. This method enriched for soluble proteins from the periplasm, and the pellet collected was fully lysed using the BugBuster kit. Insoluble protein fractions were also collected in most experiments.

At this point in the project, we also hypothesized that the amount of secreted enzyme was low because the Xcp secretion machinery was not induced in RK media. We had tried to formulate a "low-phosphate" version of RK media that simply had a lower concentration of potassium phosphate added, but this change removed the buffering capacity of the media and allowed the pH to drop significantly during cultivation. From the literature, we identified a phosphate-limited media for *P. putida* that contained an alternative buffer system (Tris) and was used to induce the secretion of native phosphatases via the Xcp pathway [464]. All experiments described from this point on were completed using variations of this media.

Constructs containing the enzyme/chaperone operon with and without a C-terminal FLAG-tag were expressed in *P. putida* EM383. Inducing expression with 1 mM IPTG still led to an increase in visible protein bands in supernatant extracts as seen by SDS-PAGE (Figure 6.8a). The level of phosphate in the media did appear to have a small effect on the amount of protein bands for the untagged enzyme construct, as phosphate-limiting conditions did lead to a lower protein amount in the supernatant for this construct. The relative amount of secreted enzyme did not seem to be greatly affected by phosphate

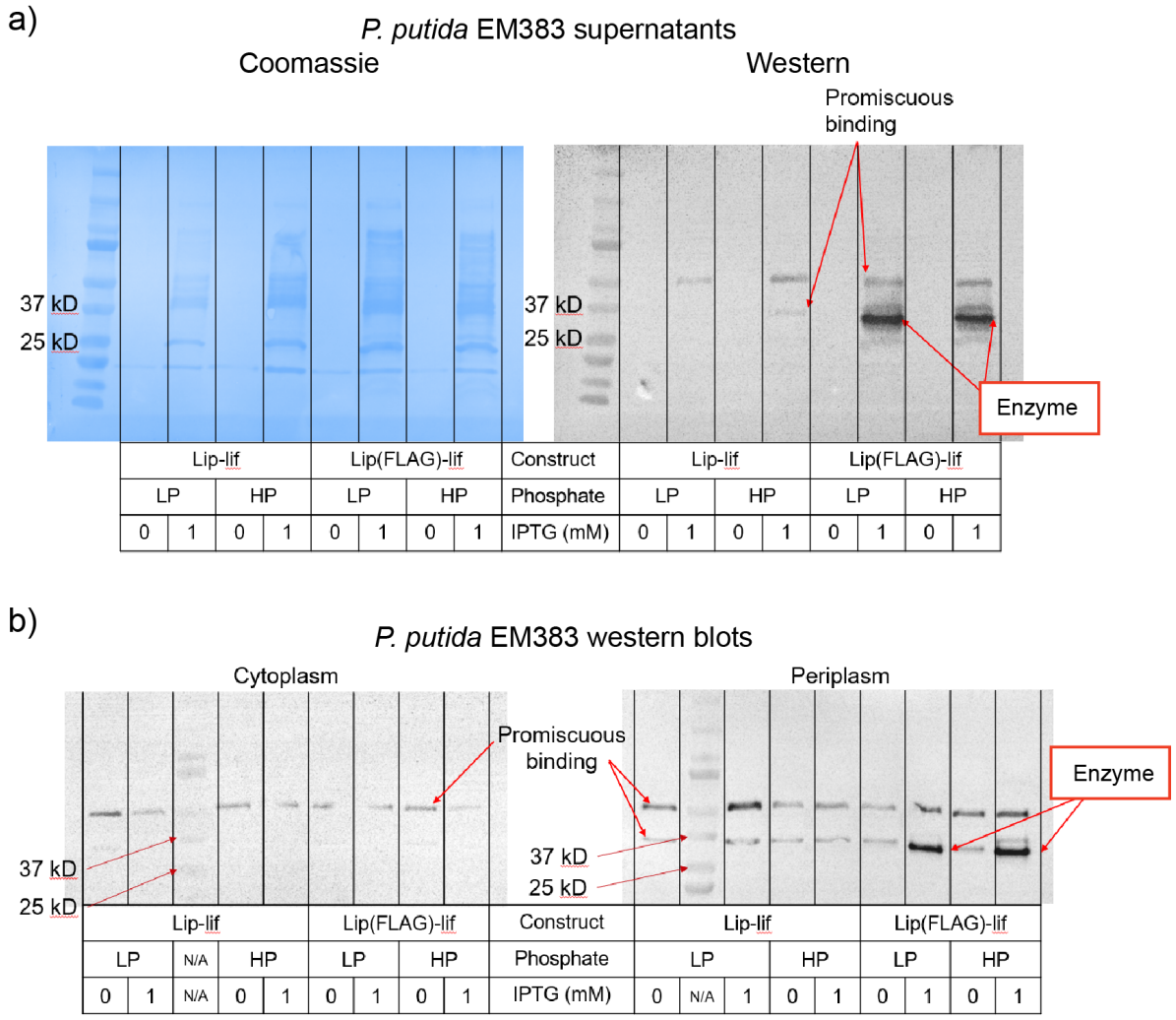
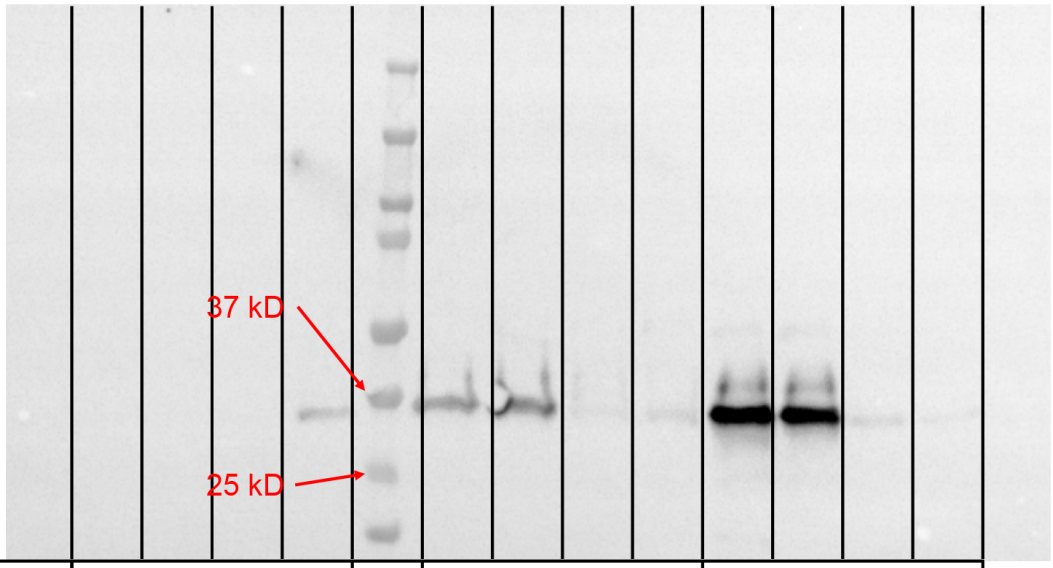


Figure 6.8: **Compartmentalization of heterologous enzyme in *P. putida* production cultures.** **a)** SDS-PAGE and western blot results of *P. putida* EM383 supernatants. **b)** Western blot results of cytoplasmic and periplasmic protein preps of *P. putida* EM383 expressing enzyme-chaperone constructs. Non-specific binding by the M2 anti-FLAG antibody was observed and are marked as "promiscuous binding." LP – phosphate-limiting media; HP – phosphate-limiting media with inorganic phosphate added. Base media was phosphate-limited media with 0.5% (w/v) glycerol. Cultures were grown overnight before preparing protein extractions.

concentration either. Interestingly, more enzyme was detected by Western blot in the periplasm than the cytoplasm in *P. putida* (Figure 6.8b). These images also did not show a major difference in enzyme amounts between samples with high- and low-phosphate media. Overall, these results suggested that enzyme was accumulating in the periplasm in the production host and that a non-optimal interaction with the Xcp machinery is the primary reason for low amounts of enzyme secretion.

Based on our observation that a C-terminal FLAG-tag on the heterologous enzyme may be inhibiting its solubility and leading to cell lysis, we investigated another tagged construct with an N-terminal FLAG-tag. The FLAG peptide was placed between the native enzyme leader and the mature enzyme sequence, so the tag was not truly on the N-terminus unless the native leader was cleaved off. However, the M2 antibody can bind in both cases, and therefore we were not able to differentiate between non-cleaved and mature enzyme by Western blot. We repeated production experiments with this new construct in phosphate-limited media, this time including formulations with glucose as the carbon source, as this was the original phosphate-limited conditions reported in the literature. These experiments again showed little to no secretion of enzyme into the supernatant from any sample (Figure 6.9). The N-terminally tagged enzyme appeared mostly in the cytoplasm and the band signal was much stronger than the C-terminally tagged enzyme in any fraction. Enzyme with a C-terminal tag appeared to be most abundant in the supernatant from cultures with glycerol-based media. It is unknown whether the difference in band signal was due to differences in protein production or in antibody binding to the tagged proteins. During these experiments, we also tested an alternative operon sequence that had been codon optimized for *E. coli*. Based on SDS-PAGE and Western blots, there was not a significant difference in enzyme production and compartmentalization between the two operon sequences.

The stronger signal in Western blots from enzyme constructs with an N-terminal FLAG tag made it difficult to analyze the C-terminal constructs. However, it is possible to see that the enzyme does appear to be compartmentalizing primarily in the supernatant when a C-terminal tag is used and the carbon source is glycerol. This suggests that the



Prep	Supernatant				NA	Periplasm				Cytoplasm			
Carbon Source	Glu	Gly	Glu	Gly	NA	Glu	Gly	Glu	Gly	Glu	Gly	Glu	Gly
FLAG terminal	N	N	C	C	NA	N	N	C	C	N	N	C	C

Figure 6.9: **Compartmentalization of N- and C-terminally tagged enzyme on different carbon sources.** Anti-FLAG Western blot of supernatant, periplasm, and cytoplasm extractions from *P. putida* EM383 cultures expressing constructs with N- and C-terminal FLAG-tags on the enzyme. N-terminally tagged enzyme was more easily detected by Western blot, but the majority was in the cytoplasm, and no enzyme was detected in the supernatant. C-terminally tagged enzyme was found mostly in the supernatant in glycerol-based media, with some appearing the periplasm and cytoplasm. Media used was phosphate-limiting media with 0.5% (w/v) glucose or glycerol. Cultures were grown overnight before preparing protein extracts. Glu – glucose; Gly – glycerol; N – N-terminal tag; C – C-terminal tag.

phosphate-limiting media is inducing *P. putida*'s secretion machinery more when the carbon source is glycerol. The N-terminal FLAG tag completely broke the processing of enzyme via the Sec and Xcp pathways, and no enzyme was detected in supernatant fractions. This result suggests that the presence of C-terminal enzyme in supernatants is due to genuine secretion by the cell rather than cell lysis and release of intracellular proteins. However, the production titers of secreted enzyme are still much lower than the native host, as enzyme is hardly detectable via SDS-PAGE. Significant strain engineering would likely be necessary to increase titers of secreted enzyme. XcpP and XcpQ are two components of the Xcp machinery that are likely to directly interact with secreted proteins, and sequence alignment shows that XcpP and XcpQ from *P. putida* have poor homology to the corresponding proteins of the enzyme's native host. It is possible that XcpP and XcpQ in *P. putida* do not fully recognize the heterologous enzyme, inhibiting efficient secretion.

a)

Strain	Plasmid transformation	Enzyme expression	Enzyme location
P. acidophila ATCC 31363	Yes	Not detected	N/A
P. fluorescens B-10	No	N/A	N/A
P. fluorescens B-11	No	N/A	N/A
P. fluorescens B-1603	No	N/A	N/A
P. fluorescens B-2550	Yes	Yes	Periplasm

b)

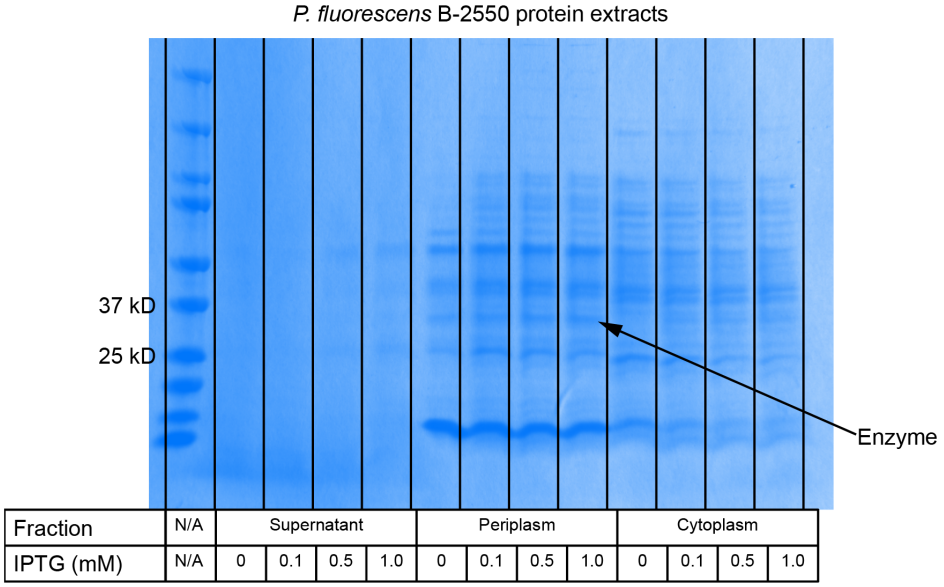


Figure 6.10: **Enzyme secretion in alternative Pseudomonads. a)** Table of alternative hosts used for enzyme expression. Enzyme expression was only detected in *P. fluorescens* B-2550. **b)** SDS-PAGE of protein fractions from cultures of B-2550. Cultures were grown in nutrient broth overnight.

6.4.6 Surveying alternative Pseudomonas hosts for enzyme secretion

After determining the limitations of heterologous enzyme secretion in *P. putida*, we set out to identify and test alternative Pseudomonads as heterologous hosts. The USDA Agricultural Research culture collection (NRRL) contains several strains of *P. fluorescens*, and a significant fraction of them have been shown to secrete enzyme. Other strains of *P. fluorescens* have also been developed for heterologous protein production [459]. We chose four strains from NRRL, two of which exhibited enzyme secretion and two that did not (Figure 6.10a, Table D.2). In addition, we included *P. acidophila* ATCC 31363 in the panel of alternative hosts.

Before attempting to produce heterologous enzyme in these strains, we first needed to demonstrate transformation of the RK2-based expression vectors. The electroporation

protocol that we used for DNA transformation into *P. putida* only led to successful plasmid transformation into *P. fluorescens* B-2550 (Figure 6.10a). We also attempted an electroporation protocol developed for *E. coli* from Cold Spring Harbor [466]. This enabled plasmid transformation into *P. acidophila* ATCC 31363, but we were unable to transform any RK2-based plasmids into the remaining three strains. We attempted to detect enzyme expression from ATCC 31363 and B-2550, and a band corresponding to enzyme was only visible from cultures of B-2550 (Figure 6.10b). Similar to *P. putida*, the enzyme appeared to accumulate primarily in the periplasm. It is possible that expression and secretion could be improved after testing different cultivation parameters, such as media formulation, so that the secretion machinery in the alternative hosts is induced. However, these parameters are less understood for these strains.

6.5 Conclusions

We were unable to produce and secrete significant amounts of heterologous enzyme, likely because conditions that induce the Xcp secretion pathway are not fully understood or the heterologous enzyme did not interact favorably with Xcp proteins. Transfer of the enzyme across the inner membrane, facilitated by the chaperone and heterologous Sec pathway, was not the limiting step towards secreting the enzyme into the supernatant. We detected accumulation of enzyme into the periplasm of both *P. putida* and *P. fluorescens* B-2550, suggesting that this step is less affected by the identity of the heterologous host. If work is continued in *Pseudomonas* bacteria, substantial progress will likely require engineering the secretory pathways of the host, such as expressing heterologous Xcp proteins. This strategy will require reliable synthetic biology tools, so *P. putida* would be the most ideal strain due to the available genetic parts characterized for this bacterium.

6.6 Acknowledgments

The authors would like to thank our corporate collaborators for proposing the project and providing funding for the research discussed here.

Chapter 7

Complementation of levulinic acid catabolism in *P. putida* with mammalian enzymes

Authors and Contributors:

- **Taylor B. Cook** Designed and performed *P. putida* complementation experiments.
- **Edrees H. Rashan** Designed complementation experiments, provided ACAD sequences.
- **David J. Pagliarini** Provided guidance, edited, and reviewed the manuscript.
- **Dr. Brian F. Pfleger** Provided guidance, edited, and reviewed the manuscript.

7.1 Abstract

Levulinic acid is a γ-keto acid that is a common by-product from acid hydrolysis of lignocellulosic biomass. Recently, the pathway in *Pseudomonas putida* KT2440 responsible for the catabolism of this compound was elucidated. The enzymatic reactions include the reduction of the γ-keto group followed by isomerization of the γ-hydroxy group to form a β-hydroxyacyl-CoA, which is then converted to acetyl-CoA and propionyl-CoA via β-oxidation. The isomerization steps involve a phosphorylated intermediate, where the phosphate acts as a leaving group to promote the isomerization reactions. Similar catabolic pathways exist in mammalian cells. For example, *Mus musculus* (house mouse) encodes two enzymes that are members of the acyl-CoA dehydrogenase family, and each enzyme contains two domains that have homology to the enzymes responsible for isomerization of the hydroxyl group. We proposed that these enzymes catalyze similar reactions in their native organism. To generate evidence for the activity of these enzymes, we attempted to express each enzyme in strains of *P. putida* where the homologous enzymes in the levulinic acid pathway were deleted from the chromosome. We monitored the growth of these strains on solid minimal media with levulinic acid as the sole carbon source and observed that ACAD10 expression restored growth for the strain with a deletion of *lvaA*, the gene encoding the phosphotransferase responsible for formation of the phosphorylated intermediate. Neither ACAD10 or ACAD11 restored growth in strains missing LvaC, the enzyme responsible for isomerization.

7.2 Introduction

Levulinic acid is an atypical carbon source for microbial growth. It is a five-carbon γ-keto acid that is derived from the non-enzymatic hydrolysis of hexose sugars in biomass [472]. There is considerable interest in this compound because there are several routes for converting it into valuable chemicals, including biofuels [473], plasticizers [474], and solvents [475]. Even though microorganisms that could use levulinic acid as a carbon source were known, the enzymes responsible for catabolism were unknown until recently, when the catabolic pathway for levulinic acid in *P. putida* KT2440 was elucidated [169]. The authors demonstrated that levulinic acid catabolism involves a 4-phosphoacyl-CoA intermediate before conversion into a 3-hydroxyacyl-CoA that is processed through β-oxidation. This work was a great step towards applying metabolic engineering strategies to develop bioconversions of levulinic acid into high-value chemicals [476].

There are also issues related to human health that involve levulinic acid and its derivatives. Calcium levulinate has been sold as a dietary supplement for calcium administration [477]. Research has also implicated a phosphorylated intermediate in levulinic acid catabolism in rat livers, like the pathway elucidated in *P. putida* [478]. A reduced derivative of levulinic acid, 4-hydroxypentanoate has been revealed as a drug of abuse, similar to 4-hydroxybutyrate [479]. When calcium levulinate has been administered in patients with β-ketothiolase deficiency, an accumulation of 4-hydroxypentanoate was observed [480]. These findings have motivated research into how levulinic acid and other γ-keto acids are catabolized in mammalian systems. Acyl-CoA dehydrogenases (ACAD), which catalyze enoyl-CoA formation from acyl-CoA substrates, are most commonly implicated in these pathways. Mitochondrial ACADs are responsible for the β-oxidation of long-chain fatty acids, which is important for energy homeostasis [481].

Two enzymes that belong to the family of mammalian ACADs are ACAD10 and ACAD11. In humans, these enzymes are highly expressed in the brain and are active on very long chain fatty acyl-CoAs (C20-C26) [482]. Both enzymes have been reported as key determinants in various diseases. Mutations in ACAD10 were shown to be associated with type 2 diabetes in different human populations[483], and ACAD11 has been implicated in

cancer cell survival upon glucose starvation [484]. Their role in fatty acid metabolism is not surprising due to their annotations as ACADs, but both also contain an N-terminal kinase domain whose function in cell metabolism is unclear [485]. These domains are similar to aminoglycoside phosphotransferases, a kinase family mostly found in bacteria.

The aforementioned domains in ACAD10 and ACAD11 have sequence similarity to two enzymes in levulinic acid catabolism. The kinase domain is homologous to LvaA, a phosphotransferase that catalyzes formation of 4-phosphovaleroyl-CoA, and the ACAD domain is similar to LvaC, an enzyme that is also annotated as an ACAD that converts 4-phosphovaleryl-CoA to 3-hydroxyvaleroyl-CoA. Our collaborators from the Pagliarini research group have proposed that LvaA and LvaC are orthologues to these mammalian ACADs and that ACAD10 and ACAD11 catalyze a similar sequence of reactions that isomerize 4-hydroxy-*n*-acyl-CoAs to 3-hydroxy-*n*-acyl-CoAs. We planned to test these enzymes in various cell contexts, including in *P. putida*. We constucted expression vectors of ACAD10 and ACAD11 and demonstrated that ACAD10 can partially complement levulinic acid catabolism in *P. putida*.

7.3 Materials & Methods

7.3.1 Chemicals, strains, and plasmids

The bacterial strains and plasmids described in this chapter are listed in Table 7.1. *E. coli* was grown in LB at 37C and *P. putida* was grown in LB at 30C. When necessary, media was supplemented with antibiotics, including kanamycin (50 μg/mL, Kan50) and gentamycin (30 μg/mL, Gent30). Plasmids designed for this work were all constructed using Gibson assembly [382].

All chemicals were obtained from Sigma-Aldrich or Fisher Scientific. 4-Hydroxyvalerate and 4-hydroxyoctanoate were made from the saponification of their corresponding lactones, γ-valerolactone and γ-octalactone. Solutions containing 2 M lactone were adjusted to pH 12 using 10 M NaOH and incubated for 1 hour. For use in bacterial growth con-

Table 7.1: Strains and plasmids used in this work.

Strain or plasmid	Genotype and relevant characteristics	Source
Strains		
E. coli DH5α	F- Φ80*lacZ*ΔM15 Δ*(lacZYA-argF)* U169 *reacAI endAI hsdR17* (r_{k-} m_{k+}) *phoA supE*44 *thi*-1 *gyrA*96 *relA*1 λ-	Invitrogen
P. putida KTU	KT2440 Δ*upp*	[158]
P. putida KTU Δ*lvaA*	KT2440 Δ*upp* Δ*lvaA*	[169]
P. putida KTU Δ*lvaAB*	KT2440 Δ*upp* Δ*lvaA* Δ*lvaB*	[169]
P. putida KTU Δ*lvaC*	KT2440 Δ*upp* Δ*lvaC*	[169]
P. putida KTU Δ*lvaABC*	KT2440 Δ*upp* Δ*lvaA* Δ*lvaB* Δ*lvaC*	This study
Plasmids		
pBAD35	Empty expression vector with P_{araBAD}, BBR1 origin, and KanR	[169]
pBAD35-lvaA	Expression vector for *lvaA* with P_{araBAD}, BBR1 origin, and KanR	[169]
pBAD35-lvaB	Expression vector for *lvaB* with P_{araBAD}, BBR1 origin, and KanR	[169]
pBAD35-lvaC	Expression vector for *lvaC* with P_{araBAD}, BBR1 origin, and KanR	[169]
pBAD35-nd34ACAD10	Expression vector for ACAD10 with first 34 N-terminal amino acids omitted with P_{araBAD}, BBR1 origin, and KanR	This work
pBAD35-ACAD11	Expression vector for ACAD11 with P_{araBAD}, BBR1 origin, and KanR	This work
pBAD35-nd34ACAD10-D48A	Expression vector for D48A inactivation mutant of ACAD10 with P_{araBAD}, BBR1 origin, and KanR	This work
pBAD35-nd34ACAD10-D463A	Expression vector for D463A inactivation mutant of ACAD10 with P_{araBAD}, BBR1 origin, and KanR	This work
pBAD35-nd34ACAD10-D1040N	Expression vector for D1040N inactivation mutant of ACAD10 with P_{araBAD}, BBR1 origin, and KanR	This work
pBAD35-ACAD11-D220A	Expression vector for D220A inactivation mutant of ACAD11 with P_{araBAD}, BBR1 origin, and KanR	This work
pBAD35-ACAD11-D753N	Expression vector for D753N inactivatino mutant of ACAD11 with P_{araBAD}, BBR1 origin, and KanR	This work
pRK2T-Gm-Ptrc-lvaABDE	Expression vector for lvaABDE P_{trc}, RK2 origin, and GmR	This work
pRK2T-Gm-Ptrc-lvaDE	Expression vector for lvaDE P_{trc}, RK2 origin, and GmR	This work

ditions, the pH of the 4-hydroxyvalerate and 4-hydroxyoctanoate stocks were adjusted with 5 M HCl.

7.3.2 LVA pathway complementation growth assays

Strains of *P. putida* were grown on solid minimal media containing various carbon sources. Solid medium was MOPS minimal agar containing 20 mM glucose, levulinic acid, 4-hydroxyvalerate, or 4-hydroxyoctanoate. The pH was adjusted to 7.2 before sterilization.

Strains containing expression vectors for *lva* or ACAD genes were grown overnight in LB with the appropriate antibiotics. The next morning, 1 mL of cultures were washed one time with 1 mL phosphate-buffered saline (PBS) and diluted 10-fold. Five microliters were applied to agar plates with a pipette and allowed to dry on the bench. Once the plates were dry, they were placed in the 30°C or 37°C incubator for 1-3 days. Images were taken once a day until incubations were complete.

7.4 Results & Discussion

Levulinic acid catabolism in *P. putida* is catalyzed by five enzymes encoded in the *lva* operon. First, LvaE convert levulinic acid to its CoA thioester, and then LvaD reduces the γ-keto group to a hydroxyl group, resulting in 4-hydroxyvaleryl-CoA (Figure 7.1a). Next, LvaA in complex with LvaB phosphorylates the 3-hydroxyl group. LvaC then catalyzes the last sequence of reactions, whereby pentenoyl-CoA is formed after removal of the phosphate group of 4-phosphovaleryl-CoA and subsequently hydrated to from 3-hydroxyvaleryl-CoA. LvaB is a small protein that binds to LvaA and is required for phosphotransferase activity. *P. putida* is capable of processing 3-hydroxyvaleryl-CoA through fatty acid β-oxidation, where the substrates are broken down into acetyl-CoA and propionyl-CoA.

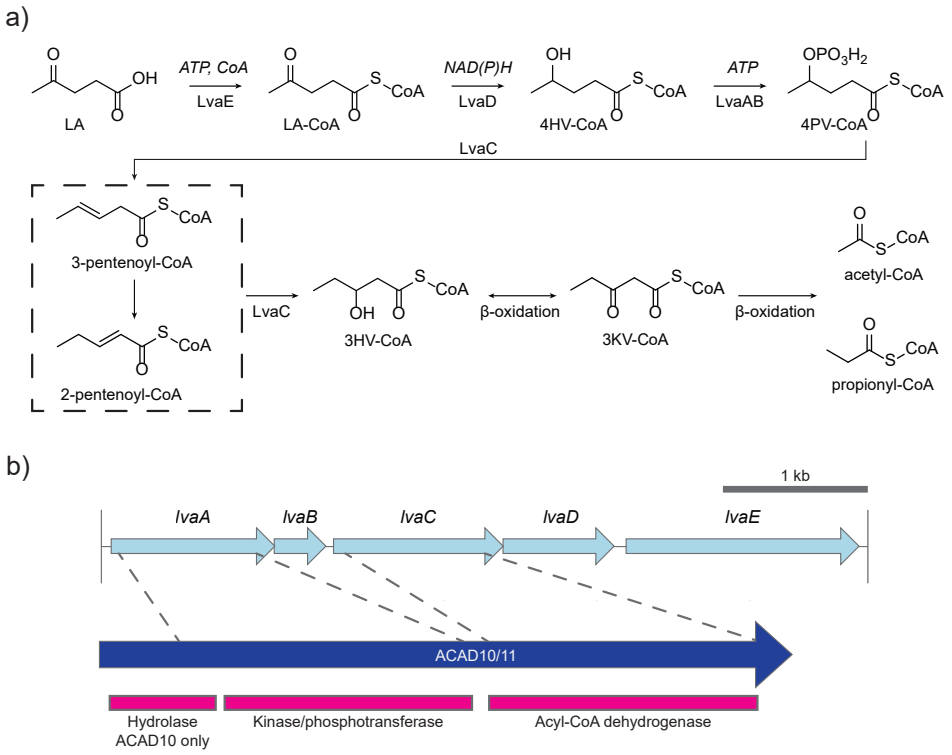


Figure 7.1: **Catabolic pathway for levulinic acid in *P. putida* a)** Enzymatic reactions required for levulinic acid catabolism in *P. putida*. **b)** Genetic structure of *lva* operon encoding enzymes in levulinic acid catabolism. The ACAD enzymes contain two domains that are homologous to LvaA and LvaC, respectively.

ACAD10 contains three distinct catalytic domains: an N-terminal hydrolase domain, followed by a kinase/phosphotransferase domain, and then a C-terminal ACAD domain (Figure 7.1b). ACAD11 contains only the kinase/phosphotransferase and ACAD do-

mains. The kinase domains of ACAD10 and ACAD11 have 23.9% and 24.3% amino acid identity to LvaA, respectively, and the ACAD domains have 33.8% and 33.5% amino acid identity to LvaC. The similarity between aminoglycoside phosphotranferases and eukaryotic kinases was originally observed through structural homology [486], so the low sequence identity between LvaA and the kinase domains is expected. It is not known if the hydrolase domain is involved in the formation of 3-hydroxy-*n*-acyl-CoAs. *In vitro* assays with ACAD11 demonstrated phosphorylation of 4-hydroxyvaleryl-CoA, but so far, soluble production of ACAD10 has not been achieved (correspondance with E. H. Rashan).

To determine if ACAD10 and ACAD11 were capable of replacing LvaA and/or LvaC in levulinic acid catabolism, we obtained single-knockout strains of *P. putida* for *lvaA*, *lvaB*, and *lvaC*. The cDNA for ACAD10 and ACAD11 that was originally codon optimized for *E. coli* expression was then cloned into the pBAD35 backbone. We cloned a truncated version of ACAD10 where the first 34 amino acids were removed because this sequence was predicted to be a mitochondrial transfer peptide by TargetP [487]. We also constructed inactivation mutants of both ACADs. The D463A and D1040N mutations in ACAD10 inactivated the kinase and ACAD domains, respectively. The analogous mutations in ACAD11 were D220A and D753N. To explore the role of the hydrolase domain in ACAD10, we attempted to clone a D48A mutant and a G50D mutant, both mutations in the active site, but we were unable to clone the G50D mutant.

We introduced each expression plasmid into the three *P. putida* strains mentioned above and grew them in minimal media with various carbon sources. Initial experiments were in liquid culture, but the knockout strains were unable to grow in these conditions, even if the complementary expression plasmid was provided. Instead, we prepared agar plates of the selection media and tested strains by transferring liquid cultures to solid media. Four different carbon sources were included in the experiment: levulinic acid, 4-hydroxyvaleric acid, 4-hydroxyoctanoic acid, and glucose. Media with levulinic acid or 4-hydroxyvaleric acid would reveal if the mammalian ACADs could complement portions of levulinic acid catabolism. Adding 4-hydroxyoctanoic acid as a carbon source would

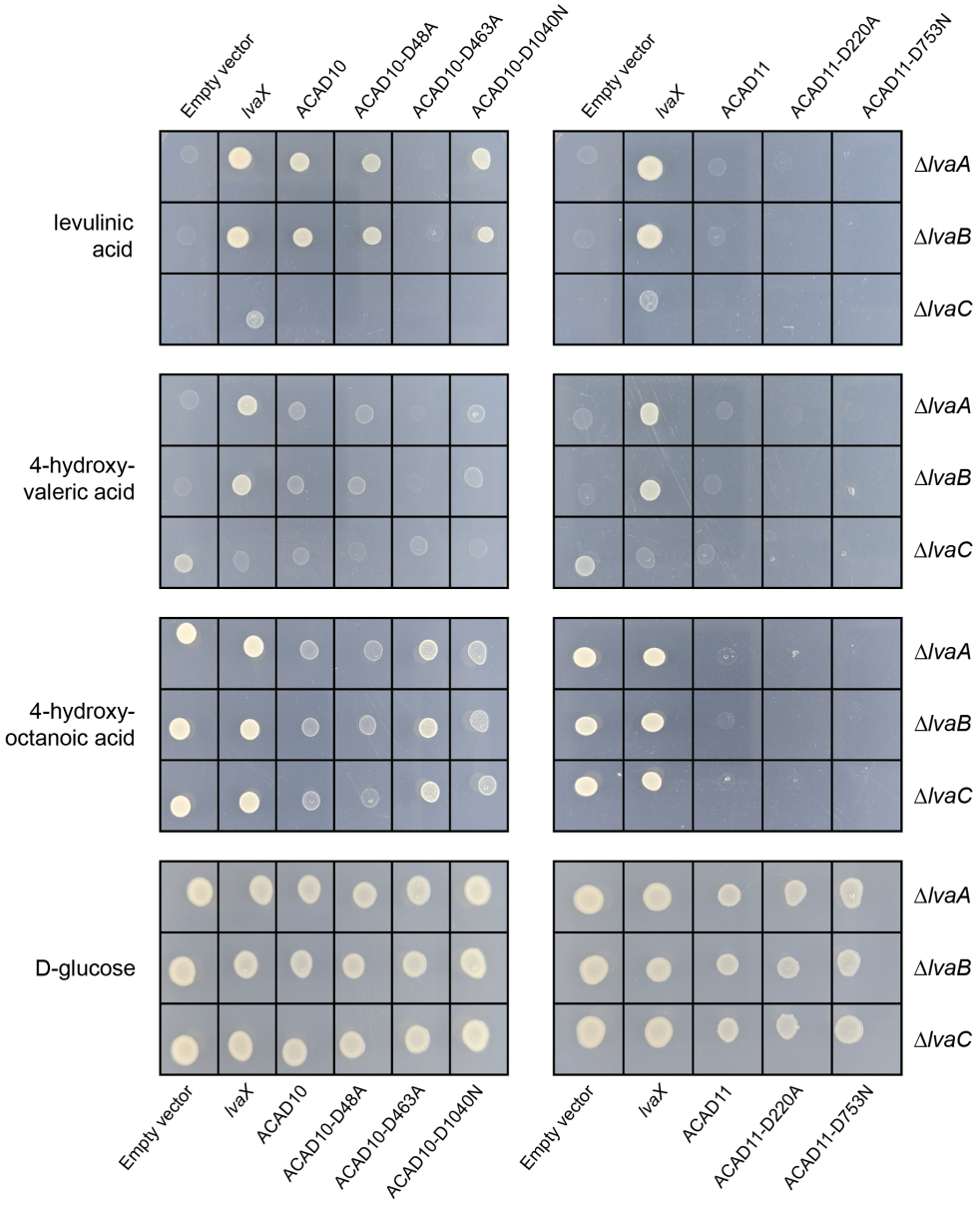


Figure 7.2: **Complementing knockout strains of genes in levulinic acid catabolism with ACAD10 and ACAD11.** Cells were grown on minimal agar with the stated carbon source (left panel). The gene knockout is indicated by the right panel. The top and bottom panels state the expression vector in the corresponding strain. *lvaX* - expression vector contains the same gene that is deleted in the host strain. All plasmids were on the pBAD35 backbone.

reveal if the enzymes have different enzymatic activities on a longer-chain substrate. In most cases, knockout strains with complementary expression vectors grew on levulinic acid and 4-hydroxyvaleric acid. However, complementation of the *lvaC* knockout with plasmid expression resulted in poor growth compared to the other knockout strains (Figure 7.2). Notably, ACAD10 expression was able to restore growth on levulinic acid and 4-hydroxyvalerate in the *lvaA* and *lvaB* knockout strains. The D463A mutant was not able to confer growth, demonstrating that ACAD10 was restoring growth through kinase activity. The D48A mutant also grew in these conditions, so it is unlikely that the hydrolase domain is involved levulinic acid catabolism in these experiments. ACAD11 did not restore growth on levulinic acid and 4-hydroxyvalerate in any knockout strain, and neither ACAD restored growth in the *lvaC* knockout strains, suggesting that the ACAD domains are not functional in *P. putida*.

P. putida can natively catabolize 4-hydroxoctanoic acid using a pathway independent of the *lva* genes, so minimal agar with this carbon source did not create a growth selection. However, we observed differences in growth among the strains expressing the ACAD constructs. ACAD10 constructs with an active kinase domain led to a noticeable growth defect on 4-hydroxyoctanoic acid. ACAD10 is likely catalyzing the formation of 4-phosphooctanoyl-CoA, and the observed growth defect could be from *P. putida* being unable to catabolize this product. Expression of any ACAD11 construct prevented growth, so the growth defect is not due to enzymatic activity and suggests that

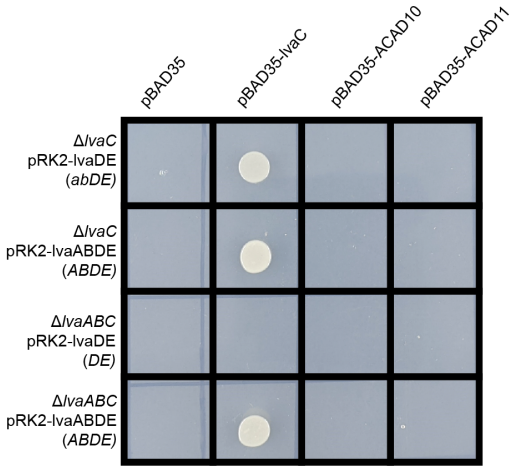


Figure 7.3: **Complementation of *lvaC* knockout strains with secondary expression plasmids for *lva* operon when growing on levulinic acid.** The labels in the left panel indicate the gene(s) knocked out in the host strain and the genes from the *lva* operon expressed on the secondary plasmid. The letters in parentheses list the genes expressed either chromosomally or on the plasmid. Lower case - the genes are expressed on the chromosome only. Upper case - the genes are expressed on the secondary plasmid. The pBAD35-based expression vectors are described in the top panel. An equivalent agar plate with glucose as the carbon source showed growth from all strains.

ACAD11 expression is not leading to functional or properly folded enzyme. Considering that soluble ACAD11 was purified from *E. coli* production cultures, growth-rescue experiments with an *E. coli* host may be more likely to show kinase activity from ACAD11. A derivative strain of *E. coli* LS5218 is able to grow on levulinic acid when expressing the *lva* operon and could be an alternative host to *P. putida* for complementation experiments [476]. These results demonstrate that the kinase domain of ACAD10 is active in *P. putida*, and it is confirmed to be active on 4-hydroxyvalerate and potentially 4-hydroxyoctanoate.

So far, the ACAD domains of both mammalian enzymes do not appear functional. However, The relatively poor complementation of the *lvaC* knockout by the gene itself suggested that the lack of growth could be explained by something other than lack of activity. The *lva* operon has a gene upstream of *lvaA* encoding the transcriptional regulator, LvaR. This protein represses *lva* expression, and appears to be antagonized by 4-phosphovaleryl-CoA [488]. The *lva* operon may have evolved so that this metabolite accumulates so that gene expression is induced, but overexpressing LvaC or an equivalent enzyme may increase the rate of consumption, leading to poor expression of the operon. To account for this possibility, we constructed secondary expression plasmids containing either *lvaABDE* or *lvaDE*. We introduced these plasmids along with the ACAD10 and ACAD11 expression vectors to *P. putida* Δ*lvaC* and tested growth of the resulting strains on minimal agar with levulinic acid. Complementation of the *lvaC* knockout by pBAD35-lvaC was greatly improved, demonstrating that the *lva* operon was poorly expressed in earlier experiments (Figure 7.3). Unfortuantely, neither ACAD10 or ACAD11 expression rescued growth, showing that the ACAD domains are not functional in *P. putida*. Mammalian ACADs typically use electron-transferring flavoproteins (ETF) as an electron carrier [489, 490]. *P. putida* likely lacked a compatible electron carrier, preventing ACAD activity.

7.5 Conclusions

Although the results are limited, there is some evidence that ACAD10 and ACAD11 are capable of processing 4-hydroxy-*n*-acyl-CoAs in a manner similar to levulinic acid catabolism in *P. putida*. Through growth experiments with *P. putida* and *in vitro* assays, we demonstrated that ACAD10 and ACAD11 phosphorylate the hydroxy group in 4-hydrxyvaleryl-CoA. More experiments are necessary to demonstrate isomerization to 3-hydroxy-*n*-acyl-CoAs by the ACAD domains. Future investigations should focus on the role of electron carriers in ACAD activity. Interactions between ACADs and ETFs are essential for tying fatty acid metabolism to oxidative phosphorylation in mammalian cells, which likely limits the utility of bacterial hosts for mammalian ACADs.

Chapter 8

Conclusions & Future Directions

Authors and Contributors:

- **Taylor B. Cook** Wrote text, designed and implemented experiments.
- **Aditya R. Aliani** Generated and analyzed Tn5 libraries in *P. putida*
- **Dr. Brian F. Pfleger** Provided guidance, edited, and reviewed the manuscript.

8.1 Summary of Thesis Research

Drug development in the future will continue to depend on natural product discovery, but new discovery platforms are needed to keep up with modern demands. These new methods are primarily enabled by increased access to genetic information from next-generation sequencing technologies. In order to fulfill the metabolic potential of this relatively new resource, researchers will rely on synthetic biology and metabolic engineering. Modern drug discovery platforms will involve engineering microbial hosts, pathways, and biosynthetic enzymes to synthesize novel natural products. Throughout this thesis, we explored all three of these aspects for engineering polyketide and nonribosomal peptide biosynthesis. Specifically, methods for production host engineering, metabolic pathway optimization, and modification of enzyme substrate specificity are evaluated for their potential to generate compounds of interest. In the examples discussed below, we aimed to standardize engineering techniques with well-characterized synthetic biology tools and with high-throughput selection schemes.

Humans have a long history of manipulating biological systems for their benefit. This natural ingenuity eventually transformed into the fields of medicinal chemistry and bioprocess engineering, which were instrumental in drug development and distribution in the twentieth century. This approach relied on the discovery and optimization of existing compounds and production strains. In **Chapter 1**, we introduced strategies for drug discovery based on molecular biology and compared rational and evolutionary techniques within the fields of synthetic biology and metabolic engineering. Understanding and manipulating molecular biology processes is essential for constructing production strains. When a system of interest is well-characterized, rational methods and systems biology are powerful tools for designing and testing variants. For poorly understood systems, laboratory evolution allows the processing of large libraries and isolation of variants that meet the design target. We conclude this chapter by explaining the synthetic biology and metabolic engineering challenges inherent to natural product biosynthesis.

The major challenges of engineering natural product biosynthesis, particularly for polyketides and nonribosomal peptides, lie in the lack of *a priori* knowledge of production

hosts and biosynthetic pathways. In **Chapter 2**, we outline attempts to circumvent the disadvantages of native hosts by engineering heterologous hosts to produce compounds of interest. Ideal heterologous hosts have a collection of well-characterized genetic tools, and we highlight efforts to domesticate *P. putida* as a metabolic engineering chassis. Heterologous hosts enable the expression and elucidation of uncharacterized biosynthetic gene clusters (BGCs). Manipulating natural product biosynthesis through protein engineering also requires expression hosts with genetic tools, so heterologous hosts are often the ideal host for these projects. This review also demonstrated that the application of cutting-edge genetic tools and metabolic engineering methods towards natural product biosynthesis has been limited.

Next, we presented several genetic tools that were generated for use in *P. putida* KT2440, a popular chassis strain for bioprocesses (**Chapter 3**). Several inducible promoter systems were compared for heterologous gene expression, and we discovered that *lacI* expression needed to be optimized to improve expression from *lac* promoters. Copy numbers for three common broad-host range origins were determined, with considerable implications on future metabolic engineering projects. Two of the three origins were found to have high copy numbers, which led to a growth defect in *P. putida* cultures. Lastly, we implemented a Cas9-assisted homologous recombination method for creating genomic deletions in *P. putida*, which out-performed other genome editing methods that the lab used previously.

The primary motivation for characterizing genetic tools for *P. putida* was to apply them towards engineering secondary metabolism. In **Chapter 4**, we expanded our genome editing tools for chromosomally integrating large BGCs, which facilitated the heterologous production of two natural products, prodigiosin and glidobactin A. Promoter and RBS optimization was used to improve expression of the heterologous BGCs, resulting in the highest prodigiosin titer of 1.1 g/L. We found that media formulations that elicit a strong response of carbon-catabolite repression to be worse for heterologous production. Through chromosomal modifications, we further increased glidobactin A titers by overexpressing key accessory proteins and deleting a competing pathway. MLP ex-

pression was crucial for improving production of glidobactin A from the MLP-dependent pathway from *S. brevitalea*. Providing a heterologous PPTase also improved production for both glidobactin BGCs. Lastly, we investigated the influence of PHA biosynthesis as a competing pathway in production strains. Deleting the PHA synthases slightly improved heterologous production in glycerol-based media if a regulatory protein was also deleted. These mutations had a greater effect on glidobactin A production and increased the highest titer to 470 mg/L glidobactin A.

An intriguing avenue for engineering natural product biosynthesis is through modifying substrate specificity in key biosynthetic enzymes. In **Chapter 5**, we explored various approaches to engineering substrate activation in NRPSs, the findings of which provided lessons for future engineering attempts. First, we summarized a collaboration with the Thomas Lab where we employed directed evolution to study the active-site residues in the adenylation domain of an NRPS. This work revealed that L-serine-specific adenylation domains in bacteria share a functional sequence space that extends beyond what has been observed in Nature, but these domains likely possess a strong bias towards their native substrate. Attempts to alter substrate specificity of other adenylation domains based on these findings were not successful. An alternative strategy for modifying adenylation domains is through subdomain swaps, and we briefly investigated this strategy for engineering the pyoverdine NRPS in *P. putida*. It was possible to construct functional chimeric enzymes with this approach, but they performed poorly compared to wild-type. Optimization of the subdomain boundary locations are needed to fully assess this strategy for NRPS engineering.

Using our experience in engineering *P. putida* as a heterologous host, we attempted to produce an extracellular enzyme (**Chapter 6**). Expression and secretion of this enzyme requires a specific chaperone that is natively expressed in the same operon as the enzyme. A panel of plasmid-based constructs were compared for heterologous expression in *P. putida*, and we found that the combination of the RK2 origin with the *trc* promoter resulted in the highest inducible gene expression. Interestingly, the *rha* promoter had extremely high basal expression in most cultivation conditions. Next, we showed that

P. putida secretes some native proteins into the media, and expression of the extracellular enzyme resulted in an increase in the number of proteins in culture supernatants. SDS-PAGE and Western blot analysis of production strains demonstrated that enzyme accumulated in the periplasm (in *P. putida* and in an alternative strain of *P. fluorescens*), suggesting that proteins in the Xcp pathway did not optimally recognize heterologous enzyme. In **Chapter 7**, we similarly express two mammalian ACADS in *P. putida* to determine if they could functionally replace steps in levulinic acid catabolism. Both ACADs contain a kinase domain with structural homology to bacterial phosphotransferases, and one enzyme, ACAD10, was able to complement a knockout of the phosphotranferase required for levulinic acid catabolism.

8.2 Future Directions

There are several avenues for building upon the work described herein. We discuss these potential projects in the following sections, with a focus on expected challenges and the impacts completion would have on their respective fields.

8.2.1 Discovery and Production of Cyanobacterial Natural Products

The most prudent application of synthetic biology towards producing novel drug candidates is the expression of uncharacterized BGCs in heterologous hosts. It is common for microorganisms to contain multiple BGCs with the potential to produce secondary metabolites, and it is not always possible to determine the corresponding metabolite for every BGC [34, 491]. Native hosts can be modified to activate BGCs of interest, but this strategy is not viable for hosts that are genetically incalcitrant or are difficult to cultivate in laboratory conditions. This is especially true for marine bacteria and cyanobacteria, which received attention recently for their secondary metabolite potential [416]. However, marine bacteria may not be amenable to robust growth in laboratory media and cyanobacteria typically have longer doubling times compared to heterotrophs,

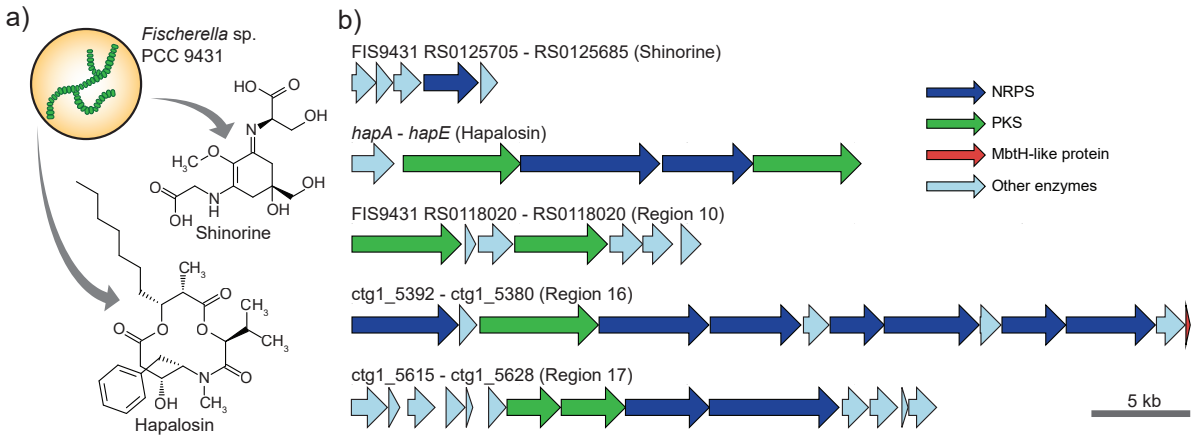


Figure 8.1: ***Fischerella* sp. PCC 9431 contains several BGCs encoding PKSs and NRPSs.** **a)** *Fischerella* sp. PCC 9431 natively produces the lipopeptide, hapalosin, and has the potential to produce the sunscreen, shinorine. **b)** There are five BGCs encoding PKSs and NRPSs, including those responsible for shinorine and hapalosin biosynthesis. Region numbers refer to the numbering of the BGCs from output from antiSMASH analysis in Figure E.1.

so elucidating BGCs from marine bacteria often requires heterologous hosts.

One such cyanobacterium is *Fischerella* sp. PCC 9431 (formerly known as *Hapalosiphon welwitschii* UTEX B1830), a member of filamentous cyanobacteria that contains several orphan BGCs [492]. Analysis of this strain's genome with antiSMASH [493] identified seventeen regions with the potential of producing secondary metabolites, five of which contained genes for PKSs and NRPSs (Figures 8.1, E.1). Two of these BGCs contain genes likely to produce the UV-absorbing compound, shinorine, and hapalosin, a drug candidate for chemotherapies [87, 278]. Shinorine biosynthesis involves a single NRPS module and hapalosin is synthesized by a hybrid PKS/NRPS (Figure 8.1b).

The BGCs illustrated in Figure 8.1b have several features that are relevant to findings from Chapter 4. The largest BGC is approximately 42 kb, and all BGCs encode genes that appear to be expressed in the same direction in a single operon, so rapid construction of production strains through one-step integrations should be possible. Although, a previous attempt to express the *hap* BGC identified a likely terminator sequence between the first two genes of the BGC [87]. No regions encode a PPTase, similar to the two glidobactin BGCs discussed here. Sfp from *B. subtilis* has been shown to poorly activate carrier proteins from a *Fischerella* species compared to cyanobacterial PPTases [250], so alternative PPTases will likely need to be investigated for optimizing heterologous production. The largest BGC also encodes an MLP that is likely required for activity of

some of the seven adenylation domains in the pathway. This is the only MLP encoded in the genome of *Fischerella* sp. PCC 9431, so interactions between this MLP and other NRPSs in the native host can also be investigated.

This proposed work could lead to the discovery of novel natural products and would establish *P. putida* as an alternative host for producing cyanobacterial natural products. There is considerable interest in developing heterotrophic hosts to produce these compounds because their faster growth rates allow for higher productivity compared to cyanobacterial hosts. Researchers have achieved modest titers engineering both *E. coli* and *S. cerevisiae* as heterologous hosts for cyanobacterial natural products [494, 495]. The various NRPSs encoded among these BGCs will also provide valuable information on substrate specificity, as several of the encoded A domains do not have strong predictions for their preferred substrate.

8.2.2 Systematic evaluation of chromosomal loci for heterologous expression

Future metabolic engineering projects with *P. putida* will require chromosomal integrations due to issues with plasmid stability and plasmid-induced growth defects. Multiple techniques have been used to integrate genetic constructs into the *P. putida* chromosome, and each one enables a different level of control over the location. Tn5 transposition integrates DNA into random positions [230], Tn7 transposition integrates directly downstream of the conserved gene *glmS* [383], and homologous recombination allows the user to choose the location by designing regions of homology in the integration construct [231]. Recently, Chaves and colleagues showed that heterologous expression can vary at different chromosomal loci by comparing seven integration sites throughout the *P. putida* chromosome [496]. However, due to the diversity in integration techniques available for *P. putida*, researchers have often used locations that were not evaluated in this study. A higher-resolution analysis of heterologous expression across the chromosome is needed to evaluate common integration loci and determine if more optimal loci exist.

Tn5 transposition is a powerful tool for generating libraries where each mutant has

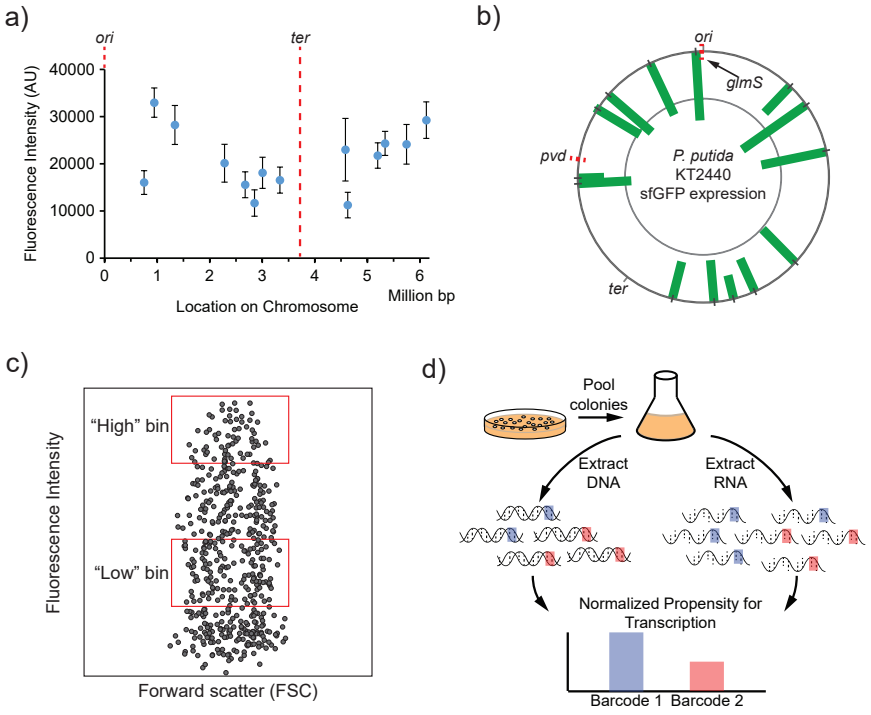


Figure 8.2: **Location of chromosomal integration affects heterologous expression in *P. putida*** **a)** A sample of *P. putida* Tn5 mutants expressing sfGFP shows variable levels of fluorescence, including locations for *glmS* and the *pvd* operon. **b)** Circular representation of the data presented in panel A. **c)** Fluorescence-activated cell sorting (FACS) with pooled libraries of strains containing barcoded expression constructs would provide more statistically significant data on heterologous protein production throughout the *P. putida* chromosome. **d)** Alternatively, NGS technologies can provide information on the effect of integration loci on transcription.

a transposon inserted into a unique location on the chromosome [22, 497]. We used this technique to create a modest library of *P. putida* strains expressing sfGFP at various chromosomal loci. Next, mutants were evaluated in 96-well plates, and 14 representative mutants were sequenced to determine their integration site (Figure 8.2a). We also observed variability in expression among the mutants, and heterologous expression was generally lower near the replication terminus (*ter*) of the chromosome. Sequencing individual mutants required a two-step PCR method that is not easily scaled for high-throughput analysis [384], so alternative methods are necessary to analyze larger libraries.

A solution for scaling up analysis of Tn5 mutants is to create pooled libraries with barcoded transposons, where each mutant has a short (approx. 15 bp) and unique sequence included with the expression construct. The barcodes can be linked to the chromosomal locus and to the transposon performance. For example, a pooled library can be analyzed by fluorescence-activate cell sorting (FACS) (Figure 8.2c). Individual cells can be binned

based on their fluorescence intensity, and next-gen sequencing would then provide the location of each mutant, link it to a barcode, and finally reveal the prevalence of each barcode in each bin. This method is advantageous because it captures how location can affect protein production, but the readout resolution is limited by the size and number of bins that are technically feasible.

Another pooled library approach that would allow for high-resolution evaluation of individual mutants is to use only next-gen sequencing technologies. DNA sequencing would again link each barcode to a chromosomal locus, but afterwards, DNA sequencing and RNA sequencing can be used in parallel to determine the amount of transcript relative to the amount of corresponding DNA (Figure 8.2d). This is a powerful technique that revealed several chromosomal features affecting heterologous expression in *E. coli* [497].

8.2.3 Modifying the Substrate Specificity of a Bi-specific NRPS Module

While there are some successes among NRPS engineering strategies, most still have considerable limitations. For example, domain substitution strategies are often limited by the substrate specificity of neighboring domains or require sequence identity between alternative domains. However, the A domain swap strategy reported by Calcott and co-workers appears to be the most versatile so far [126]. The authors discovered that the N-terminus of the C-A linker was the optimal N-terminal boundary for A domain substitutions (Figure 8.3a), resulting in chimeras that outperformed those constructed from C-A di-domain and A subdomain substitutions. Furthermore, they demonstrated that this strategy yielded functional chimeras from parent enzymes with relatively low homology.

This strategy could be applied towards the NRPS responsible for capreomycin biosynthesis, which involves a bi-specific module that activates both L-serine and L-alanine [437]. The L-serine derivative (capreomycin IA) has stronger antibiotic activity against *Mycobacterium tuberculosis*, so the activity of capreomycin extracts could be improved through A domain substitutions. There is an abundance of L-serine-specific NRPSs across

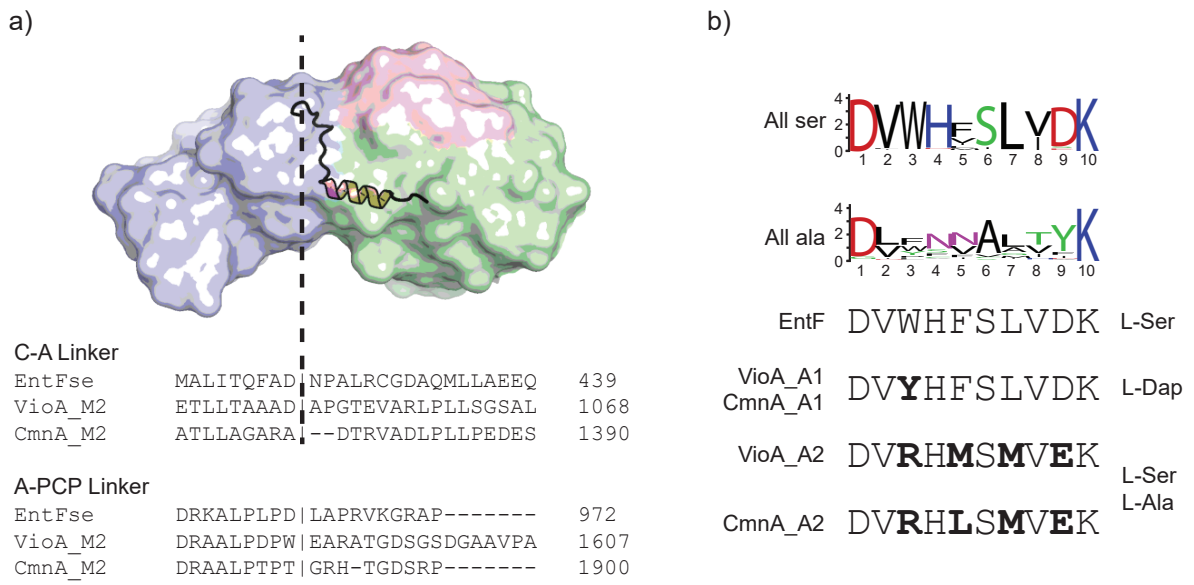


Figure 8.3: **An adenylation domain substitution would enable directed evolution of bi-specific domain in capreomycin biosynthesis. a)** The N-terminus of the C-A linker has been found to be the optimal boundary for generating functional NRPS chimeras. Protein structure shows the C (light blue) and A (light green) domains of EntF, with the A domain bound to its cognate MLP (pink). The A-PCP linker is not shown in the structure. **b)** Two homologous A domains in each of the capreomycin and viomycine NRPSs have similar specificity codes to L-serine-specific A domains. The first A domain (VioA_A1/CmnA_A1) recognizes L-2,3-diaminopropionate (L-Dap), and only differs in one residue in its specificity code compared to EntF. The second A domain (VioA_A2/CmnA_A2), recognizes L-serine and L-alanine, and has four different residues in its specificity code.

mutiple genera, including *Pseudomonas* and *Streptomyces*, that would provide candidate A domains for chimeric enzymes.

Alternatively, A domain substitutions could provide an opportunity to investigate substrate specificity of bi-specific A domains via directed evolution. As demonstrated in Chapter 5, enterobactin synthesis is a powerful selection platform for evolving L-serine-specific A domains. Capreomycin and viomycin biosynthesis contain homologous A domains that activate L-serine and L-alanine (VioA_A2/CmnA_A2), and a chimeric EntF containing either of these A domains would enable the isolation of mutants with increased specificity towards L-serine. The specificity codes of these domains resemble that of EntF and other L-serine codes rather than L-alanine codes, and they differ from the EntF code by four residues (Figure 8.3b). Initial directed evolution experiments would focus on these four residues, but mutations throughout the remainder of the A domain may also be necessary for evolving specificity.

Intriguingly, the first A domains in VioA and CmnA have a specificity code that

more closely resembles L-serine codes, with only one residue differing from the code of EntF (Figure 8.3b). Notably, no variants with the VioA_A1/CmnA_A1 specificity code appeared in the EntF libraries discussed in Chapter 5. Both of these domains activate L-2,3-diaminopropionate, which is structurally similar to L-serine, and have a small amount of activity on L-serine *in vitro*. Directed evolution on these A domains in EntF chimeras would lead to the discovery of mutations that confer altered specificity towards L-serine. Due to the similar specificity codes of these domains and L-serine-specific domains, these mutations would likely appear outside of the active site residues.

Regardless of which A domains are targeted, directed evolution in EntF chimeras is a versatile approach to isolate rare mutations that alter substrate specificity in NRPSs. The EntF scaffold limits directed evolution studies to L-serine activation, but this approach could be applied to other siderophore systems containing NRPSs, such as bacillibactin from *B. subtilis* and the pyoverdines from fluorescent Pseudomonads. Expanding this strategy to other siderophore systems will be crucial to improving our understanding of A domain specificity. While in the short term, A domain substitutions appear to be a versatile approach to engineering NRPSs to synthesize natural product derivatives, the directed evolution of important residues in A domains will provide fundamental knowledge on A domain activity.

8.3 Grand Challenges

Natural product discovery is in a new phase that is dependent on metagenomics and genome mining to find new BGCs. In order to access the novel drug candidates encoded in these sequences, researchers need tools for synthetic biology and metabolic engineering. Adoption of cutting-edge techniques within these fields towards natural product biosynthesis has been limited, primarily because most bacterial hosts capable of producing these compounds have relatively fewer genetic tools. Significant efforts have been made towards the domestication of non-model organisms, which can be applied towards native and heterologous hosts for producing secondary metabolites. Future projects should focus

on rapidly developing hosts for producing target compounds, as ideal hosts for products from various genera still have not been identified [498]. Researchers should also take advantage of the natural diversity available in microbiology and metagenomics by sampling various production hosts or biosynthetic pathways for accessing novel compounds [499, 309]. These approaches will be necessary to identify new classes of medically-relevant compounds, which are more impactful for providing drug candidates for clinical trials, similar to how the discovery of lipopeptides and arylomyicins has led to new classes of antibiotics in recent years [500, 407].

While enzyme engineering is a more academically fascinating approach to producing natural product derivatives, considerable progress is needed before an impact is made towards clinical trials and beyond. Modular enzymes often show product plasticity, and successful attempts to engineer substrate specificity are rare and have caveats that prevent wide-spread application. In the case of building chimeric enzymes, systematic approaches to determine the optimal domain boundaries are necessary to develop universal strategies for enzyme engineering [126]. Alternatively, mutating the specificity-conferring residues of gatekeeper domains will require laboratory evolution in order to isolate desired mutants. Bacterial PKSs and NRPSs have primarily evolved through domain and module substitutions, duplications and deletions [128, 458], so the natural diversity present in these enzymes will not provide the information required to guide these efforts. Switching substrate specificity through random mutations will likely require a significant restructuring of the active site, and this process may resemble "walking bac" towards a sequence that is similar to an ancestral generalist before re-evolving towards a desired specificity [501]. Lastly, a major limitation in attempts to engineer PKSs and NRPSs is that strategies for improving protein solubility and activity are virtually non-existant, and improving functional expression of these enzymes when scaling-up production cultures would improve bioprocesses centered around natural products.

Appendix A

Genetic Tools for Reliable Gene Expression and Recombineering in *P. putida*

A.1 Supplementary Figures and Tables

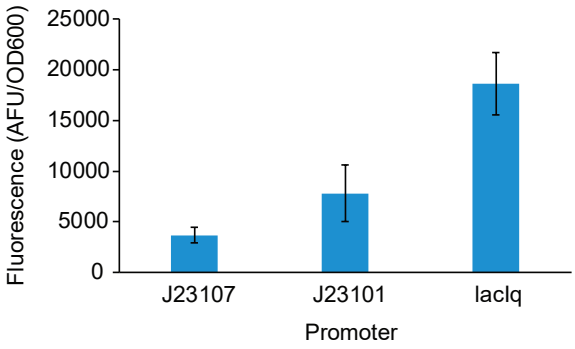


Figure A.1: **Expression of a fluorescent reporter protein from P_{lacIq} and promoters J23101 and J23107 from the Anderson promoter library on a BBR1-based plasmid. Fluorescence values are adjusted so that the empty vector control has a value of 0. Error bars represent one standard deviation.**

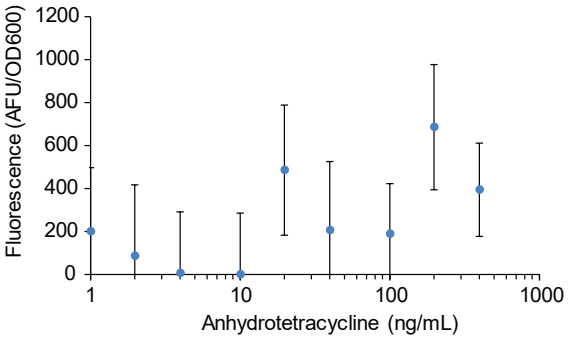


Figure A.2: **Induction curve for 2k promoter system in *P. putida* KT2440. Cultures were inoculated from overnight cultures and allowed to grow for 24 hours in the presence of varying concentrations of anhydrotetracycline.**

Table A.1: Copy numbers of broad host range plasmids in *E. coli* MG1655 and *P. putida* KT2440 in stationary phase. Values were calculated on a per-cell basis.

Origin	*E. coli* copy number	*P. putida* copy number
BBR1	3.0 ± 0.6	120 ± 20
BBR1-UP	24 ± 3	370 ± 70
BBR1-B5	3.2 ± 1.6	230 ± 60
RK2	1.7 ± 0.5	100 ± 20
RSF1010	4.4 ± 2.2	280 ± 50

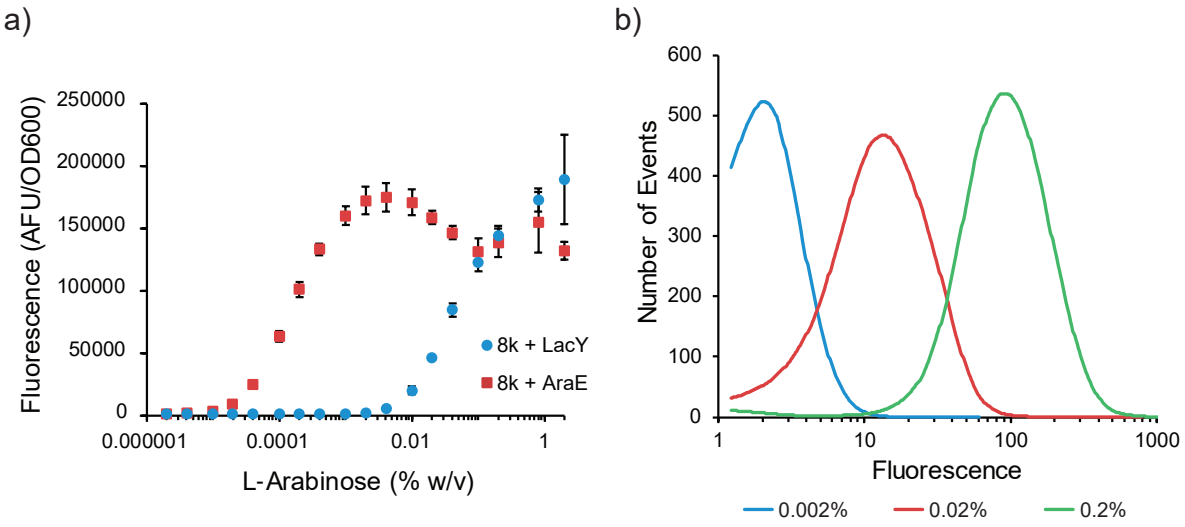


Figure A.3: **P_{araB} expression in the presence of AraE and flow cytometry analysis of P_{araB} expression in *P. putida*. a) Fluorescence of *P. putida* populations replicating pBb(B5)8k-sfGFP with and without *araE* expression from pRK2-AraE. Error bars represent one standard deviation. b) Histogram of *P. putida* populations replicating pBBR8k-GFPuv with different arabinose concentrations. Number of events is plotted as a running average with a window of 40 data points.**

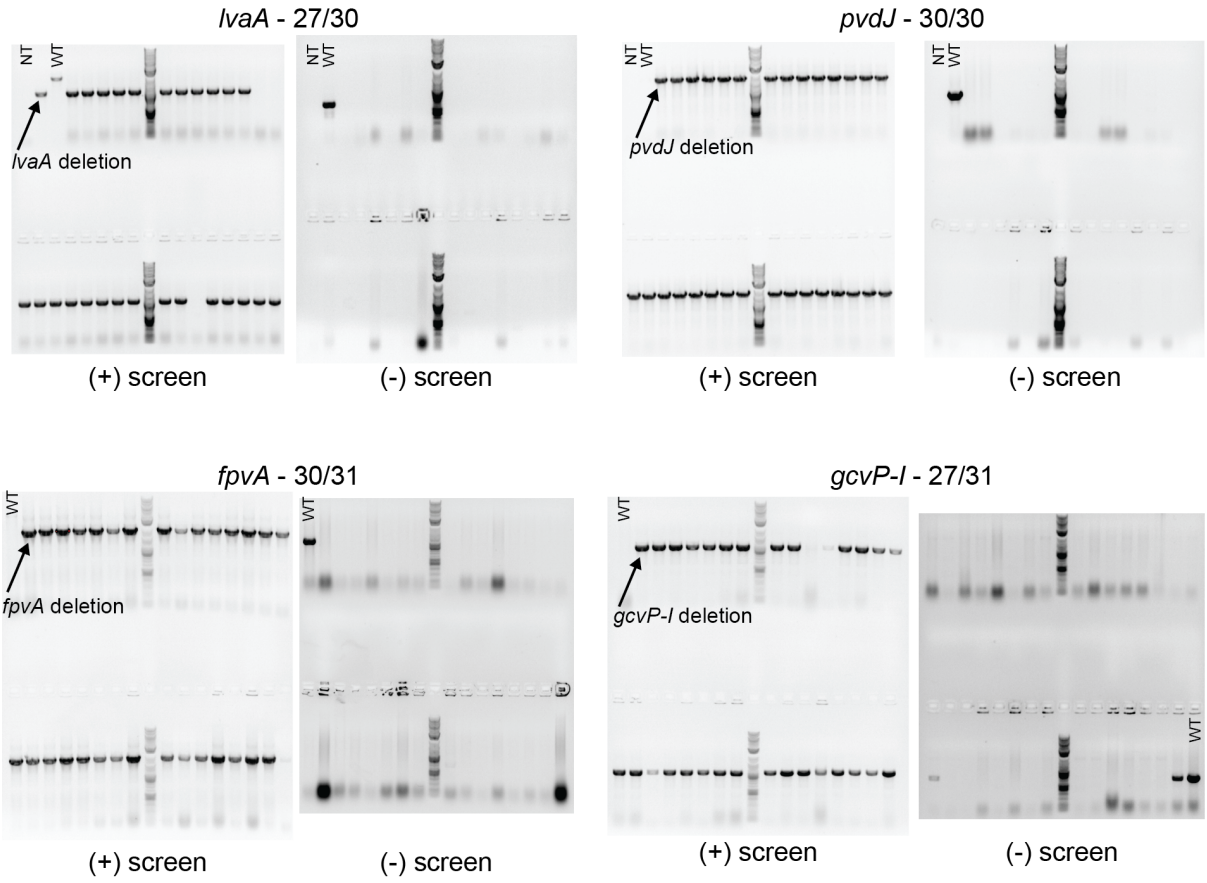


Figure A.4: **Colony PCR screens used to determine editing efficiency of four *P. putida* knockouts using the 2-step λRed/Cas9 recombineering protocol. Positive screens detected presence of the deletion. Negative screens detected presence of the wild type sequence. NT = no template control. WT = wild type control. All sample lanes not labelled as NT or WT are colonies picked from the indicated recombineering experiment.**

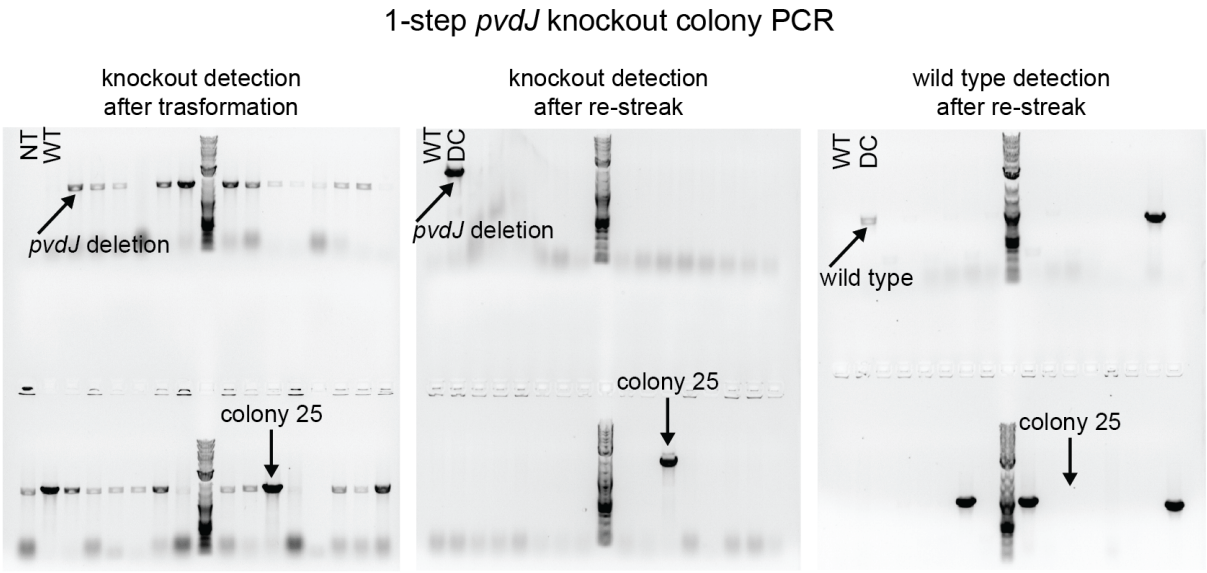


Figure A.5: **Colony PCR screens used to track loss of deletion when isolating colonies from transformants. The *pvdJ* knockout is used here as an example. Most transformants initially test positive for deletion, but after re-streaking, colonies lose the deletion genotype. Most colonies do not retain the wild type sequence. NT = no template control. WT = wild type control. DC = deletion control.**

Appendix B

Heterologous expression of biosynthetic gene clusters in *P. putida*

B.1 Supplementary Figures and Tables

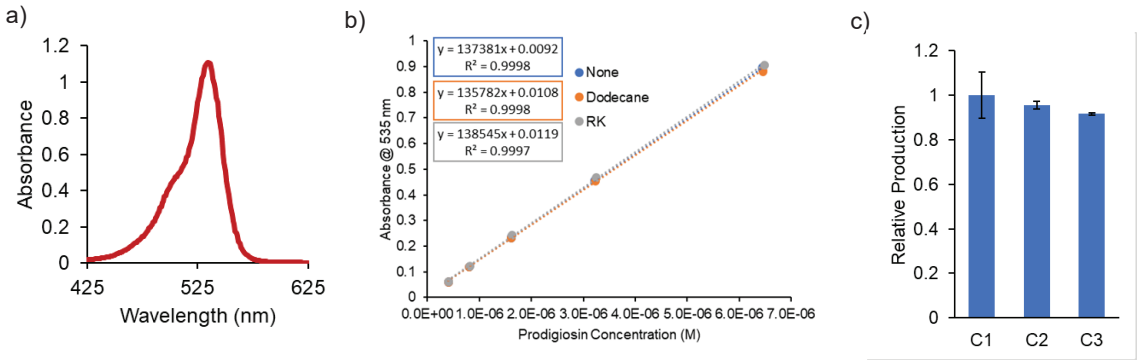


Figure B.1: **Verification of prodigiosin quantification. a)** Absorbance spectrum of prodigiosin extracted from *P. putida* in acidified ethanol. **b)** Determination of molar extinction coefficient with various "contaminants" in solvent. A prodigiosin standard was diluted in acidified ethanol containing either no contaminant, 1% (v/v) dodecane, or 1% (v/v) RK media. The addition of dodecane or RK media did not have a noticeable effect on prodigiosin quantification. The slopes of the standard curves correspond to the molar extinction coefficient of prodigiosin at 535 nm for each condition. Prodigiosin titers reported in this manuscript were calculated using the reported value of 139,800 M^{-1} cm^{-1}. **c)** Demonstration of consistent production titers from independently cloned prodigiosin producers (all equivalent to PIG01). The two-step protocol for integrating the pig cluster with P_{rha} was repeated three times and a verified colony from each attempt was isolated. Each strain had a similar production titers.

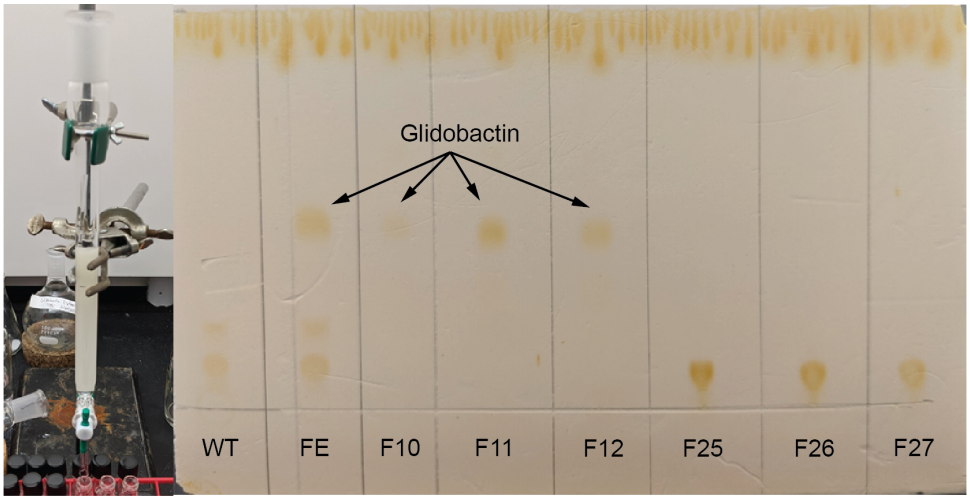


Figure B.2: **Silica gel chromatography for glidobactin purification and thin-layer chromatography (TLC) analysis.** Select fractions from chromatography purification of crude glidobactins were analyzed by TLC. Fractions F10-F12 had a white solid after the solvent was removed and were the fractions combined for creating the glidobactin standard. Fractions F25-F27 had a brown solid after removing the solvent and correspond to a native compound from *P. putida*. WT=wild-type *P. putida* KT2440. FE=full extraction from production strains before purification. F10-F27=fractions collected from eluent.

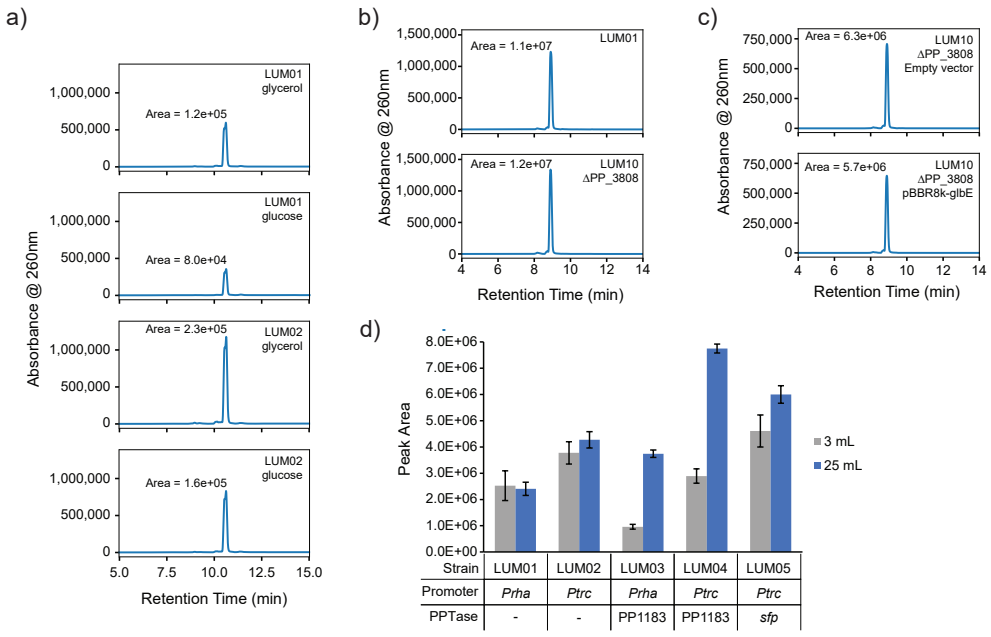


Figure B.3: **Additional effects on glidobactin production. a)** HPLC traces of glidobactin extracts from production cultures in RK glycerol vs. glucose. Extracts were created through acetone-ethyl acetate protocol. **b)** Changes in MLP expression do not effect glidobactin production in *P. putida* LUM01. Samples in panels **(b)** and **(c)** were analyzed on HPLC with a different mobile phase gradient (5% mobile phase B to 100% B over 15 minutes), hence the different retention time for glidobactin A. Extracts were created through acetone-ethyl acetate extraction. **d)** Effect of culture volume on glidobactin production from strains overexpressing a PPTase. Media was RK glycerol and 25-mL cultures were in 250-mL non-baffled flasks and 3-ml cultures were in tubes. Extracts were created through butanol extraction.

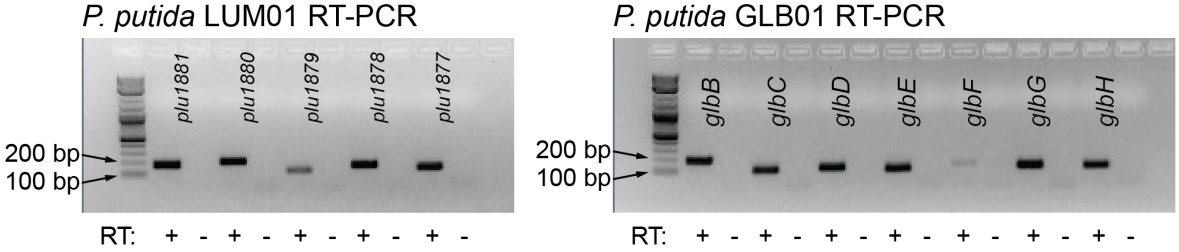


Figure B.4: **RT-PCR results from LUM01 and GLB01 confirming transcription through the entire operon for each gene cluster.** Strains were grown in typical growth conditions (250 mL flasks, RK glycerol) and cells were harvested for RNA extraction at 24 hours of growth.

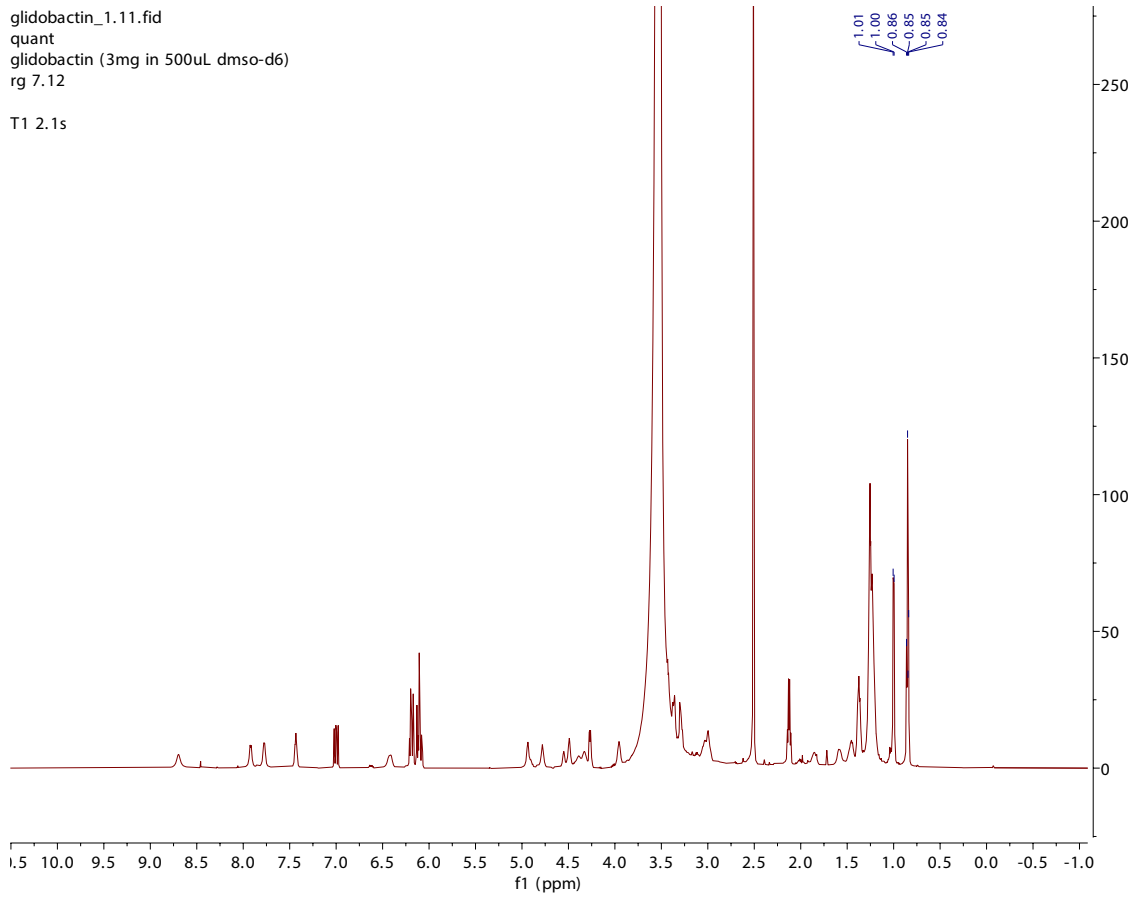


Figure B.5: **^{1}H-NMR spectrum for purified glidobactin A.** The doublet peak corresponding to the methyl group on the L-threonine residue (ppm=1.05) was used to quantify the concentration of glidobactin in the sample.

Table B.1: Genes and their encoded proteins from the prodigiosin BGC from *S. marcescens* ATCC274.

Gene	**Length (aa)**	**Protein accession number**	**Putative encoded protein**
pigA	385	WP_161545167	Acyl-CoA/acyl-ACP dehydrogenase
pigB	671	WP_161545166	FAD-dependent oxidoreductase
pigC	888	WP_161544455	Phosphoenolpyruvate synthase
pigD	904	Q5W251	Thiamine diphosphate dependent 3-acetyloctanal synthase
pigE	853	CAH55650	Aminotransferase
pigF	338	WP_004940222	Methyltransferase domain-containing protein
pigG	87	WP_004940224	Peptidyl carrier protein
pigH	648	CAH55653	Aminotransferase
pigI	490	CAH55654	L-prolyl-AMP ligase
pigJ	762	WP_060444127	Polyketide synthase
pigK	104	WP_004940233	RedY-like protein
pigL	215	WP_161544452	4'-phosphopantetheinyl transferase
pigM	352	WP_060444125	Hypothetical protein
pigN	242	WP_016928822	Oxidoreductase

Table B.2: Genes and their encoded proteins from the glidobactin A BGCs from *P. luminescens* TT01 *S. brevitalea* sp. nov. DSM7029.

Gene or ORF	**Length (aa)**	**Protein accession number**	**Amino acid percent identity**	**Putative encoded protein**
glbA	337	WP_053013498	N/A	LysR-type regulator
N/A	N/A	N/A		
glbB	286	AKJ29081	52.1%	Lysine 4-hydroxylase[a]
PLU_RS09380	272	Q7N5R2		
glbC	4,181	WP_053013497	52.7%	Hybrid non-ribosomal peptide synthetase and polyketide synthase
PLU_RS09375	4,160	CAE14173		
glbD	436	WP_083438241	50.7%	Major facilitator superfamily transporter
PLU_RS09370	421	WP_011146144		
glbE	73	WP_047194803	N/A	MbtH family protein
N/A	N/A	N/A		
glbF	1,083	WP_083438240	51.0%	Non-ribosomal peptide synthetase
PLU_RS09365	1,065	WP_011146143		
glbG	121	WP_047194802	47.9%	Nuclear transport factor 2 family protein
PLU_RS09360	119	WP_109791494		
glbH	472	AKJ29075	N/A	2-nitropropane dioxygenase
N/A	N/A	N/A		

[a]Annotated as hypothetical protein, but verified in Amatuni and Renata [502].

Table B.3: Editing efficiencies of chromosomal deletions and exogenous DNA insertions in *P. putida* KT2440

(a) Editing efficiency of chromosomal deletions via ssDNA recombineering in *P. putida* KT2440

Gene(s)	**Function**	**Location on chromosome**	**Length of deletion (bp)**	**Oligo**	**Recovery time**	**Correct clones**
sotB	Suggar efflux transporter	3,404,994	1,227	Lagging strand	2 hr	4/15 (27%)
				Lagging strand	3 hr	6/15 (40%)
				Lagging strand	6 hr	8/15 (53%)
				Leading strand	2 hr	0/15 (0%)
glpR	Transcriptional regulator	4,948,189	714	Lagging strand	3 hr	13/15 (86%)
phaG	Hydroxyacyl-ACP:CoA transacylase	4,574,340	855	Lagging strand	3 hr	0/15 (0%)
phaC1ZC2	PHA polymerases and depolymerase	482,351	4,287	Lagging strand	3 hr	0/31 (0%)
PP_3808	MbtH family protein	1,846,105	195	Lagging strand	3 hr	6/7 (86%)

(b) Editing efficiency of chromosomal deletions via circular dsDNA recombination in *P. putida* KT2440

Gene(s)	**Function**	**Location on chromosome**	**Length of deletion (bp)**	**Recovery time**	**Correct clones**
phaC1ZC2	PHA polymerases and depolymerase	482,351	4,287	3 hr	15/30 (50%)
pvdIJD	Non-ribosomal peptide synthetases	1,388,034	24,920	3 hr	7/7 (100%)

(c) Editing efficiency of BGC insertions via circular dsDNA recombination in *P. putida* KT2440

BGC	**Product**	**Location on chromosome**	**Length of insertion (bp)**	**Recovery time**	**Correct clones**
pig	Prodigiosin	1,388,034	23,247	3 hr	13/15 (86%)
lum	Glidobactin A	1,388,034	22,255	3 hr	15/15 (100%)
glb	Glidobactin A	1,388,034	22,367	3 hr	15/15 (100%)
ant	Anthraquinones	1,388,034	12,506	3 hr	15/15 (100%)

Appendix C

Characterization and manipulation of bacterial siderophore systems

C.1 Supplementary Figures and Tables

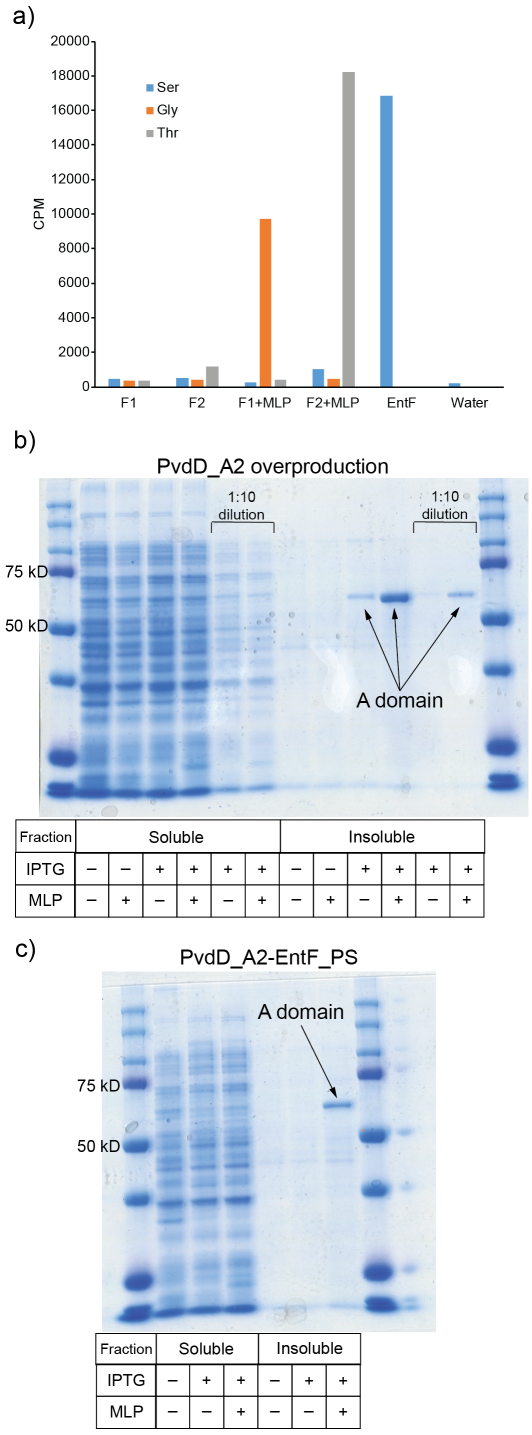


Figure C.1: **A domains in siderophore systems are dependent on cognate MLPs for expression and activity.** **a)** *In vitro* activity of the two A domains in DhbF, F1 and F2, both require addition of their cognate MLP in order to activate their respective amino acid substrate **b)** A domain in the second module of PvdD has higher production in *E. coli* Δ*ybdZ* when MLP (PA_2412) is co-expressed. **c)** PvdD_A2 chimera with EntF subdomain also has increased production when co-expressed with PA_2412 MLP.

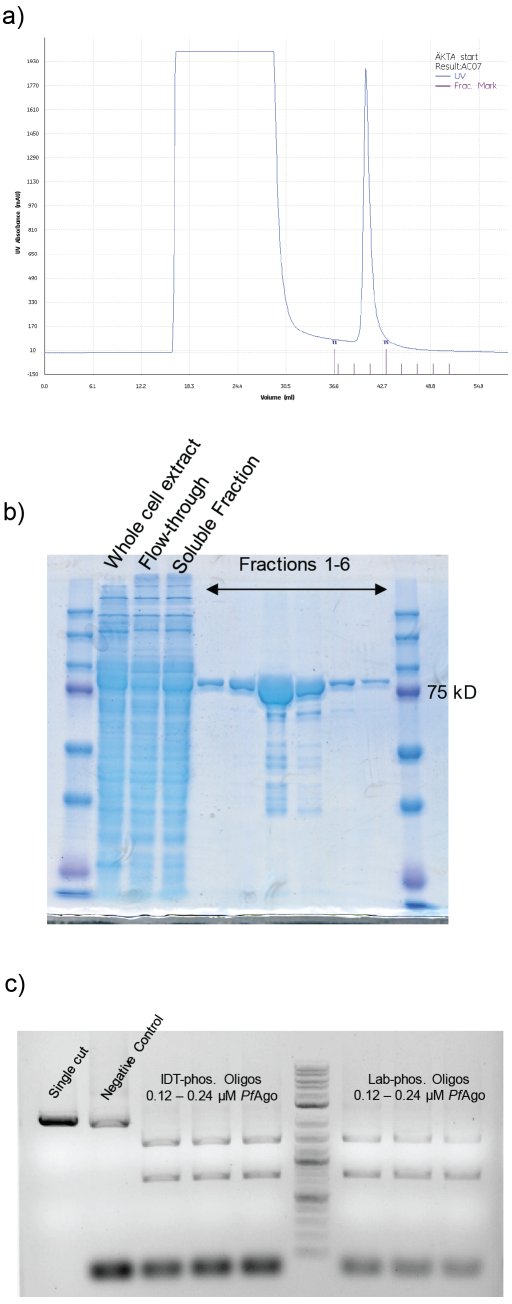


Figure C.2: ***Pf* Ago is a programmable restriction enzyme a)** Purification of Strep-tagged *Pf* Ago with Akta Start. **b)** SDS-PAGE analysis of fractions from purification protocol. *Pf* Ago ran slightly over 75 kD. **c)** Digestion of linear DNA with purified *Pf* Ago. Oligos were designed to target the single guide RNA CDS of pgRNAcm-pvdDShell2. Plasmid DNA was linearized by ApaLI digest. *Pf* Ago and oligo concentrations ranged from 0.06-0.24 μM. Negative control contained 0.24 μM *Pf* Ago/oligos and linearized pgRNAcm-fpvA.

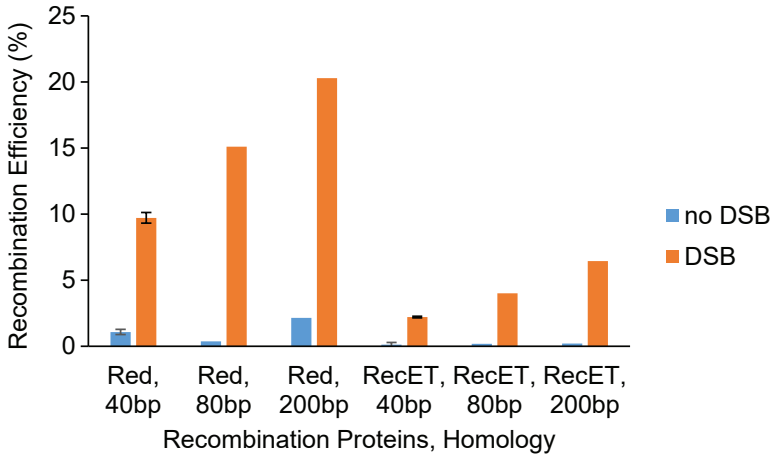


Figure C.3: **Cas9-assisted LCHR is more efficient with λRed compared to RecET enzymes.** Recombination efficiency of insertion of gentamycin resistance marker (GmR) into target plasmid. GmR was provided as a linear PCR product with 40, 80, or 200 bp of homology to target plasmid. Cloning strains were GB05-red (Red) or GB05-dir (RecET), which expressed the recombinases on the chromosome. Cas9 was expressed on the plasmid pSC101-Cas9. Double-stranded break (DSB) was induced by transforming pgRNAcm-RBS, which targeted the target plasmid. "No DSB" samples were transformed with pgRNAcm-4242, which did not target any plasmid or location in the chromosome. The number of transformants with the desired GmR insertion was determined by plating on LB agar with gentamycin. The recombination efficiency is defined as the sample CFU/mL divided by the CFU/mL of the corresponding "DSB" plated without gentamycin.

Appendix D

Production of a heterologous extracellular enzyme in *P. putida*

D.1 Supplementary Figures and Tables

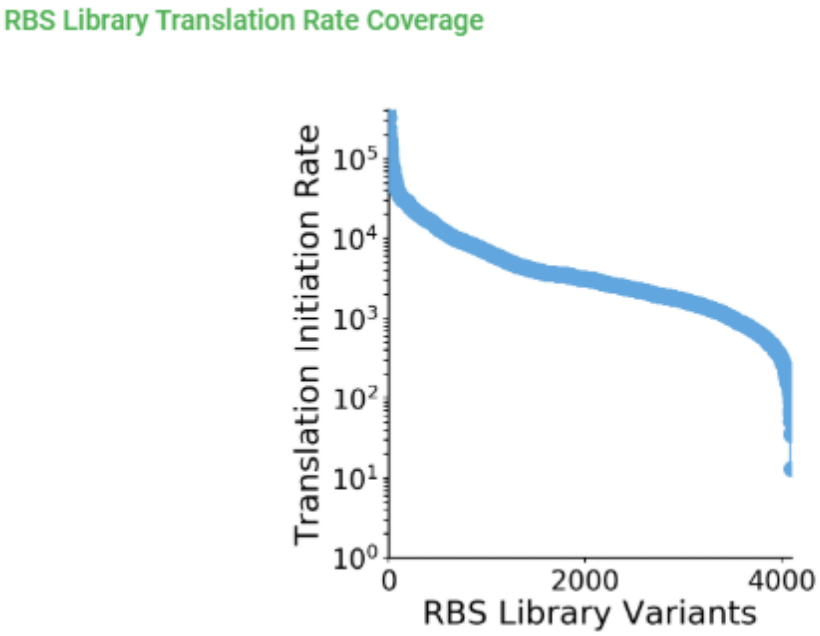


Figure D.1: **Range of predicted translation initiation rates for RBS library.** Library was calculated using the truncated peptide leader. The 5' untranslated region (UTR) was varied only in the basepairs corresponding to the Shine-Dalgarno sequence (6-bp sequence ending 8 bp upstream of the start codon). The library input was gaattcaaaagatcttttaagNNNNNNatatacat. The maximum translation rate was extraordinarily high. We typically see around 20,000 AU as the maximum translation rate. This suggests that the codon usage for the 5' end of the CDS is compatible wit that of *P. putida*. The Shine-Dalgarno sequences for RBS1 and RBS2 are GAGGAA and GAGAAC, respectively. RBS1 and RBS2 have a predicted translation initiation rate of 52,100 and 13,000 for the full enzyme leader construct, and 123,000 and 30,600 for the truncated leader construct.

Table D.1: Predicted secretion signals for various enzyme sequences using SignalP. Leader peptides are recognized by the Sec pathway, which will recruit proteins to the periplasm.

Protein	**SignalP Prediction**	**Sec Score**	**Tat Score**
Enzyme homolog	Sec	0.8075	0.1540
Native enzyme	Sec	0.7955	0.1547
Leaderless enzyme	None	0.0105	0.0005
Enzyme with truncated leader	Sec	0.973	0.185

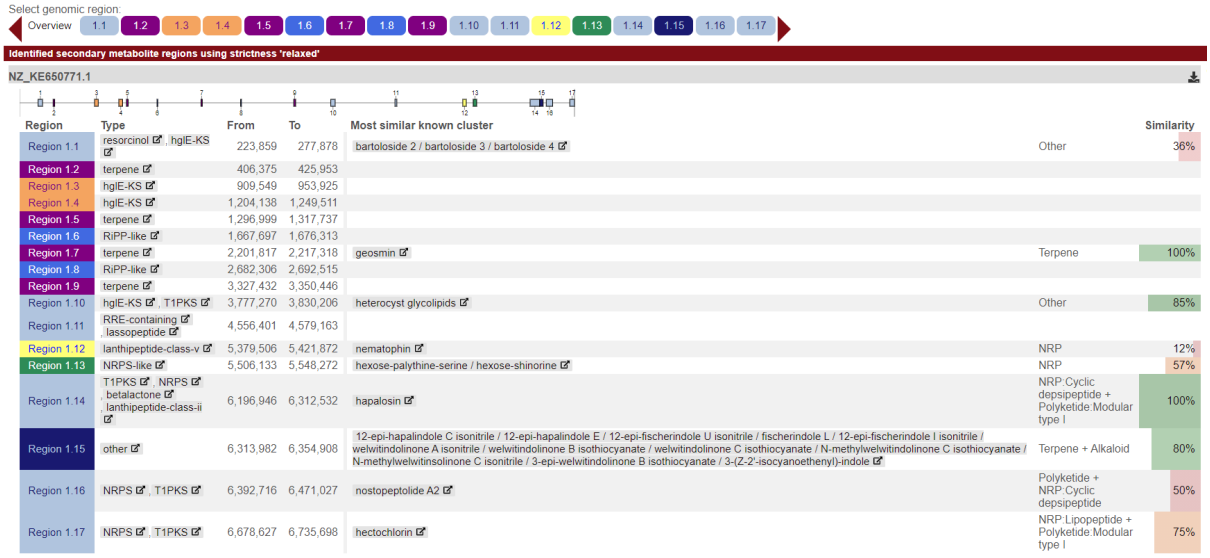
Table D.2: *P. fluorescens* strains from NRRL considered for heterologous expression. Strains listed in bold were investigated in this project. N/A – strain was isolated after time of publication for source on enzyme secretion.

Strain	**Secretes enzyme**	**Notes**
B-10	No	ATCC 948; no recombinant DNA examples in literature
B-11	Yes	ATCC 949; no recombinant DNA examples
B-1029	No	NCTC 1385; not much available literature
B-1603	Yes	ATCC 14150; no recombinant DNA examples
B-1609	Yes	Strain 12; No recombinant DNA examples
B-1796	Yes	Not much available literature
B-1800	Yes	Strain 35f
B-1801	No	Strain 41f
B-1802	No	Strain 43f
B-1964	Yes	Strain S-177
B-2322	Yes	Strain HCCB 29
B-2550	No	Strain A3.12; no recombinant DNA examples
B-3178	No	ATCC 17926; no recombinant DNA examples
B-4192	No	ATCC 25289; no recombinant DNA examples
B-4290	Yes	
B-29629	N/A	Recent isolate

Appendix E

Supplementary Information: Future Directions

E.1 Supplementary Figures and Tables



Region	Type	From	To	Most similar known cluster		Similarity
Region 1.1	resorcinol, hglE-KS	223,859	277,878	bartoloside 2 / bartoloside 3 / bartoloside 4	Other	36%
Region 1.2	terpene	406,375	425,953			
Region 1.3	hglE-KS	909,549	953,925			
Region 1.4	hglE-KS	1,204,138	1,249,511			
Region 1.5	terpene	1,296,999	1,317,737			
Region 1.6	RiPP-like	1,667,697	1,676,313			
Region 1.7	terpene	2,201,817	2,217,318	geosmin	Terpene	100%
Region 1.8	RiPP-like	2,682,306	2,692,515			
Region 1.9	terpene	3,327,432	3,350,446			
Region 1.10	hglE-KS, T1PKS	3,777,270	3,830,206	heterocyst glycolipids	Other	85%
Region 1.11	RRE-containing, lassopeptide	4,556,401	4,579,163			
Region 1.12	lanthipeptide-class-v	5,379,506	5,421,872	nematophin	NRP	12%
Region 1.13	NRPS-like	5,506,133	5,548,272	hexose-palythine-serine / hexose-shinorine	NRP	57%
Region 1.14	T1PKS, NRPS, betalactone, lanthipeptide-class-ii	6,196,946	6,312,532	hapalosin	NRP:Cyclic depsipeptide + Polyketide:Modular type I	100%
Region 1.15	other	6,313,982	6,354,908	12-epi-hapalindole C isonitrile / 12-epi-hapalindole E / 12-epi-fischerindole U isonitrile / fischerindole L / 12-epi-fischerindole I isonitrile / welwitindolinone A isonitrile / welwitindolinone B isothiocyanate / welwitindolinone C isothiocyanate / N-methylwelwitindolinone C isothiocyanate / N-methylwelwitinsolinone C isonitrile / 3-epi-welwitindolinone B isothiocyanate / 3-(Z-2'-isocyanoethenyl)-indole	Terpene + Alkaloid	80%
Region 1.16	NRPS, T1PKS	6,392,716	6,471,027	nostopeptolide A2	Polyketide + NRP:Cyclic depsipeptide	50%
Region 1.17	NRPS, T1PKS	6,678,627	6,735,698	hectochlorin	NRP:Lipopeptide + Polyketide:Modular type I	75%

Figure E.1: ***Fischerella* sp. PCC 9431 contains 17 total BGCs.** Other represented classes of secondary metabolites include terpenes, lanthipeptides, and other RiPPs.

References

[1] Luisa Alba-Lois and Claudia Segal-Kischinevzky. Yeast fermentation and the making of beer and wine. *Nature Education*, 3(9):17, 2010.

[2] Stanley A. Plotkin and Susan L. Plotkin. The development of vaccines: how the past led to the future. *Nature Reviews Microbiology*, 9(12):889–893, December 2011.

[3] Xuan-mei Huang, Zhi-jie Yang, Qing Xie, Zi-kang Zhang, Hua Zhang, and Junying Ma. Natural products for treating colorectal cancer: A mechanistic review. *Biomedicine & Pharmacotherapy*, 117:109142, September 2019.

[4] Atanas G Atanasov, Sergey B Zotchev, Verena M Dirsch, and Claudiu T Supuran. Natural products in drug discovery: advances and opportunities. *Nature Reviews Drug Discovery*, 20(3):200–216, March 2021.

[5] Björn Over, Stefan Wetzel, Christian Grütter, Yasushi Nakai, Steffen Renner, Daniel Rauh, and Herbert Waldmann. Natural-product-derived fragments for fragment-based ligand discovery. *Nature Chemistry*, 5(1):21–28, 2013.

[6] David J. Newman and Gordon M. Cragg. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *Journal of Natural Products*, 83(3):770–803, March 2020.

[7] Alexander Scriabine. Discovery and development of major drugs currently in use. In *Pharmaceutical innovation: revolutionizing human health*, volume 2, pages 148–270. Chemical Heritage Foundation, 1999.

[8] Robert Gaynes. The Discovery of Penicillin—New Insights After More Than 75 Years of Clinical Use. *Emerging Infectious Diseases*, 23(5):849–853, May 2017.

[9] Kim Lewis. Recover the lost art of drug discovery. *Nature*, 485(7399):439–440, May 2012.

[10] Richard E Nelson, Kelly M Hatfield, Hannah Wolford, Matthew H Samore, R Douglas Scott, Sujan C Reddy, Babatunde Olubajo, Prabasaj Paul, John A Jernigan, and James Baggs. National Estimates of Healthcare Costs Associated With Multidrug-Resistant Bacterial Infections Among Hospitalized Patients in the United States. *Clinical Infectious Diseases*, 72(Supplement_1):S17–S26, January 2021.

[11] Víctor de Lorenzo. From the selfish gene to selfish metabolism : Revisiting the central dogma. *BioEssays*, 36(3):226–235, March 2014.

[12] Juan Nogales, Joshua Mueller, Steinn Gudmundsson, Francisco J. Canalejo, Estrella Duque, Jonathan Monk, Adam M. Feist, Juan Luis Ramos, Wei Niu, and Bernhard O. Palsson. High-quality genome-scale metabolic modelling of Pseudomonas putida highlights its broad metabolic capabilities. *Environmental Microbiology*, 22(1):255–269, January 2020.

[13] Paul Opgenorth, Zak Costello, Takuya Okada, Garima Goyal, Yan Chen, Jennifer Gin, Veronica Benites, Markus de Raad, Trent R Northen, Kai Deng, Samuel Deutsch, Edward E. K. Baidoo, Christopher J. Petzold, Nathan J. Hillson, Hector Garcia Martin, and Harry R Beller. Lessons from Two Design–Build–Test–Learn Cycles of Dodecanol Production in Escherichia coli Aided by Machine Learning. *ACS Synthetic Biology*, 8(6):1337–1351, June 2019.

[14] Cameron Cotten and Jennifer L. Reed. Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering. *Biotechnology Journal*, 8(5):595–604, 2013.

[15] Deepanwita Banerjee, Thomas Eng, Andrew K Lau, Yusuke Sasaki, Brenda Wang, Yan Chen, Jan-philip Prahl, Vasanth R Singan, Robin A Herbert, Yuzhong Liu, Deepti Tanjore, Christopher J Petzold, Jay D Keasling, and Aindrila Mukhopadhyay. Genome-scale metabolic rewiring improves titers rates and yields of the non-native product indigoidine at scale. *Nature Communications*, 11(1):5385, December 2020.

[16] Grégory Boël, Reka Letso, Helen Neely, W Nicholson Price, Kam-ho Wong, Min Su, Jon D Luff, Mayank Valecha, John K Everett, Thomas B Acton, Rong Xiao, Gaetano T Montelione, Daniel P Aalberts, and John F Hunt. Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature*, 529(7586):358–363, 2016.

[17] Gina C. Gordon, Travis C. Korosh, Jeffrey C. Cameron, Andrew L. Markley, Matthew B. Begemann, and Brian F. Pfleger. CRISPR interference as a titratable, trans-acting regulatory tool for metabolic engineering in the cyanobacterium Synechococcus sp. strain PCC 7002. *Metabolic Engineering*, 38:170–179, 2016.

[18] Andrew D Garst, Marcelo C Bassalo, Gur Pines, Sean A Lynch, Andrea L Halweg-edwards, Rongming Liu, Liya Liang, Zhiwen Wang, Ramsey Zeitoun, William G Alexander, and Ryan T Gill. Genome-wide mapping of mutations at single-nucleotide resolution for protein , metabolic and genome engineering. *Nature Biotechnology*, 35(1):48–55, 2017.

[19] Christopher R Mehrer, Matthew R Incha, Mark C Politz, and F P Brian. Anaerobic production of medium-chain fatty alcohols via a β-reduction pathway. *Metabolic Engineering*, 48(April):63–71, 2018.

[20] Bradley S. Evans, Yunqiu Chen, William W. Metcalf, Huimin Zhao, and Neil L. Kelleher. Directed evolution of the nonribosomal peptide synthetase AdmK generates new andrimid derivatives in vivo. *Chemistry and Biology*, 18(5):601–607, 2011.

[21] Troy E Sandberg, Michael J Salazar, Liam L Weng, Bernhard O Palsson, and Adam M. Feist. The emergence of adaptive laboratory evolution as an efficient tool for biological discovery and industrial biotechnology. *Metabolic Engineering*, 56:1–16, December 2019.

[22] Kelly M Wetmore, Morgan N Price, Robert J Waters, Jacob S Lamson, Jennifer He, Cindi A Hoover, Matthew J Blow, James Bristow, Gareth Butland, Adam P Arkin, and Adam Deutschbauer. Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons. *mBio*, 6(3):1–15, May 2015.

[23] Elizabeth M.J. Gillam, Janine N Copp, and David F Ackerley. *Directed Evolution Library Creation: Methods and Protocols*. Humana Press, 2nd edition, 2014.

[24] Huimin Zhao. Directed evolution of novel protein functions. *Biotechnology and Bioengineering*, 98(2):313–317, October 2007.

[25] Taylor B. Cook and Brian F. Pfleger. Leveraging synthetic biology for producing bioactive polyketides and non-ribosomal peptides in bacterial heterologous hosts. *MedChemComm*, 10(5):668–681, 2019.

[26] Gordon M. Cragg and David J. Newman. Natural products: A continuing source of novel drug leads. *Biochimica et Biophysica Acta - General Subjects*, 1830(6):3670–3695, 2013.

[27] Walter Sneader. *Drug prototypes and their exploitation*. J. Wiley and sons, 1996.

[28] János Bérdy. Thoughts and facts about antibiotics: Where we are now and where we are heading. *Journal of Antibiotics*, 65(8):385–395, 2012.

[29] Timothy A Wencewicz. New antibiotics from Nature's chemical inventory. *Bioorganic & Medicinal Chemistry*, 24:6227–6252, 2016.

[30] Gordon M. Cragg, Paul G. Grothaus, and David J. Newman. Impact of Natural Products on Developing New Anti-Cancer Agents. *Chemical Reviews*, 109(7):3012–3043, 2009.

[31] David H Sherman. The Lego-ization of polyketide biosynthesis. *Nature biotechnology*, 23(9):1083–1084, 2005.

[32] Chris A. Dejong, Gregory M. Chen, Haoxin Li, Chad W. Johnston, McLean R. Edwards, Philip N. Rees, Michael A. Skinnider, Andrew L.H. Webster, and Nathan A. Magarvey. Polyketide and nonribosomal peptide retro-biosynthesis and global gene cluster matching. *Nature Chemical Biology*, 12(12):1007–1014, 2016.

[33] Samay Pande and Christian Kost. Bacterial Unculturability and the Formation of Intercellular Metabolic Networks. *Trends in Microbiology*, 25(5):349–361, 2017.

[34] S. D. Bentley, K. F. Chater, A.-M. Cerdeño-Tárraga, G. L. Challis, N R Thomson, K D James, D E Harris, M A Quail, H. Kieser, D Harper, A Bateman, S Brown, G. Chandra, C. W. Chen, M Collins, A Cronin, A Fraser, A Goble, J Hidalgo, T Hornsby, S Howarth, C.-H. Huang, T. Kieser, L Larke, L Murphy,

K Oliver, S. O'Neil, E Rabbinowitsch, M.-A. Rajandream, K Rutherford, S Rutter, K Seeger, D Saunders, S Sharp, R. Squares, S Squares, K Taylor, T Warren, A. Wietzorrek, J Woodward, B G Barrell, J Parkhill, and D. A. Hopwood. Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature*, 417(6885):141–147, May 2002.

[35] Elizabeth A. Felnagle, Michelle R. Rondon, Andrew D. Berti, Heidi A. Crosby, and Michael G. Thomas. Identification of the biosynthetic gene cluster and an additional gene for resistance to the antituberculosis drug capreomycin. *Applied and Environmental Microbiology*, 73(13):4162–4170, 2007.

[36] Lei Li, Yawei Zhao, Lijun Ruan, Sheng Yang, Mei Ge, Weihong Jiang, and Yinhua Lu. A stepwise increase in pristinamycin II biosynthesis by Streptomyces pristinaespiralis through combinatorial metabolic engineering. *Metabolic Engineering*, 29:12–25, 2015.

[37] Yihua Chen, Evelyn Wendt-pienkowski, and Ben Shen. Identification and Utility of FdmR1 as a Streptomyces Antibiotic Regulatory Protein Activator for Fredericamycin Production in Streptomyces griseus ATCC 49344 and Heterologous Hosts. *Journal of Bacteriology*, 190(16):5587–5596, 2008.

[38] Karsten Temme, Dehua Zhao, and Christopher A Voigt. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *Proceedings of the National Academy of Sciences*, 109(18):2–7, 2012.

[39] Sarah J Kodumal, Kedar G Patel, Ralph Reid, Hugo G Menzella, Mark Welch, and Daniel V Santi. Total synthesis of long DNA sequences : Synthesis of a contiguous 32-kb polyketide synthase gene cluster. *Proceedings of the National Academy of Sciences*, 101(44):15573–15578, 2004.

[40] Bradley S Moore and Christian Hertweck. Biosynthesis and attachment of novel bacterial polyketide synthase starter units. *Natural product reports*, 19(1):70–99, 2002.

[41] Yolande A. Chan, Angela M. Podevels, Brian M. Kevany, and Michael G. Thomas. Biosynthesis of polyketide synthase extender units. *Nat. Prod. Rep.*, 26(1):90–114, 2009.

[42] Stephan A. Sieber and Mohamed A. Marahiel. Learning from Nature's Drug Factories: Nonribosomal Synthesis of Macrocyclic Peptides. *Journal of Bacteriology*, 185(24):7036–7043, 2003.

[43] David E. Cane and Christopher T. Walsh. The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. *Chemistry and Biology*, 6(12):319–325, 1999.

[44] Ben Shen. Polyketide biosynthesis beyond the type I, II and III polyketide synthase paradigms. *Current Opinion in Chemical Biology*, 7(2):285–295, 2003.

[45] Michael A. Fischbach and Christopher T. Walsh. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic machinery, and mechanisms. *Chemical Reviews*, 106(8):3468–3496, 2006.

[46] Sabine Meyer, Jan-christoph Kehr, Andi Mainz, Daniel Dehm, Daniel Petras, Roderich D Su, and Elke Dittmann. Biochemical Dissection of the Natural Diversification of Microcystin Provides Lessons for Synthetic Biology of NRPS. *Cell Chemical Biology*, 23:462–471, 2016.

[47] Sacha J. Pidot, Hui Hong, Torsten Seemann, Jessica L. Porter, Marcus J. Yip, Artem Men, Matthew Johnson, Peter Wilson, John K. Davies, Peter F. Leadlay, and Timothy P. Stinear. Deciphering the genetic basis for polyketide variation among mycobacteria producing mycolactones. *BMC Genomics*, 9(1):462, 2008.

[48] Marc G. Chevrette, Fabian Aicheler, Oliver Kohlbacher, Cameron R. Currie, and Marnix H. Medema. SANDPUMA: Ensemble predictions of nonribosomal peptide chemistry reveal biosynthetic diversity across Actinobacteria. *Bioinformatics*, 33(20):3202–3210, 2017.

[49] Eric J N Helfrich, Reiko Ueoka, Alon Dolev, Michael Rust, Roy A Meoded, Agneya Bhushan, Gianmaria Califano, Rodrigo Costa, Muriel Gugger, Christoph Steinbeck, Pablo Moreno, and Jörn Piel. Automated structure prediction of trans-acyltransferase polyketide synthase products. *Nature Chemical Biology*, 15(8):813–821, August 2019.

[50] R McDaniel, a Thamchaipenet, C Gustafsson, H Fu, M Betlach, and G Ashley. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proceedings of the National Academy of Sciences of the United States of America*, 96(5):1846–1851, 1999.

[51] J. Piel, C. Hertweck, P. R. Shipley, D. M. Hunt, M. S. Newman, and B. S. Moore. Cloning, sequencing and analysis of the enterocin biosynthesis gene cluster from the marine isolate 'Streptomyces maritimus': Evidence for the derailment of an aromatic polyketide synthase. *Chemistry and Biology*, 7(12):943–955, 2000.

[52] Daniel Mendez-Perez, Matthew B. Begemann, and Brian F. Pfleger. Modular Synthase-Encoding Gene Involved in α-Olefin Biosynthesis in Synechococcus sp. Strain PCC 7002. *Applied and Environmental Microbiology*, 77(12):4264–4267, June 2011.

[53] Luis E N Quadri, Paul H Weinreb, Ming Lei, Michiko M Nakano, Peter Zuber, and Christopher T Walsh. Characterization of Sfp, a Bacillus subtilis Phosphopantetheinyl Transferase for Peptidyl Carrier Protein Domains in Peptide Synthetases †. *Biochemistry*, 37(6):1585–1595, February 1998.

[54] Jörn Piel. Biosynthesis of polyketides by trans-AT polyketide synthases. *Natural Product Reports*, 27(7):996–1047, 2010.

[55] Chaitan Khosla, Yinyan Tang, Alice Y. Chen, Nathan A. Schnarr, and David E. Cane. Structure and Mechanism of the 6-Deoxyerythronolide B Synthase. *Annual Review of Biochemistry*, 76(1):195–221, June 2007.

[56] B. Dunn and C. Khosla. Mechanism and specificity of an acyltransferase domain from a modular polyketide synthase. *Food, Pharmaceutical and Bioengineering Division 2013 - Core Programming Area at the 2013 AIChE Annual Meeting: Global Challenges for Engineering a Sustainable Future*, page 388, 2013.

[57] Fa Zhang, Ting Shi, Huining Ji, Imtiaz Ali, Shuxin Huang, Zixin Deng, Qing Min, Linquan Bai, Yilei Zhao, and Jianting Zheng. Structural Insights into the Substrate Specificity of Acyltransferases from Salinomycin Polyketide Synthase. *Biochemistry*, 58(27):2978–2986, July 2019.

[58] Chaitan Khosla, Daniel Herschlag, David E Cane, and Christopher T Walsh. Assembly Line Polyketide Synthases: Mechanistic Insights and Unsolved Problems. *Biochemistry*, 53(18):2875–2883, May 2014.

[59] Xinqiang Xie, Ashish Garg, Adrian T. Keatinge-Clay, Chaitan Khosla, and David E. Cane. Epimerase and Reductase Activities of Polyketide Synthase Ketoreductase Domains Utilize the Same Conserved Tyrosine and Serine Residues. *Biochemistry*, 55(8):1179–1186, March 2016.

[60] Christian Bisang, Paul F. Long, Jesús Cortés, James Westcott, John Crosby, Anne Lise Matharu, Russell J. Cox, Thomas J. Simpson, James Staunton, and Peter F. Leadlay. A chain initiation factor common to both modular and aromatic polyketide synthases. *Nature*, 401(6752):502–505, 1999.

[61] Yuji Shinohara, Fumitaka Kudo, and Tadashi Eguchi. A Natural Protecting Group Strategy To Carry an Amino Acid Starter Unit in the Biosynthesis of Macrolactam Polyketide Antibiotics. *Journal of the American Chemical Society*, 133(45):18134–18137, November 2011.

[62] Eric J. N. Helfrich and Jörn Piel. Biosynthesis of polyketides by trans-AT polyketide synthases. *Natural Product Reports*, 33(2):231–316, 2016.

[63] Lauren B. Pickens and Yi Tang. Decoding and engineering tetracycline biosynthesis. *Metabolic Engineering*, 11(2):69–75, 2009.

[64] Frank Gross, Æ Nora Luniak, Æ Olena Perlova, and Nikolaos Gaitatzis Æ Holger Jenke-kodama. Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. *Archives of Microbiology*, 185:28–38, 2006.

[65] Yun Ding, Joey Paolo Ting, Jinsha Liu, Shams Al-Azzam, Priyanka Pandya, and Sepideh Afshar. Impact of non-proteinogenic amino acids in the discovery and development of peptide therapeutics. *Amino Acids*, 52(9):1207–1226, September 2020.

[66] Torsten Stachelhaus, Henning D. Mootz, and Mohamed A. Marahiel. The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry and Biology*, 6(8):493–505, 1999.

[67] Marc Röttig, Marnix H. Medema, Kai Blin, Tilmann Weber, Christian Rausch, and Oliver Kohlbacher. NRPSpredictor2 - A web server for predicting NRPS adenylation domain specificity. *Nucleic Acids Research*, 39(SUPPL. 2):1–6, 2011.

[68] Elizabeth A. Felnagle, John J. Barkei, Hyunjun Park, Angela M. Podevels, Matthew D. McMahon, Donald W. Drott, and Michael G. Thomas. MbtH-Like Proteins as Integral Components of Bacterial Nonribosomal Peptide Synthetases. *Biochemistry*, 49(41):8815–8817, October 2010.

[69] Shalini Saha and Steven E. Rokita. An Activator of an Adenylation Domain Revealed by Activity but Not Sequence Homology. *ChemBioChem*, 17(19):1818–1823, October 2016.

[70] Lusong Luo, Rahul M Kohli, Megumi Onishi, Uwe Linne, Mohamed A Marahiel, and Christopher T Walsh. Timing of Epimerization and Condensation Reactions in Nonribosomal Peptide Assembly Lines : Kinetic Analysis of Phenylalanine Activating Elongation Modules of Tyrocidine Synthetase B †. *Biochemistry*, 41:9184–9196, 2002.

[71] Carl J Balibar, Frédéric H. Vaillancourt, and Christopher T Walsh. Generation of D Amino Acid Residues in Assembly of Arthrofactin by Dual Condensation/Epimerization Domains. *Chemistry & Biology*, 12(11):1189–1200, November 2005.

[72] Jingru Li and Susan E. Jensen. Nonribosomal Biosynthesis of Fusaricidins by Paenibacillus polymyxa PKB1 Involves Direct Activation of a d-Amino Acid. *Chemistry & Biology*, 15(2):118–127, February 2008.

[73] Jacques Ravel and Pierre Cornelis. Genomics of pyoverdine-mediated iron uptake in pseudomonads. *Trends in Microbiology*, 11(5):195–200, May 2003.

[74] Lin Zhong, Xiaotong Diao, Na Zhang, Fengwei Li, Haibo Zhou, Hanna Chen, Xianping Bai, Xintong Ren, Youming Zhang, Dalei Wu, and Xiaoying Bian. Engineering and elucidation of the lipoinitiation process in nonribosomal peptide biosynthesis. *Nature Communications*, 12(1):296, December 2021.

[75] Dirk Schwarzer, Henning D. Mootz, Uwe Linne, and Mohamed A. Marahiel. Regeneration of misprimed nonribosomal peptide synthetases by type II thioesterases. *Proceedings of the National Academy of Sciences*, 99(22):14083–14088, October 2002.

[76] Matt J Jaremko, Tony D Davis, Joshua C Corpuz, and Michael D Burkart. Type II non-ribosomal peptide synthetase proteins: structure, mechanism, and protein–protein interactions. *Natural Product Reports*, 2019.

[77] Henning D. Mootz, Dirk Schwarzer, and Mohamed A. Marahiel. Ways of assembling complex natural products on modular nonribosomal peptide synthetases. *ChemBioChem*, 3(6):490–504, 2002.

[78] Hajo Kries, David L. Niquille, and Donald Hilvert. A Subdomain Swap Strategy for Reengineering Nonribosomal Peptides. *Chemistry & Biology*, 22(5):640–648, 2015.

[79] Hugo G Menzella, Ralph Reid, John R Carney, Sunil S Chandran, Sarah J Reisinger, Kedar G Patel, David A Hopwood, and Daniel V Santi. Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. *Nature Biotechnology*, 23(9):1171–1176, 2005.

[80] Hailong Wang, Zhen Li, Ruonan Jia, Yu Hou, Jia Yin, Xiaoying Bian, Aiying Li, Rolf Müller, A Francis Stewart, Jun Fu, and Youming Zhang. RecET direct cloning and Red$\alpha\beta$ recombineering of biosynthetic gene clusters, large operons or single genes for heterologous expression. *Nature Protocols*, 11(7):1175–1190, July 2016.

[81] Behnam Enghiad and Huimin Zhao. Programmable DNA-Guided Artificial Restriction Enzymes. *ACS Synthetic Biology*, 6(5):752–757, May 2017.

[82] Wenjun Jiang and Ting F. Zhu. Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nature Protocols*, 11(5):960–975, 2016.

[83] Hailong Wang, Zhen Li, Ruonan Jia, Jia Yin, Aiying Li, Liqiu Xia, Yulong Yin, Rolf Müller, Jun Fu, A Francis Stewart, and Youming Zhang. ExoCET : exonuclease in vitro assembly combined with RecET recombination for highly efficient direct DNA cloning from complex genomes. *Nucleic Acids Research*, 46(5):e28, 2018.

[84] Jun Fu, Xiaoying Bian, Shengbaio Hu, Hailong Wang, Fan Huang, Philipp M Seibert, Alberto Plaza, Liqiu Xia, Rolf Müller, a Francis Stewart, and Youming Zhang. Full-length RecE enhances linear-linear homologous recombination and facilitates direct cloning for bioprospecting. *Nature Biotechnology*, 30(5):440–446, May 2012.

[85] Vladimir Larionov, Natalya Kouprina, Joan Graves, X N Chen, Julie R Korenberg, and Michael A Resnick. Specific cloning of human DNA as yeast artificial chromosomes by transformation-associated recombination. *Proceedings of the National Academy of Sciences*, 93(January):491–496, 1996.

[86] Ernst Weber, Carola Engler, Ramona Gruetzner, Stefan Werner, and Sylvestre Marillonnet. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS ONE*, 6(2):e16765, 2011.

[87] Paul M. D'Agostino and Tobias A M Gulder. Direct Pathway Cloning Combined with Sequence- and Ligation-Independent Cloning for Fast Biosynthetic Gene Cluster Refactoring and Heterologous Expression. *ACS Synthetic Biology*, 7(7):1702–1708, July 2018.

[88] Zengyi Shao, Hua Zhao, and Huimin Zhao. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Research*, 37(2):e16, 2009.

[89] Jia Jia Zhang and Bradley S. Moore. Site-directed mutagenesis of large biosynthetic gene clusters via oligonucleotide recombineering and CRISPR/Cas9 targeting. *ACS Synthetic Biology*, page acssynbio.0c00265, June 2020.

[90] Kei Kudo, Takuya Hashimoto, Junko Hashimoto, Ikuko Kozone, Noritaka Kagaya, Reiko Ueoka, Takehiro Nishimura, Mamoru Komatsu, Hikaru Suenaga, Haruo Ikeda, and Kazuo Shin-ya. In vitro Cas9-assisted editing of modular polyketide synthase genes to produce desired natural product derivatives. *Nature Communications*, 11(1):2–4, 2020.

[91] Markiyan Oliynyk, Murray J.B. Brown, Jesus Cortes, James Staunton, and Peter F. Leadlay. A hybrid modular polyketide synthase obtained by domain swapping. *Chemistry and Biology*, 3(10):833–839, 1996.

[92] Anand Ranganathan, Máire Timoney, Matthew Bycroft, Jesús Cortés, Iain P. Thomas, Barrie Wilkinson, Laurenz Kellenberger, Ulf Hanefeld, Ian S. Galloway, James Staunton, and Peter F. Leadlay. Knowledge-based design of bimodular and

trimodular polyketide synthases based on domain and module swaps: A route to simple statin analogues. *Chemistry and Biology*, 6(10):731–741, 1999.

[93] Li Tang, Hong Fu, and Robert McDaniel. Formation of functional heterologous complexes using subunits from the picromycin, erythromycin and oleandomycin polyketide synthases. *Chemistry and Biology*, 7(2):77–84, 2000.

[94] Rajesh S. Gokhale, Stuart Y. Tsuji, David E. Cane, and Chaitan Khosla. Dissecting and exploiting intermodular communication in polyketide synthases. *Science*, 284(5413):482–485, 1999.

[95] Andrew F.A. Marsden, Barrie Wilkinson, Jesús Cortés, Nicholas J. Dunster, James Staunton, and Peter F. Leadlay. Engineering broader specificity into an antibiotic-producing polyketide synthase. *Science*, 279(5348):199–202, 1998.

[96] Philipp Krastel, Silvio Roggo, Markus Schirle, Nathan T. Ross, Francesca Perruccio, Peter Aspesi, Thomas Aust, Kathrin Buntin, David Estoppey, Brigitta Liechty, Felipa Mapa, Klaus Memmert, Howard Miller, Xuewen Pan, Ralph Riedl, Christian Thibaut, Jason Thomas, Trixie Wagner, Eric Weber, Xiaobing Xie, Esther K. Schmitt, and Dominic Hoepfner. Nannocystin A: An Elongation Factor 1 Inhibitor from Myxobacteria with Differential Anti-Cancer Properties. *Angewandte Chemie - International Edition*, 54(35):10149–10154, 2015.

[97] Edward Kalkreuter, Jared M. CroweTipton, Andrew N. Lowell, David H. Sherman, and Gavin J. Williams. Engineering the Substrate Specificity of a Modular Polyketide Synthase for Installation of Consecutive Non-Natural Extender Units. *Journal of the American Chemical Society*, 141(5):1961–1969, February 2019.

[98] William B. Porterfield, Nannalin Poenateetai, and Wenjun Zhang. Engineered Biosynthesis of Alkyne-Tagged Polyketides by Type I PKSs. *iScience*, 23(3):100938, 2020.

[99] Joseph A. Chemler, Ashootosh Tripathi, Douglas A. Hansen, Mark O'Neil-Johnson, Russell B. Williams, Courtney Starks, Sung Ryeol Park, and David H. Sherman. Evolution of Efficient Modular Polyketide Synthases by Homologous Recombination. *Journal of the American Chemical Society*, 137(33):10603–10609, 2015.

[100] Takeshi Miyazawa, Melissa Hirsch, Zhicheng Zhang, and Adrian T Keatinge-clay. An in vitro platform for engineering and harnessing modular polyketide synthases. *Nature Communications*, 11(80):1–7, 2020.

[101] D. L. Stassi, S. J. Kakavas, K. A. Reynolds, G. Gunawardana, S. Swanson, D. Zeidner, M. Jackson, H. Liu, A. Buko, and L. Katz. Ethyl-substituted erythromycin derivatives produced by directed metabolic engineering. *Proceedings of the National Academy of Sciences of the United States of America*, 95(13):7305–7309, 1998.

[102] Ralph Reid, Misty Piagentini, Eduardo Rodriguez, Gary Ashley, Nina Viswanathan, John Carney, Daniel V. Santi, C. Richard Hutchinson, and Robert McDaniel. A model of structure and catalysis for ketoreductase domains in modular polyketide synthases. *Biochemistry*, 42(1):72–79, 2003.

[103] Andrew Hagen, Sean Poust, Tristan De Rond, Jeffrey L. Fortman, Leonard Katz, Christopher J. Petzold, and Jay D. Keasling. Engineering a Polyketide Synthase for in Vitro Production of Adipic Acid. *ACS Synthetic Biology*, 5(1):21–27, 2016.

[104] Thibault Annaval, Cédric Paris, Peter F. Leadlay, Christophe Jacob, and Kira J. Weissman. Evaluating ketoreductase exchanges as a means of rationally altering polyketide stereochemistry. *ChemBioChem*, 16(9):1357–1364, 2015.

[105] Constance B. Bailey, Marjolein E. Pasman, and Adrian T. Keatinge-Clay. Substrate structure-activity relationships guide rational engineering of modular polyketide synthase ketoreductases. *Chemical Communications*, 52(4):792–795, 2016.

[106] Erin E. Drufva, Nolan R. Spengler, Elijah G. Hix, and Constance B. Bailey. Site-Directed Mutagenesis of Modular Polyketide Synthase Ketoreductase Domains for Altered Stereochemical Control. *ChemBioChem*, 22(7):1122–1150, April 2021.

[107] Xiaoan Ruan, Ana Pereda, Diane L. Stassi, David Zeidner, Richard G. Summers, Marianna Jackson, Annapur Shivakumar, Stephan Kakavas, Michael J. Staver, Stefano Donadio, and Leonard Katz. Acyltransferase domain substitutions in erythromycin polyketide synthase yield novel erythromycin derivatives. *Journal of Bacteriology*, 179(20):6416–6425, 1997.

[108] Marcus Hans, Andreas Hornung, Agnieszka Dziarnowski, David E. Cane, and Chaitan Khosla. Mechanistic analysis of acyl transferase domain exchange in polyketide synthase modules. *Journal of the American Chemical Society*, 125(18):5366–5374, 2003.

[109] Maja Klaus, Lynn Buyachuihan, and Martin Grininger. Ketosynthase Domain Constrains the Design of Polyketide Synthases. *ACS Chemical Biology*, 15(9):2422–2432, 2020.

[110] Mark C. Walker, Benjamin W. Thuronyi, Louise K. Charkoudian, Brian Lowry, Chaitan Khosla, and Michelle C. Y. Chang. Expanding the Fluorine Chemistry of Living Systems Using Engineered Polyketide Synthase Pathways. *Science*, 341(6150):1089–1094, September 2013.

[111] Zhixia Ye, Ewa M. Musiol, Tilmann Weber, and Gavin J. Williams. Reprogramming acyl carrier protein interactions of an acyl-CoA promiscuous trans-acyltransferase. *Chemistry and Biology*, 21(5):636–646, 2014.

[112] Samantha M. Carpenter and Gavin J. Williams. Extender Unit Promiscuity and Orthogonal Protein Interactions of an Aminomalonyl-ACP Utilizing Trans-Acyltransferase from Zwittermicin Biosynthesis. *ACS Chemical Biology*, 1(Figure 1):3361–3373, 2018.

[113] Christopher D. Reeves, Sumati Murli, Gary W. Ashley, Misty Piagentini, C. Richard Hutchinson, and Robert McDaniel. Alteration of the Substrate Specificity of a Modular Polyketide Synthase Acyltransferase Domain through Site-Specific Mutations. *Biochemistry*, 40(51):15464–15470, December 2001.

[114] Francesca Del Vecchio, Hrvoje Petkovic, Steven G. Kendrew, Lindsey Low, Barrie Wilkinson, Rachel Lill, Jesús Cortés, Brian A.M. Rudd, Jim Staunton, and Peter F. Leadlay. Active-site residue, domain and module swaps in modular polyketide synthases. *Journal of Industrial Microbiology and Biotechnology*, 30(8):489–494, 2003.

[115] Kenny Bravo-Rodriguez, Stephan Klopries, Kyra R M Koopmans, Uschi Sundermann, Samir Yahiaoui, Julia Arens, Susanna Kushnir, Frank Schulz, and Elsa Sanchez-Garcia. Substrate Flexibility of a Mutated Acyltransferase Domain and Implications for Polyketide Biosynthesis. *Chemistry and Biology*, 22(11):1425–1430, 2015.

[116] Edward Kalkreuter, Kyle S. Bingham, Aaron M. Keeler, Andrew N. Lowell, Jennifer J. Schmidt, David H. Sherman, and Gavin J. Williams. Computationally-guided exchange of substrate selectivity motifs in a modular polyketide synthase acyltransferase. *Nature Communications*, 12(1):2193, December 2021.

[117] A. Schneider, T. Stachelhaus, and M. A. Marahiel. Targeted alteration of the substrate specificity of peptide synthetases by rational module swapping. *Molecular and General Genetics*, 257(3):308–318, 1998.

[118] Henning D Mootz, Dirk Schwarzer, and Mohamed A Marahiel. Construction of hybrid peptide synthetases by module and domain fusions. *Proceedings of the National Academy of Sciences*, 97(11):5848–5853, 2000.

[119] Peter J. Belshaw. Aminoacyl-CoAs as Probes of Condensation Domain Selectivity in Nonribosomal Peptide Synthesis. *Science*, 284(5413):486–489, April 1999.

[120] David E. Ehmann, John W. Trauger, Torsten Stachelhaus, and Christopher T. Walsh. Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chemistry and Biology*, 7(10):765–772, 2000.

[121] Mark J. Calcott, Jeremy G. Owen, Iain L. Lamont, and David F. Ackerley. Biosynthesis of novel pyoverdines by domain substitution in a nonribosomal peptide synthetase of pseudomonas aeruginosa. *Applied and Environmental Microbiology*, 80(18):5723–5731, 2014.

[122] Alan Tanovic, Stefan a Samel, Lars-Oliver Essen, and Mohamed a Marahiel. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science (New York, N.Y.)*, 321(5889):659–663, 2008.

[123] Michael a Fischbach, Jonathan R Lai, Eric D Roche, Christopher T Walsh, and David R Liu. Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. *Proceedings of the National Academy of Sciences*, 104(29):11951–11956, July 2007.

[124] Kenan A. J. Bozhüyük, Florian Fleischhacker, Annabell Linck, Frank Wesche, Andreas Tietze, Claus-Peter Niesert, and Helge B. Bode. De novo design and engineering of non-ribosomal peptide synthetases. *Nature Chemistry*, 10(3):275–281, March 2018.

[125] Kenan A J Bozhüyük, Annabell Linck, Andreas Tietze, Janik Kranz, Frank Wesche, Sarah Nowak, Florian Fleischhacker, Yan-ni Shi, Peter Grün, and Helge B Bode. Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nature Chemistry*, 11(7):653–661, July 2019.

[126] Mark J. Calcott, Jeremy G. Owen, and David F. Ackerley. Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. *Nature Communications*, 11(1):4554, December 2020.

[127] Max Cru, Axel Zeeck, Ivonne Höfer, Max Crüsemann, Markus Radzom, Bernadette Geers, Daniel Flachshaar, Xiaofeng Cai, Axel Zeeck, and Jörn Piel. Insights into the biosynthesis of hormaomycin, an exceptionally complex bacterial signaling metabolite. *Chemistry and Biology*, 18(3):381–391, 2011.

[128] Max Crüsemann, Christoph Kohlhaas, and Jörn Piel. Evolution-guided engineering of nonribosomal peptide synthetase adenylation domains. *Chemical Science*, 4(3):1041–1045, 2013.

[129] Aleksa Stanišić, Annika Hüsken, and Hajo Kries. HAMA: a multiplexed LC-MS/MS assay for specificity profiling of adenylate-forming enzymes. *Chemical Science*, 10(44):10395–10399, 2019.

[130] David L Niquille, Douglas A Hansen, Takahiro Mori, David Fercher, Hajo Kries, and Donald Hilvert. Nonribosomal biosynthesis of backbone-modified peptides. *Nature Chemistry*, 10(March), 2018.

[131] Aleksa Stanišić, Annika Hüsken, Philipp Stephan, David L Niquille, Jochen Reinstein, and Hajo Kries. An Engineered Nonribosomal Peptide Synthetase Shows Opposite Amino Acid Loading and Condensation Specificity. *ChemRxiv*, 2021.

[132] Katrin Eppelmann, Torsten Stachelhaus, and Mohamed A. Marahiel. Exploitation of the selectivity-conferring code of nonribosomal peptide synthetases for the rational design of novel peptide antibiotics. *Biochemistry*, 41(30):9718–9726, 2002.

[133] Milda Kaniusaite, Tiia Kittilä, Robert J. A. Goode, Ralf B. Schittenhelm, and Max J. Cryle. Redesign of Substrate Selection in Glycopeptide Antibiotic Biosynthesis Enables Effective Formation of Alternate Peptide Backbones. *ACS Chemical Biology*, 15(9):2444–2455, September 2020.

[134] Benoit Villiers and Florian Hollfelder. Directed evolution of a gatekeeper domain in nonribosomal peptide synthesis. *Chemistry and Biology*, 18(10):1290–1299, 2011.

[135] Hajo Kries, Rudolf Wachtel, Anja Pabst, Benedikt Wanner, David Niquille, and Donald Hilvert. Reprogramming nonribosomal peptide synthetases for "clickable" amino acids. *Angewandte Chemie - International Edition*, 53(38):10105–10108, 2014.

[136] David L. Niquille, Ines B. Folger, Sophie Basler, and Donald Hilvert. Biosynthetic Functionalization of Nonribosomal Peptides. *Journal of the American Chemical Society*, 2021.

[137] Jeremy G. Owen, Janine N. Copp, and David F. Ackerley. Rapid and flexible biochemical assays for evaluating 4'-phosphopantetheinyl transferase activity. *Biochemical Journal*, 436(3):709–717, June 2011.

[138] Luis E N Quadri, Jason Sello, Thomas A Keating, Paul H Weinreb, and Christopher T Walsh. Identification of a Mycobacterium tuberculosis gene cluster encoding the biosynthetic enzymes for assembly of the siderophore mycobactin. *Chemistry & Biology*, 5(11):631–645, 1998.

[139] Tatjana Taubitz and Lutz Heide. Role of MbtH-like Proteins in the Adenylation of Tyrosine during Aminocoumarin and Vancomycin Biosynthesis. *The Journal of Biological Chemistry*, 286(42):36281–36290, 2011.

[140] Bradley R. Miller, Eric J. Drake, Ce Shi, Courtney C. Aldrich, and Andrew M. Gulick. Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic domain architecture. *Journal of Biological Chemistry*, 291(43):22559–22571, 2016.

[141] Vigdis Torsvik, Jostein Goksyr, and Frida Lise Daae. High Diversity in DNA of Soil Bacteria. *Applied and Environmental Microbiology*, 56(3):782–787, 1990.

[142] Jeffrey W Craig, Fang-yuan Chang, and Sean F Brady. Natural Products from Environmental DNA Hosted in Ralstonia metallidurans. *ACS Chemical Biology*, 4(1):23–28, 2009.

[143] Gui-yang-sheng Wang, Edmund Graziani, Barbara Waters, Wubin Pan, Xiang Li, Joe Mcdermott, Guido Meurer, Geeta Saxena, Raymond J Andersen, and Julian Davies. Novel Natural Products from Soil DNA Libraries in a Streptomycete Host. *Organic Letters*, 2(16):2401–2404, 2000.

[144] Jeremy G Owen, Zachary Charlop-Powers, Alexandra G Smith, Melinda A Ternei, Paula Y Calle, Boojala Vijay B Reddy, Daniel Montiel, and Sean F Brady. Multiplexed metagenome mining using short DNA sequence tags facilitates targeted discovery of epoxyketone proteasome inhibitors. *PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF THE UNITED STATES OF AMERICA*, 112(14):4221–4226, April 2015.

[145] Boojala Vijay B. Reddy, Aleksandr Milshteyn, Zachary Charlop-Powers, and Sean F. Brady. eSNaPD: a versatile, web-based bioinformatics platform for surveying and mining natural product biosynthetic diversity from metagenomes. *Chemistry & biology*, 21(8):1023–1033, 2014.

[146] J. W. Craig, F.-Y. Chang, J. H. Kim, S. C. Obiajulu, and S. F. Brady. Expanding Small-Molecule Functional Metagenomics through Parallel Screening of Broad-Host-Range Cosmid Environmental DNA Libraries in Diverse Proteobacteria. *Applied and Environmental Microbiology*, 76(5):1633–1641, 2010.

[147] David Cole Stevens, Kyle R Conway, Nelson Pearce, Luis Roberto Villegas-Penaranda, Anthony G Garza, and Christopher N Boddy. Alternative Sigma Factor Over-Expression Enables Heterologous Expression of a Type II Polyketide Biosynthetic Pathway in Escherichia coli. *PLOS ONE*, 8(5):e64858, May 2013.

[148] Jia Jia Zhang, Xiaoyu Tang, Michelle Zhang, Darlene Nguyen, and Bradley S. Moore. Broad-Host-Range Expression Reveals Native and Host Regulatory Elements That Influence Heterologous Antibiotic Production in Gram-Negative Bacteria. *MBio*, 8(5):e01291–17, 2017.

[149] Daniel E Schaufelberger, Mary P Koleck, John A Beutler, Angel M Vatakis, A Belinda Alvarado, Pamela Andrews, L V Marzo, G M Muschik, John Roach, John T Ross, B Lebherz, Marie P Reeves, Robert M Eberwein, Lori L Rodgers, Robert P Testerman, and M Snader. The Large-Scale Isolation of Bryostatin 1 from Bugula neritina Following Current Good Manufacturing Practices. *Journal of Natural Products*, 54(5):1265–1270, 1991.

[150] Barry M Trost and Guangbin Dong. Total synthesis of bryostatin 16 using atom-economical and chemoselective approaches. *Nature*, 456(November):485–488, 2008.

[151] Elvis Legala Ongey and Peter Neubauer. Lanthipeptides: chemical synthesis versus in vivo biosynthesis as tools for pharmaceutical production. *Microbial Cell Factories*, 15(1):97, 2016.

[152] Hengqian Ren, Bin Wang, and Huimin Zhao. Breaking the silence: new strategies for discovering novel natural products. *Current Opinion in Biotechnology*, 48:21–27, December 2017.

[153] Hee Ju Nah, Hye Rim Pyeon, Seung Hoon Kang, Si Sun Choi, and Eung Soo Kim. Cloning and heterologous expression of a large-sized natural product biosynthetic gene cluster in Streptomyces species. *Frontiers in Microbiology*, 8(MAR):394, 2017.

[154] Emily Freed, Jacob Fenster, Sharon L. Smolinski, Julie Walker, Calvin A. Henard, Ryan Gill, and Carrie A. Eckert. Building a genome engineering toolbox in non-model prokaryotic microbes. *Biotechnology and Bioengineering*, 115(9):2120–2138, 2018.

[155] Marnix H Medema, Rainer Breitling, and Eriko Takano. *Synthetic Biology in Streptomyces Bacteria*, volume 497. Elsevier Inc., 1 edition, 2011.

[156] Pablo I Nikel, Esteban Martínez-García, and Víctor de Lorenzo. Biotechnological domestication of pseudomonads using synthetic biology. *Nature reviews. Microbiology*, 12(5):368–79, 2014.

[157] Xin Tian Li, Lynn C. Thomason, James A. Sawitzke, Nina Costantino, and Donald L. Court. Positive and negative selection using the tetA-sacB cassette: Recombineering and P1 transduction in Escherichia coli. *Nucleic Acids Research*, 41(22):e204, 2013.

[158] Nadja Graf and Josef Altenbuchner. Development of a Method for Markerless Gene Deletion in Pseudomonas putida. *Applied and Environmental Microbiology*, 77(15):5549–5552, August 2011.

[159] Martin Jinek, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A Doudna, and Emmanuelle Charpentier. A Programmable Dual-RNA – Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, 337(August):816–822, 2012.

[160] Bernd Zetsche, Jonathan S Gootenberg, Omar O Abudayyeh, Aviv Regev, Eugene V Koonin, and Feng Zhang. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, 163(3):759–771, 2015.

[161] He Huang, Guosong Zheng, Weihong Jiang, Haifeng Hu, and Yinhua Lu. One-step high-efficiency CRISPR / Cas9-mediated genome editing in Streptomyces. *Acta Biochimica et Biophysica Sinica*, 47(March):231–243, 2015.

[162] Ryan E. Cobb, Yajie Wang, and Huimin Zhao. High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synthetic Biology*, 4(6):723–728, 2015.

[163] Haoran Zhang, Brett A. Boghigian, and Blaine A. Pfeifer. Investigating the role of native propionyl-CoA and methylmalonyl-CoA metabolism on heterologous polyketide production in Escherichia coli. *Biotechnology and Bioengineering*, 105(3):567–573, 2010.

[164] Haruo Ikeda, Kazuo Shin-ya, and Satoshi Omura. Genome mining of the Streptomyces avermitilis genome and development of genome-minimized hosts for heterologous expression of biosynthetic gene clusters. *Journal of Industrial Microbiology & Biotechnology*, 41:233–250, 2014.

[165] Andreas Domröse, Andreas S Klein, Jennifer Hage-Hülsmann, Stephan Thies, Vera Svensson, Thomas Classen, Jörg Pietruszka, Karl-Erich Jaeger, Thomas Drepper, and Anita Loeschcke. Efficient recombinant production of prodigiosin in Pseudomonas putida. *Frontiers in Microbiology*, 6(September):972, September 2015.

[166] Kutlu Ozergin-Ulgen and Ferda Mavituna. Oxygen transfer and uptake in Streptomyces coelicolor A3(2) culture in a batch bioreactor. *Journal of Chemical Technology and Biotechnology*, 73(3):243–250, 1998.

[167] Maria Martin-Pascual, Christos Batianis, Lyon Bruinsma, Enrique Asin-Garcia, Luis Garcia-Morales, Ruud A. Weusthuis, Richard van Kranenburg, and Vitor A.P. Martins dos Santos. A navigation guide of synthetic biology tools for Pseudomonas putida. *Biotechnology Advances*, 49:107732, July 2021.

[168] Andreas Ankenbauer, Richard A. Schäfer, Sandra C. Viegas, Vânia Pobre, Björn Voß, Cecília M. Arraiano, and Ralf Takors. Pseudomonas putida KT2440 is naturally endowed to withstand industrial-scale stress conditions. *Microbial Biotechnology*, 13(4):1145–1161, July 2020.

[169] Jacqueline M. Rand, Tippapha Pisithkul, Ryan L. Clark, Joshua M. Thiede, Christopher R. Mehrer, Daniel E. Agnew, Candace E. Campbell, Andrew L. Markley, Morgan N. Price, Jayashree Ray, Kelly M. Wetmore, Yumi Suh, Adam P. Arkin, Adam M. Deutschbauer, Daniel Amador-Noguez, and Brian F. Pfleger. A metabolic pathway for catabolizing levulinic acid in bacteria. *Nature Microbiology*, 2(12):1624–1634, December 2017.

[170] Davinia Salvachúa, Allison Z Werner, Isabel Pardo, Martyna Michalska, Brenna A Black, Bryon S. Donohoe, Stefan J. Haugen, Rui Katahira, Sandra Notonier, Kelsey J. Ramirez, Antonella Amore, Samuel O. Purvine, Erika M. Zink, Paul E.

Abraham, Richard J. Giannone, Suresh Poudel, Philip D. Laible, Robert L. Hettich, and Gregg T. Beckham. Outer membrane vesicles catabolize lignin-derived aromatic compounds in Pseudomonas putida KT2440. *Proceedings of the National Academy of Sciences*, 117(17):9302–9310, April 2020.

[171] Harald Gross and Joyce E Loper. Genomics of secondary metabolite production by Pseudomonas spp. *Natural product reports*, 26(11):1408–1446, 2009.

[172] Katherine Gregory, Laura A Salvador, Shukria Akbar, Barbara I Adaikpoh, and D Cole Stevens. Survey of Biosynthetic Gene Clusters from Sequenced Myxobacteria Reveals Unexplored Biosynthetic Potential. *Microorganisms*, 7(6):181, June 2019.

[173] Dongwon Park, Girish Swayambhu, and Blaine A. Pfeifer. Heterologous biosynthesis as a platform for producing new generation natural products. *Current Opinion in Biotechnology*, 66:123–130, December 2020.

[174] Sonja Christina Troeschel, Stephan Thies, Olga Link, Catherine Isabell Real, Katja Knops, Susanne Wilhelm, Frank Rosenau, and Karl-Erich Jaeger. Novel broad host range shuttle vectors for expression in Escherichia coli, Bacillus subtilis and Pseudomonas putida. *Journal of Biotechnology*, 161(2):71–79, October 2012.

[175] J. Lin and D. R. Helinski. Analysis of mutations in trfA, the replication initiation gene of the broad-host-range plasmid RK2. *Journal of Bacteriology*, 174(12):4110–4119, 1992.

[176] Taek Lee, Rachel A Krupa, Fuzhong Zhang, Meghdad Hajimorad, William J Holtz, Nilu Prasad, Sung Lee, and Jay D Keasling. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *Journal of Biological Engineering*, 5(1):12, 2011.

[177] K. Nagahari and K. Sakaguchi. RSF1010 plasmid as a potentially useful vector in Pseudomonas species. *Journal of Bacteriology*, 133(3):1527–1529, 1978.

[178] Daniel C. Volke, Justine Turlin, Viviënne Mol, and Pablo I. Nikel. Physical decoupling of XylS/ Pm regulatory elements and conditional proteolysis enable precise control of gene expression in Pseudomonas putida. *Microbial Biotechnology*, 13(1):222–232, January 2020.

[179] Joshua R. Elmore, Anna Furches, Gara N. Wolff, Kent Gorday, and Adam M. Guss. Development of a high efficiency integration system and promoter library for rapid modification of Pseudomonas putida KT2440. *Metabolic Engineering Communications*, 5(April):1–8, December 2017.

[180] JC Anderson. Anderson Promoter Library Registry of Standard Biological Parts, 2006.

[181] Patricia Calero, Sheila Ingemann Jensen, and Alex Toftgaard Nielsen. Broad host range ProUSER vectors enable fast characterization of inducible promoters and optimization of p-coumaric acid production in Pseudomonas putida KT2440. *ACS Synthetic Biology*, 5(7):acssynbio.6b00081, 2016.

[182] Victor de Lorenzo, Silvia Fernández, Marta Herrero, Ute Jakubzik, and Kenneth N. Timmis. Engineering of alkyl- and haloaromatic-responsive gene expression with mini-transposons containing regulated promoters of biodegradative pathways of Pseudomonas. *Gene*, 130(1):41–46, 1993.

[183] Sven Panke, V\'\ictor de Lorenzo, Arnë Kaiser, Bernard Witholt, and Marcel G Wubbolts. Engineering of a Stable Whole-Cell Biocatalyst Capable of (S)-Styrene Oxide Formation for Continuous Two-Liquid-Phase Applications. *Applied and Environmental Microbiology*, 65(12):5619–5623, 1999.

[184] Marcel Jeske and Josef Altenbuchner. The Escherichia coli rhamnose promoter rhaP BAD is in Pseudomonas putida KT2440 independent of Crp–cAMP activation. *Applied Microbiology and Biotechnology*, 85(6):1923–1933, February 2010.

[185] Nadja Graf and Josef Altenbuchner. Functional characterization and application of a tightly regulated MekR/P mekA expression system in Escherichia coli and Pseudomonas putida. *Applied Microbiology and Biotechnology*, 97(18):8239–8251, 2013.

[186] Sarah Lieder, Pablo I Nikel, Víctor de Lorenzo, and Ralf Takors. Genome reduction boosts heterologous gene expression in Pseudomonas putida. *Microbial Cell Factories*, 14(1):23, 2015.

[187] Belén Calles, Ángel Goñi-Moreno, and Víctor Lorenzo. Digitalizing heterologous gene expression in Gram-negative bacteria with a portable ON/OFF module. *Molecular Systems Biology*, 15(12):1–16, December 2019.

[188] Sue Zanne Tan, Christopher R. Reisch, and Kristala L. J. Prather. A Robust CRISPR Interference Gene Repression System in Pseudomonas. *Journal of Bacteriology*, 200(7):e00575–17, January 2018.

[189] Yueyue Zhou, Lu Lin, Heng Wang, Zhichao Zhang, Jizhong Zhou, and Nianzhi Jiao. Development of a CRISPR/Cas9n-based tool for metabolic engineering of Pseudomonas putida for ferulic acid-to-polyhydroxyalkanoate bioconversion. *Communications Biology*, 3(1):98, December 2020.

[190] Howard M Salis, Ethan A Mirsky, and Christopher A Voigt. Automated design of synthetic ribosome binding sites to control protein expression. *Nature biotechnology*, 27(10):946–50, October 2009.

[191] Iasson E.P. Tozakidis, Lena M. Lüken, Alina Üffing, Annika Meyers, and Joachim Jose. Improving the autotransporter-based surface display of enzymes in Pseudomonas putida KT2440. *Microbial Biotechnology*, 13(1):176–184, 2020.

[192] Daniel Christoph Volke, Karel Olavarría, and Pablo Iván Nikel. Cofactor Specificity of Glucose-6-Phosphate Dehydrogenase Isozymes in Pseudomonas putida Reveals a General Principle Underlying Glycolytic Strategies in Bacteria. *mSystems*, 6(2):1–21, March 2021.

[193] Tomas Aparicio, Akos Nyerges, Esteban Martínez-García, and Víctor de Lorenzo. High-Efficiency Multi-site Genomic Editing of Pseudomonas putida through Thermoinducible ssDNA Recombineering. *iScience*, 23(3):100946, March 2020.

[194] Kazuki Sato, Kenichi Ushioda, Keiji Akiba, Yoshimi Matsumoto, Hideaki Maseda, Tasuke Ando, Emiko Isogai, Taiji Nakae, and Hiroshi Yoneyama. Development of a Novel Antimicrobial Screening System Targeting the Pyoverdine-Mediated Iron Acquisition System and Xenobiotic Efflux Pumps. *Molecules*, 20(5):7790–7806, April 2015.

[195] Sei ichi Nakamura, Yoshimitsu Oda, Tsutomu Shimada, Iwashiro Oki, and Kanji Sugimoto. SOS-inducing activity of chemical carcinogens and mutagens in Salmonella typhimurium TA1535/pSK1002: examination with 151 chemicals. *Mutation Research Letters*, 192(4):239–246, 1987.

[196] Tomás Aparicio, Víctor de Lorenzo, and Esteban Martínez-García. CRISPR/Cas9-Based Counterselection Boosts Recombineering Efficiency in Pseudomonas putida. *Biotechnology Journal*, 13(5):1700161, May 2018.

[197] Zhixin Wu, Zhongqiu Chen, Xinyue Gao, Jing Li, and Guangdong Shang. Combination of ssDNA recombineering and CRISPR-Cas9 for Pseudomonas putida KT2440 genome editing. *Applied Microbiology and Biotechnology*, 103(6):2783–2795, March 2019.

[198] Jun Sun, Qingzhuo Wang, Yu Jiang, Zhiqiang Wen, Lirong Yang, Jianping Wu, and Sheng Yang. Genome editing and transcriptional repression in Pseudomonas putida KT2440 via the type II CRISPR system. *Microbial Cell Factories*, 17(1):41, 2018.

[199] Ekaterina Kozaeva, Nicolas T Wirth, and Pablo I Nikel. Accelerated genome engineering of Pseudomonas putida by I-SceI—mediated recombination and CRISPR-Cas9 counterselection. *Microbial Biotechnology*, pages 1–17, 2019.

[200] Tomas Aparicio, Akos Nyerges, István Nagy, Csaba Pal, Esteban Martínez-García, and Víctor Lorenzo. Mismatch repair hierarchy of Pseudomonas putida revealed by mutagenic ssDNA recombineering of the pyrF gene. *Environmental Microbiology*, 22(1):45–58, January 2020.

[201] Sebastian Zobel, Ilaria Benedetti, Lara Eisenbach, Victor De Lorenzo, Nick Wierckx, and Lars M Blank. Tn7-Based Device for Calibrated Heterologous Gene Expression in Pseudomonas putida. *ACS Synthetic Biology*, 4(12):1341–1351, 2015.

[202] Joshua R Elmore, Gara N Dexter, Ryan Francis, Lauren Riley, Jay Huenemann, Henri Baldino, Adam M Guss, and Robert Egbert. The SAGE genetic toolkit enables highly efficient, iterative site-specific genome engineering in bacteria. *bioRxiv*, 2020.

[203] Wei Niu, Howard Willett, Joshua Mueller, Xinyuan He, Levi Kramer, Bin Ma, and Jiantao Guo. Direct biosynthesis of adipic acid from lignin-derived aromatics using engineered Pseudomonas putida KT2440. *Metabolic Engineering*, 59(February):151–161, May 2020.

[204] Michael Kohlstedt, Sören Starck, Nadja Barton, Jessica Stolzenberger, Mirjam Selzer, Kerstin Mehlmann, Roland Schneider, Daniel Pleissner, Jan Rinkel, Jeroen S Dickschat, Joachim Venus, Jozef B J H Van Duuren, and Christoph Wittmann.

From lignin to nylon: Cascaded chemical and biochemical conversion using metabolically engineered Pseudomonas putida. *Metabolic Engineering*, 47(March):279–293, 2018.

[205] Davinia Salvachúa, Thomas Rydzak, Raquel Auwae, Annette De Capite, Brenna A Black, Jason T Bouvier, Nicholas S Cleveland, Joshua R. Elmore, Anna Furches, Jay D Huenemann, Rui Katahira, William E. Michener, Darren J Peterson, Holly Rohrer, Derek R Vardon, Gregg T Beckham, and Adam M Guss. Metabolic engineering of Pseudomonas putida for increased polyhydroxyalkanoate production from lignin. *Microbial Biotechnology*, 13(1):290–298, January 2020.

[206] Frank Gross, Michael W Ring, Olena Perlova, Jun Fu, Susan Schneider, Klaus Gerth, Silvia Kuhlmann, A Francis Stewart, Youming Zhang, and Rolf Müller. Metabolic engineering of Pseudomonas putida for methylmalonyl-CoA biosynthesis to enable complex heterologous secondary metabolite formation. *Chemistry & biology*, 13(12):1253–64, 2006.

[207] Mitchell G Thompson, Luis E Valencia, Jacquelyn M. Blake-Hedges, Pablo Cruz-Morales, Alexandria E Velasquez, Allison N Pearson, Lauren N Sermeno, William A Sharpless, Veronica T Benites, Yan Chen, Edward E.K. Baidoo, Christopher J Petzold, Adam M Deutschbauer, and Jay D Keasling. Omics-driven identification and elimination of valerolactam catabolism in Pseudomonas putida KT2440 for increased product titer. *Metabolic Engineering Communications*, 9(August):e00098, December 2019.

[208] Yuqian Gao, Thomas L. Fillmore, Nathalie Munoz, Gayle J. Bentley, Christopher W. Johnson, Joonhoon Kim, Jamie A. Meadows, Jeremy D. Zucker, Meagan C. Burnet, Anna K. Lipton, Aivett Bilbao, Daniel J. Orton, Young-Mo Kim, Ronald J. Moore, Errol W. Robinson, Scott E. Baker, Bobbie-Jo M. Webb-Robertson, Adam M. Guss, John M. Gladden, Gregg T. Beckham, Jon K. Magnuson, and Kristin E. Burnum-Johnson. High-Throughput Large-Scale Targeted Proteomics Assays for Quantifying Pathway Proteins in Pseudomonas putida KT2440. *Frontiers in Bioengineering and Biotechnology*, 8(December):1–13, December 2020.

[209] Gayle J Bentley, Niju Narayanan, Ramesh K Jha, Davinia Salvachúa, Joshua R Elmore, George L Peabody, Brenna A Black, Kelsey Ramirez, Annette De Capite, William E Michener, Allison Z Werner, Dawn M Klingeman, Heidi S Schindel, Robert Nelson, Lindsey Foust, Adam M Guss, Taraka Dale, Christopher W Johnson, and Gregg T Beckham. Engineering glucose metabolism for enhanced muconic acid production in Pseudomonas putida KT2440. *Metabolic Engineering*, 59(January):64–75, May 2020.

[210] Hongbo Liu, Hao Jiang, Bradley Haltli, Kerry Kulowski, Elwira Muszynska, Xidong Feng, Mia Summers, Mairead Young, Edmund Graziani, Frank Koehn, Guy T Carter, and Min He. Rapid Cloning and Heterologous Expression of the Meridamycin Biosynthetic Gene Cluster Using a Versatile Escherichia coli - Streptomyces Artificial Chromosome Vector, pSBAC. *Journal of Natural Products*, 72(3):389–395, 2009.

[211] Chao Chen, Xinqing Zhao, Yingyu Jin, Zongbao Kent, and Joo-won Suh. Rapid construction of a Bacterial Artificial Chromosomal (BAC) expression vector using designer DNA fragments. *Plasmid*, 76:79–86, 2014.

[212] Laxmi Prasad Thapa, Tae-jin Oh, Hei Chan Lee, Kwangkyoung Liou, Je Won Park, Yeo Joon Yoon, and Jae Kyung Sohng. Heterologous expression of the kanamycin biosynthetic gene cluster (pSKC2) in Streptomyces venezuelae YJ003. *Applied Genetics and Molecular Biotechnology*, 76:1357–1364, 2007.

[213] Jun Fu, Silke C. Wenzel, Olena Perlova, Junping Wang, Frank Gross, Zhiru Tang, Yulong Yin, A. Francis Stewart, Rolf Müller, and Youming Zhang. Efficient transfer of two large secondary metabolite pathway gene clusters into heterologous hosts by transposition. *Nucleic Acids Research*, 36(17):e113–e113, October 2008.

[214] Patrick Videau, Kaitlyn N. Wells, Arun J. Singh, Jessie Eiting, Philip J. Proteau, and Benjamin Philmus. Expanding the Natural Products Heterologous Expression Repertoire in the Model Cyanobacterium Anabaena sp. Strain PCC 7120: Production of Pendolmycin and Teleocidin B-4. *ChemRxiv*, pages 1–43, 2019.

[215] Weihua Chen and Zhongjun Qin. Development of a gene cloning system in a fast-growing and moderately thermophilic Streptomyces species and heterologous expression of Streptomyces antibiotic biosynthetic gene clusters. *BMC Microbiology*, 11(1):243, 2011.

[216] Keshav K. Nepal and Guojun Wang. Streptomycetes: Surrogate hosts for the genetic manipulation of biosynthetic gene clusters and production of natural products. *Biotechnology Advances*, 37:1–20, 2019.

[217] Ryan M Phelan, Daniel Sachs, Shayne J Petkiewicz, Jesus F Barajas, Jacquelyn M Blake-Hedges, Mitchell G Thompson, Amanda Reider Apel, Blake J Rasor, Leonard Katz, and Jay D Keasling. Development of Next Generation Synthetic Biology Tools for Use in Streptomyces venezuelae. *ACS Synthetic Biology*, 6(1):159–166, 2016.

[218] D. R. Zusman, D. M. Krotoski, and M. Cumsky. Chromosome replication in Myxococcus xanthus. *Journal of Bacteriology*, 133(1):122–129, 1978.

[219] Pablo I. Nikel and Víctor de Lorenzo. Pseudomonas putida as a functional chassis for industrial biocatalysis: From native biochemistry to trans-metabolism. *Metabolic Engineering*, 50(April):142–155, November 2018.

[220] Yongxin Li, Zhongrui Li, Kazuya Yamanaka, Ying Xu, Weipeng Zhang, and Hera Vlamakis. Directed natural product biosynthesis gene cluster capture and expression in the model bacterium Bacillus subtilis. *Scientific Reports*, 5:1–7, 2015.

[221] René H. Wijffels, Olaf Kruse, and Klaas J. Hellingwerf. Potential of industrial biotechnology with cyanobacteria and eukaryotic microalgae. *Current Opinion in Biotechnology*, 24(3):405–413, 2013.

[222] Wim F. J. Vermaas, A. William Rutherford, and Orjan Hansson. Site-directed mutagenesis in photosystem II of the cyanobacterium Synechocystis sp . PCC 6803 : Donor D is a tyrosine residue in the D2 protein. *Proceedings of the National Academy of Sciences*, 85(November):8477–8481, 1988.

[223] Jingjie Yu, Michelle Liberton, Paul F. Cliften, Richard D. Head, Jon M. Jacobs, Richard D. Smith, David W. Koppenaal, Jerry J. Brand, and Himadri B. Pakrasi. Synechococcus elongatus UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO\textsubscript{2}. *Scientific reports*, 5:8132, 2015.

[224] Nicolas Seghezzi, Patrick Amar, Brian Koebmann, Peter R Jensen, and Marie-joëlle Virolle. The construction of a library of synthetic promoters revealed some specific features of strong Streptomyces promoters. *Applied Microbiology and Biotechnology*, 90:615–623, 2011.

[225] Taylor B. Cook, Jacqueline M. Rand, Wasti Nurani, Dylan K. Courtney, Sophia A. Liu, and Brian F. Pfleger. Genetic tools for reliable gene expression and recombineering in Pseudomonas putida. *Journal of Industrial Microbiology & Biotechnology*, 45(7):517–527, July 2018.

[226] Yafeng Song, Jonas M. Nikoloff, Gang Fu, Jingqi Chen, Qinggang Li, Nengzhong Xie, Ping Zheng, Jibin Sun, and Dawei Zhang. Promoter screening from Bacillus subtilis in various conditions hunting for synthetic biology and industrial applications. *PLoS ONE*, 11(7):e0158447, 2016.

[227] Akiyoshi Higo, Atsuko Isu, Yuki Fukaya, and Toru Hisabori. Efficient Gene Induction and Endogenous Gene Repression Systems for the Filamentous Cyanobacterium Anabaena sp . PCC 7120. *Plant & Cell Physiology*, 57(2):387–396, 2016.

[228] Elias Englund, Feiyan Liang, and Pia Lindberg. Evaluation of promoters and ribosome binding sites for biotechnological applications in the unicellular cyanobacterium Synechocystis sp . PCC 6803. *Scientific Reports*, 6(36640), 2016.

[229] Olena Perlova, Jun Fu, Silvia Kuhlmann, Daniel Krug, A Francis Stewart, Youming Zhang, and Rolf Müller. Reconstitution of the myxothiazol biosynthetic gene cluster by Red/ET recombination and heterologous expression in Myxococcus xanthus. *Appl. Environ. Microbiol.*, 72(12):7485–7494, 2006.

[230] Esteban Martínez-García, Belén Calles, Miguel Arévalo-Rodríguez, and Víctor de Lorenzo. pBAM1: an all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiology*, 11(1):38, 2011.

[231] Kyeong Rok Choi, Jae Sung Cho, In Jin Cho, Dahyeon Park, and Sang Yup Lee. Markerless gene knockout and integration to express heterologous biosynthetic gene clusters in Pseudomonas putida. *Metabolic Engineering*, 47(April):463–474, May 2018.

[232] Li Ping Zhu, Xin Jing Yue, Kui Han, Zhi Feng Li, Lian Shuai Zheng, Xiu Nan Yi, Hai Long Wang, and You Ming Zhang. Allopatric integrations selectively change host transcriptomes , leading to varied expression efficiencies of exotic genes in Myxococcus xanthus. *Microbial Cell Factories*, 14:105, 2015.

[233] Mario Juhas and James W. Ajioka. Integrative bacterial artificial chromosomes for DNA integration into the Bacillus subtilis chromosome. *Journal of Microbiological Methods*, 125:1–7, 2016.

[234] Julia Roulet, Arnaud Taton, James W Golden, Ana Arabolaza, Michael D Burkart, and Hugo Gramajo. Development of a cyanobacterial heterologous polyketide production platform. *Metabolic Engineering*, 49(July):94–104, 2018.

[235] Andrew L. Markley, Matthew B. Begemann, Ryan E. Clarke, Gina C. Gordon, and Brian F. Pfleger. A synthetic biology toolbox for controlling gene expression in the cyanobacterium Synechococcus strain sp. PCC 7002. *ACS Synthetic Biology*, 4(5):595–603, 2014.

[236] Deng Liu and Himadri B. Pakrasi. Exploring native genetic elements as plug-in tools for synthetic biology in the cyanobacterium Synechocystis sp. PCC 6803. *Microbial Cell Factories*, 17(1):1–8, 2018.

[237] Ying Yang, Ye Wang, Zhi-feng Li, Ya Gong, Peng Zhang, Wen-chao Hu, Duohong Sheng, and Yue-zhong Li. Increasing on-target cleavage efficiency for CRISPR/Cas9-induced large fragment deletion in Myxococcus xanthus. *Microbial Cell Factories*, 16:142, 2017.

[238] Anna A. Toymentseva and Josef Altenbuchner. New CRISPR-Cas9 vectors for genetic modifications of Bacillus species. *FEMS Microbiology Letters*, 366(1):fny284, 2019.

[239] Marcelo C. Bassalo, Andrew D. Garst, Andrea L. Halweg-Edwards, William C. Grau, Dylan W. Domaille, Vivek K. Mutalik, Adam P. Arkin, and Ryan T. Gill. Rapid and Efficient One-Step Metabolic Pathway Integration in E. coli. *ACS Synthetic Biology*, 5:561–568, 2016.

[240] Justin Ungerer and Himadri B Pakrasi. Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria. *Scientific Reports*, 6:39681, 2016.

[241] Fumitaka Kudo, Akimasa Miyanaga, and Tadashi Eguchi. Biosynthesis of natural products containing B-amino acids. *Natural Product Reports*, 31:1056–1073, 2014.

[242] Pooja Arora, Archana Vats, Priti Saxena, Debasisa Mohanty, and Rajesh S. Gokhale. Promiscuous fatty acyl CoA ligases produce acyl-CoA and acyl-SNAC precursors for polyketide biosynthesis. *Journal of the American Chemical Society*, 127(26):9388–9389, 2005.

[243] Ross Zirkle, James M. Ligon, and István Molnár. Heterologous production of the antifugal polyketide antibiotic soraphen A of Sorangium cellulosum So ce26 in Streptomyces lividans. *Microbiology*, 150(8):2761–2774, 2004.

[244] Won Seok Jung, Eunji Kim, Young Ji Yoo, Yeon Hee Ban, Eun Ji Kim, and Yeo Joon Yoon. Characterization and engineering of the ethylmalonyl-CoA pathway towards the improved heterologous production of polyketides in Streptomyces venezuelae. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 98(8):3701–3713, April 2014.

[245] Emma J Rackham, Sabine Grüschow, Amany E Ragab, Shilo Dickens, and Rebecca J M Goss. Pacidamycin Biosynthesis : Identification and Heterologous Expression of the First Uridyl Peptide Antibiotic Gene Cluster. *ChemBioChem*, 11:1700–1709, 2010.

[246] Xuejun Zhu, Joyce Liu, and Wenjun Zhang. De novo biosynthesis of terminal alkyne-labeled natural products. *Nature Chemical Biology*, 11(2):115–120, 2015.

[247] Jeong Ho, Mamoru Komatsu, Kazuo Shin-ya, Satoshi Omura, and Haruo Ikeda. Distribution and functional analysis of the phosphopantetheinyl transferase superfamily in Actinomycetales microorganisms. *Proceedings of the National Academy of Sciences*, 115(26):6828–6833, 2018.

[248] Blaine A. Pfeifer, Suzanne J. Admiraal, Hugo Gramajo, David E. Cane, and Chaitan Khosla. Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. coli. *Science*, 291(5509):1790–1792, March 2001.

[249] Alexandra A Roberts, Janine N Copp, Mohamed A Marahiel, and Brett A Neilan. The Synechocystis sp. PCC6803 Sfp-Type Phosphopantetheinyl Transferase Does Not Possess Characteristic Broad-Range Activity. *ChemBioChem*, 6140:1869–1877, 2009.

[250] Guang Yang, Yi Zhang, Nicholas K. Lee, Monica A. Cozad, Sara E. Kearney, Hendrik Luesch, and Yousong Ding. Cyanobacterial Sfp-type phosphopantetheinyl transferases functionalize carrier proteins of diverse biosynthetic pathways. *Scientific Reports*, 7(1):11888, December 2017.

[251] Peter Meiser and Rolf Müller. Two Functionally Redundant Sfp-Type 4 '-Phosphopantetheinyl Transferases Differentially Activate Biosynthetic Pathways in Myxococcus xanthus. *ChemBioChem*, 9:1549–1553, 2008.

[252] Frank Gross, Daniela Gottschalk, and Rolf Müller. Posttranslational modification of myxobacterial carrier protein domains in Pseudomonas sp . by an intrinsic phosphopantetheinyl transferase. *Applied Microbiology and Biotechnology*, 68:66–74, 2005.

[253] Rebecca A. Schomer and Michael G. Thomas. Characterization of the Functional Variance in MbtH-like Protein Interactions with a Nonribosomal Peptide Synthetase. *Biochemistry*, 56(40):5380–5390, October 2017.

[254] Shogo Mori, Keith D Green, Ryan Choi, Garry W Buchko, and Michael G Fried. Using MbtH-Like Proteins to Alter the Substrate Profile of a Nonribosomal Peptide Adenylation Enzyme. *ChemBioChem*, 19(20):2186–2194, 2018.

[255] Karla J. Esquilín-Lebrón, Tye O. Boynton, Lawrence J. Shimkets, and Michael G. Thomas. An Orphan MbtH-Like Protein Interacts with Multiple Nonribosomal Peptide Synthetases in Myxococcus xanthus DK1622. *Journal of Bacteriology*, 200(21):e00346–18, August 2018.

[256] Chien Hung Yu, Yunkun Dang, Zhipeng Zhou, Cheng Wu, Fangzhou Zhao, Matthew S. Sachs, and Yi Liu. Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Molecular Cell*, 59(5):744–754, 2015.

[257] Supratim Mukherjee, Dimitri Stamatis, Jon Bertsch, Galina Ovchinnikova, Hema Y Katta, Alejandro Mojica, I-Min A Chen, Nikos C Kyrpides, and TBK Reddy. Genomes OnLine database (GOLD) v.7: updates and new features. *Nucleic Acids Research*, 47(October 2018):649–659, 2018.

[258] Kai Blin, Victòria Pascal Andreu, Emmanuel L C de los Santos, Francesco Del Carratore, Sang Yup Lee, Marnix H Medema, and Tilmann Weber. The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters. *Nucleic Acids Research*, 47(D1):D625–D630, 2018.

[259] Biao Tang, Yucong Yu, Junheng Liang, Youming Zhang, Xiaoying Bian, Xiaoyang Zhi, and Xiaoming Ding. Reclassification of 'Polyangium brachysporum' DSM 7029 as Schlegelella brevitalea sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 69(9):2877–2883, September 2019.

[260] Xiaoying Bian, Biao Tang, Yucong Yu, Qiang Tu, Frank Gross, Hailong Wang, Aiying Li, Jun Fu, Yuemao Shen, Yue-zhong Li, A Francis Stewart, Guoping Zhao, Xiaoming Ding, Rolf Müller, and Youming Zhang. Heterologous Production and Yield Improvement of Epothilones in Burkholderiales Strain DSM 7029. *ACS Chemical Biology*, 12(7):1805–1812, July 2017.

[261] Chenlang Liu, Fangnan Yu, Qingshu Liu, Xiaoying Bian, Shengbiao Hu, Huansheng Yang, Yulong Yin, Yuezhong Li, Yuemao Shen, Liqiu Xia, Qiang Tu, and Youming Zhang. Yield improvement of epothilones in Burkholderia strain DSM7029 via transporter engineering. *FEMS Microbiology Letters*, 365(9):fny045, May 2018.

[262] Fu Yan, David Auerbach, Yi Chai, Lena Keller, Qiang Tu, Stephan Hüttel, Amelie Glemser, Hanusch A. Grab, Thorsten Bach, Youming Zhang, and Rolf Müller. Biosynthesis and Heterologous Production of Vioprolides: Rational Biosynthetic Engineering and Unprecedented 4-Methylazetidinecarboxylic Acid Formation. *Angewandte Chemie - International Edition*, 57(28):8754–8759, 2018.

[263] Peter Cimermancic, Marnix H Medema, Jan Claesen, Kenji Kurita, Laura C Wieland Brown, Konstantinos Mavrommatis, Amrita Pati, Paul A Godfrey, Michael Koehrsen, Jon Clardy, Bruce W Birren, Eriko Takano, Andrej Sali, Roger G Linington, and Michael A Fischbach. Insights into Secondary Metabolism from a Global Analysis of Prokaryotic Biosynthetic Gene Clusters. *Cell*, 158(2):412–421, 2014.

[264] Sylvia Kunakom and Alessandra S Eusta. Burkholderia as a Source of Natural Products. *Journal of Natural Products*, 2018.

[265] Joyce Liu, Xuejun Zhu, Seong Jong Kim, and Wenjun Zhang. Antimycin-type depsipeptides: Discovery, biosynthesis, chemical synthesis, and bioactivities. *Natural Product Reports*, 33(10):1146–1165, 2016.

[266] Shie-pon Tzung, Kristine M Kim, Gorka Basañez, Chris D Giedt, Julian Simon, Joshua Zimmerberg, Kam Y J Zhang, and David M Hockenbery. Antimycin A mimics a cell-death-inducing Bcl2 homology domain 3. *Nature Cell Biology*, 3(February):183–191, 2001.

[267] Brian O Bodemann and Michael A White. Ras GTPases : Codon Bias Holds KRas Down but Not Out. *Current Biology*, 23(1):R17–R20, 2013.

[268] Shouliang Yin, Zilong Li, Xuefeng Wang, Huizhuan Wang, Xiaole Jia, Guomin Ai, Zishang Bai, Mingxin Shi, Fang Yuan, Tiejun Liu, Weishan Wang, and Keqian Yang. Heterologous expression of oxytetracycline biosynthetic gene cluster in

Streptomyces venezuelae WVR2006 to improve production level and to alter fermentation process. *Applied Microbiology and Biotechnology*, 100(24):10563–10572, 2016.

[269] I Chopra and M Roberts. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiology and molecular biology reviews : MMBR*, 65(2):232–60 ; second page, table of contents, June 2001.

[270] Will Skyrud, Joyce Liu, Divya Thankachan, Maria Cabrera, Ryan F Seipke, and Wenjun Zhang. Biosynthesis of the 15-Membered Ring Depsipeptide Neoantimycin. *ACS CHEMICAL BIOLOGY*, 13(5):1398–1406, May 2018.

[271] Yongjun Zhou, Xiao Lin, Simon R Williams, Liyun Liu, Yaoyao Shen, Shu-Ping Wang, Fan Sun, Shihai Xu, Hai Deng, Peter F Leadlay, and Hou-Wen Lin. Directed Accumulation of Anticancer Depsipeptides by Characterization of Neoantimycins Biosynthetic Pathway and an NADPH-Dependent Reductase. *ACS CHEMICAL BIOLOGY*, 13(8):2153–2160, August 2018.

[272] Wenhua Zhou, Li Hu, Liming Ying, Zhen Zhao, Paul K. Chu, and Xue-Feng Yu. A CRISPR–Cas9-triggered strand displacement amplification method for ultrasensitive DNA detection. *Nature Communications*, 9(1):5012, 2018.

[273] Lei Fang, Marc Guell, George M Church, and Blaine A Pfeifer. Heterologous erythromycin production across strain and plasmid construction. *BIOTECHNOLOGY PROGRESS*, 34(1):271–276, 2018.

[274] Christian Greunke, Elke Regina Duell, Paul Michael D Agostino, Anna Glöckle, Katharina Lamm, Tobias Alexander, and Marius Gulder. Direct Pathway Cloning (DiPaC) to unlock natural product biosynthetic potential. *Metabolic Engineering*, 47(March):334–345, 2018.

[275] Ran Peng, Ye Wang, Wan Feng, Xin Yue, Jiang Chen, Xiao Hu, and Zhi Li. CRISPR/dCas9-mediated transcriptional improvement of the biosynthetic gene cluster for the epothilone production in Myxococcus xanthus. *Microbial Cell Factories*, 17:15, 2018.

[276] Andreas Domröse, Robin Weihmann, Stephan Thies, Karl-Erich Jaeger, Thomas Drepper, and Anita Loeschcke. Rapid generation of recombinant Pseudomonas putida secondary metabolite producers using yTREX. *Synthetic and Systems Biotechnology*, 2(4):310–319, December 2017.

[277] Sung Soo Kim. Building Triketide alpha-Pyrone-Producing Yeast Platform Using Heterologous Expression of Sporopollenin Biosynthetic Genes. *Journal of microbiology and biotechnology*, 25(11):1796–1800, November 2015.

[278] Guang Yang, Monica A Cozad, Destin A Holland, Yi Zhang, Hendrik Luesch, and Yousong Ding. Photosynthetic Production of Sunscreen Shinorine Using an Engineered Cyanobacterium. *ACS Synthetic Biology*, 7(2):664–671, February 2018.

[279] Ken-ichi Harada, Kiyonaga Fujii, Takayuki Shimada, Makoto Suzuki, and Wayne W Carmichael. Two Cyclic Peptides, Anabaenopeptins, A Third Group of Bioactive Compounds from the Cyanobacterium Anabaena flos-aquae NRC 525-17. *Tetrahedron Letters*, 36(9):1511–1514, 1995.

[280] Lisa Spoof, Agata Błaszczyk, Jussi Meriluoto, Marta Cegłowska, and Hanna Mazur-Marzec. Structures and Activity of New Anabaenopeptins Produced by Baltic Sea Cyanobacteria. *Marine drugs*, 14(1):8, December 2015.

[281] Wallace E Herrell and William E Wellman. Erythromycin: A Microbial and Clinical Perspective After 30 Years of Clinical Use. *Mayo Clinic Proceedings*, 60:189–203, 1985.

[282] Leo Rouhiainen, Jouni Jokela, David P Fewer, Marina Urmann, and Kaarina Sivonen. Two Alternative Starter Modules for the Non- Ribosomal Biosynthesis of Specific Anabaenopeptin Variants in Anabaena (Cyanobacteria). *Chemistry & Biology*, 17(3):265–273, 2010.

[283] Tatyana Ammosova, Marina Jerebtsova, Monique Beullens, Yaroslav Voloshin, Patricio E Ray, Ajit Kumar, Mathieu Bollen, and Sergei Nekhai. Nuclear Protein Phosphatase-1 Regulates HIV-1 Transcription. *The Journal of Biological Chemistry*, 278(34):32189–32194, 2003.

[284] Haoran Zhang, Yong Wang, Jiequn Wu, Karin Skalina, and Blaine A. Pfeifer. Complete biosynthesis of erythromycin A and designed analogs using E. coli as a heterologous host. *Chemistry and Biology*, 17(11):1232–1240, 2010.

[285] Gergana A Vandova, Robert V O'Brien, Brian Lowry, Thomas F Robbins, Curt R Fischer, Ronald W Davis, Chaitan Khosla, Colin JB Harvey, and Maureen E Hillenmeyer. Heterologous expression of diverse propionyl-CoA carboxylases affects polyketide production in Escherichia coli. *JOURNAL OF ANTIBIOTICS*, 70(7):859–863, July 2017.

[286] Ming Jiang and Blaine A Pfeifer. Metabolic and pathway engineering to influence native and altered erythromycin production through E. coli. *METABOLIC ENGINEERING*, 19:42–49, September 2013.

[287] Daniel Mendez-Perez, Suman Gunasekaran, Victor J. Orler, and Brian F. Pfleger. A translation-coupling DNA cassette for monitoring protein translation in Escherichia coli. *Metabolic Engineering*, 14(4):298–305, 2012.

[288] P. Fumoleau, B. Coudert, N. Isambert, and E. Ferrant. Novel tubulin-targeting agents: Anticancer activity and pharmacologic profile of epothilones and related analogues. *Annals of Oncology*, 18(SUPPL. 5):9–15, 2007.

[289] Bryan Julien and Sanjay Shah. Heterologous Expression of Epothilone Biosynthetic Genes in Myxococcus xanthus. *Antimicrobial Agents and Chemotherapy*, 46(9):2772–2778, 2002.

[290] Xin-jing Yue, Xiao-wen Cui, Zheng Zhang, Wei-feng Hu, Zhi-feng Li, You-ming Zhang, and Yue-zhong Li. Effects of transcriptional mode on promoter substitution and tandem engineering for the production of epothilones in Myxococcus xanthus. *Applied Microbiology and Biotechnology*, 102:5599–5610, 2018.

[291] Corina Oßwald, Gregor Zipf, Gisela Schmidt, Josef Maier, Hubert S. Bernauer, Rolf Müller, and Silke C. Wenzel. Modular construction of a functional artificial epothilone polyketide pathway. *ACS Synthetic Biology*, 3(10):759–772, 2014.

[292] Jana Kumpfmueller, Karen Methling, Lei Fang, Blaine A Pfeifer, Michael Lalk, and Thomas Schweder. Production of the polyketide 6-deoxyerythronolide B in the heterologous host Bacillus subtilis. *APPLIED MICROBIOLOGY AND BIOTECHNOLOGY*, 100(3):1209–1220, February 2016.

[293] Michael Teuber and Johann Bader. Action of Polymyxin B on Bacterial Membranes. *Archives of Microbiology*, 109(1-2):51–58, 1976.

[294] M. Koike, K. Iida, and T. Matsuo. Electron Microscopic Studies on Mode of Action of Polymyxin. *Journal of Bacteriology*, 97(1):448–452, 1969.

[295] Se-Yu Kim, Soo-Young Park, Soo-Keun Choi, and Seung-Hwan Park. Biosynthesis of Polymyxins B, E, and P Using Genetically Engineered Polymyxin Synthetases in the Surrogate Host Bacillus subtilis. *JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY*, 25(7):1015–1025, July 2015.

[296] Mini-review O A Koksharova. Genetic tools for cyanobacteria. *Applied Microbiology and Biotechnology*, 58:123–137, 2002.

[297] Alakananda Basu, Alan P. Kozikowski, and John S. Lazo. Structural Requirements of Lyngbyatoxin A for Activation and Downregulation of Protein Kinase C. *Biochemistry*, 31:3824–3830, 1992.

[298] Takafumi Suzuki, Hozumi Motohashi, and Masayuki Yamamoto. Toward clinical application of the Keap1 – Nrf2 pathway. *Trends in Pharmacological Sciences*, 34(6):340–346, 2013.

[299] Ranko Gacesa, Karl P Lawrence, Nikolaos D Georgakopoulos, Kazuo Yabe, Walter C Dunlap, David J Barlow, Geoffrey Wells, Antony R Young, and Paul F Long. The mycosporine-like amino acids porphyra-334 and shinorine are antioxidants and direct antagonists of Keap1-Nrf2 binding. *Biochimie*, 154:35–44, 2018.

[300] Silke C. Wenzel, Frank Gross, Youming Zhang, Jun Fu, A. Francis Stewart, and Rolf Müller. Heterologous Expression of a Myxobacterial Natural Products Assembly Line in Pseudomonads via Red/ET Recombineering. *Chemistry & Biology*, 12(3):349–356, March 2005.

[301] Yi Chai, Shiping Shan, Kira J. Weissman, Shengbiao Hu, Youming Zhang, and Rolf Müller. Heterologous Expression and Genetic Engineering of the Tubulysin Biosynthetic Gene Cluster Using Red/ET Recombineering and Inactivation Mutagenesis. *Chemistry & Biology*, 19(3):361–371, March 2012.

[302] Abigail K P Harris, Neil R. Williamson, Holly Slater, Anthony Cox, Sophia Abbasi, Ian Foulds, Henrik T. Simonsen, Finian J. Leeper, and George P C Salmond. The Serratia gene cluster encoding biosynthesis of the red antibiotic, prodigiosin, shows species- and strain-dependent genome context variation. *Microbiology*, 150(11):3547–3560, November 2004.

[303] Tjaša Danevčič, Maja Borić Vezjak, Maša Zorec, and David Stopar. Prodigiosin - A Multifaceted Escherichia coli Antimicrobial Agent. *PLOS ONE*, 11(9):e0162412, September 2016.

[304] Jing Zhang, Yaling Shen, Jianwen Liu, and Dongzhi Wei. Antimetastatic effect of prodigiosin through inhibition of tumor invasion. *Biochemical Pharmacology*, 69(3):407–414, February 2005.

[305] Vanessa Soto-cerrato, Esther Llagostera, Beatriz Montaner, George L Scheffer, and Ricardo Perez-tomas. Mitochondria-mediated apoptosis operating irrespective of multidrug resistance in breast cancer cells by the anticancer agent prodigiosin. *Biochemical Pharmacology*, 68:1345–1352, 2004.

[306] Sang-bae Han, Chang Woo Lee, Yeo Dae Yoon, Jong Soon Kang, Ki Hoon Lee, Won Kee Yoon, Young Kook Kim, Kiho Lee, Song-kyu Park, and Hwan Mook Kim. Effective prevention of lethal acute graft-versus-host disease by combined immunosuppressive therapy with prodigiosin and cyclosporine A. *Biochemical Pharmacology*, 70(10):1518–1526, November 2005.

[307] Anita Loeschcke, Annette Markert, Susanne Wilhelm, Astrid Wirtz, Frank Rosenau, Karl Erich Jaeger, and Thomas Drepper. TREX: A universal tool for the transfer and expression of biosynthetic pathways in bacteria. *ACS Synthetic Biology*, 2(1):22–33, 2013.

[308] Jia Jia Zhang, Xiaoyu Tang, Tao Huan, Avena C Ross, and Bradley S Moore. Pass-back chain extension expands multimodular assembly line biosynthesis. *Nature Chemical Biology*, 16(1):42–49, January 2020.

[309] Gaoyan Wang, Zhiying Zhao, Jing Ke, Yvonne Engel, Yi-ming Shi, David Robinson, Kerem Bingol, Zheyun Zhang, Benjamin Bowen, Katherine Louie, Bing Wang, Robert Evans, Yu Miyamoto, Kelly Cheng, Suzanne Kosina, Markus De Raad, Leslie Silva, Alicia Luhrs, Andrea Lubbe, David W Hoyt, Charles Francavilla, Hiroshi Otani, Samuel Deutsch, Nancy M Washton, Edward M Rubin, Nigel J Mouncey, Axel Visel, Trent Northen, Jan-fang Cheng, Helge B Bode, and Yasuo Yoshikuni. CRAGE enables rapid activation of biosynthetic gene clusters in undomesticated bacteria. *Nature Microbiology*, 4(12):2498–2510, December 2019.

[310] Dongsoo Yang, Won Jun Kim, Seung Min Yoo, Jong Hyun Choi, Shin Hee Ha, Mun Hee Lee, and Sang Yup Lee. Repurposing type III polyketide synthase as a malonyl-CoA biosensor for metabolic engineering in bacteria. *Proceedings of the National Academy of Sciences*, 115(40):9835–9844, October 2018.

[311] Matthew R. Incha, Mitchell G. Thompson, Jacquelyn M. Blake-Hedges, Yuzhong Liu, Allison N. Pearson, Matthias Schmidt, Jennifer W. Gin, Christopher J. Petzold, Adam M. Deutschbauer, and Jay D. Keasling. Leveraging host metabolism for bisdemethoxycurcumin production in Pseudomonas putida. *Metabolic Engineering Communications*, 10(August 2019):e00119, June 2020.

[312] Hsien-chung Tseng and Kristala L J Prather. Controlled biosynthesis of odd-chain fuels and chemicals via engineered modular metabolic pathways. *Proceedings of the National Academy of Sciences*, 109(44):17925–17930, 2012.

[313] Ying-Ja Chen, Peng Liu, Alec A K Nielsen, Jennifer A N Brophy, Kevin Clancy, Todd Peterson, and Christopher A Voigt. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nature Methods*, 10(7):659–664, 2013.

[314] Kazuya Yamanaka, Kirk A Reynolds, Roland D Kersten, Katherine S Ryan, David J Gonzalez, Victor Nizet, Pieter C Dorrestein, and Bradley S Moore. Direct cloning and refactoring of a silent lipopeptide biosynthetic gene cluster yields the antibiotic taromycin A. *Proceedings of the National Academy of Sciences*, 111(5):1957–1962, February 2014.

[315] Pablo I. Nikel, Max Chavarría, Antoine Danchin, and Víctor de Lorenzo. From dirt to industrial applications: Pseudomonas putida as a Synthetic Biology chassis for hosting harsh biochemical reactions, 2016.

[316] Anita Loeschcke and Stephan Thies. Pseudomonas putida—a versatile host for the production of natural products. *Applied Microbiology and Biotechnology*, pages 6197–6214, 2015.

[317] Linde F C Kampers, Rita J M Volkers, and Vitor A P Martins. Pseudomonas putida KT2440 is HV1 certified, not GRAS. *Microbial Biotechnology*, 2019.

[318] K. E. Nelson, C. Weinel, I. T. Paulsen, R. J. Dodson, H. Hilbert, V. A P Martins dos Santos, D. E. Fouts, S. R. Gill, M. Pop, M. Holmes, L. Brinkac, M. Beanan, R. T. DeBoy, S. Daugherty, J. Kolonay, R. Madupu, W. Nelson, O. White, J. Peterson, H. Khouri, I. Hance, P. Chris Lee, E. Holtzapple, D. Scanlan, K. Tran, A. Moazzez, T. Utterback, M. Rizzo, K. Lee, D. Kosack, D. Moestl, H. Wedler, J. Lauber, D. Stjepandic, J. Hoheisel, M. Straetz, S. Heim, C. Kiewitz, J. Eisen, K. N. Timmis, A. Dusterhoft, B. Tummler, and C. M. Fraser. Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440. *Environmental Microbiology*, 4(12):799–808, December 2002.

[319] John Davison. Genetic Tools for Pseudomonads, Rhizobia, and Other Gram-Negative Bacteria. *BioTechniques*, 32(February):386–401, 2002.

[320] Jana Hoffmann and Josef Altenbuchner. Functional characterization of the mannitol promoter of Pseudomonas fluorescens DSM 50106 and its application for a mannitol-inducible expression system for Pseudomonas putida KT2440. *PLoS ONE*, 10(7):1–22, 2015.

[321] Aayushi Jain and Preeti Srivastava. Broad host range plasmids. *FEMS Microbiology Letters*, 348(2):87–96, 2013.

[322] M E Kovach, P H Elzer, D S Hill, G T Robertson, M A Farris, R M Roop, and K M Peterson. Four new derivatives of the broad host range cloning vector PBBR1MCS, carrying different antibiotic resistance cassettes. *Gene*, 166(1):175–176, 1995.

[323] U Kües and U Stahl. Replication of plasmids in gram-negative bacteria. *Microbiological reviews*, 53(4):491–516, 1989.

[324] Eliana R. De Bernardez and Prasad S. Dhurjati. Effect of a broad-host range plasmid on growth dynamics of Escherichia coli and Pseudomonas putida. *Biotechnology and Bioengineering*, 29(5):558–565, 1987.

[325] M M Bagdasarian, R Lurz, B Rückert, F C Franklin, M M Bagdasarian, J Frey, and K N Timmis. Specific-purpose plasmid cloning vectors II. Broad host range, high copy number, RSF1010-derived vectors, and a host-vector system for gene cloning in Pseudomonas. *Gene*, 16(1–3):237–247, 1981.

[326] R. Antoine and C. Locht. Isolation and molecular characterization of a novel broad-host-range plasmid from Bordetella bronchiseptica with sequence similarities to plasmids from Gram-positive organisms. *Molecular Microbiology*, 6:1785–1799, 1992.

[327] Rafael Silva-Rocha, Esteban Martínez-García, Belén Calles, Max Chavarría, Alejandro Arce-Rodríguez, Aitor de las Heras, A. David Páez-Espino, Gonzalo Durante-Rodríguez, Juhyun Kim, Pablo I. Nikel, Raúl Platero, and Víctor de Lorenzo. The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Research*, 41(D1):D666–D675, 2013.

[328] Jamie E. Prior, Michael D. Lynch, and Ryan T. Gill. Broad-host-range vectors for protein expression across gram negative hosts. *Biotechnology and Bioengineering*, 106(2):326–332, 2010.

[329] Luan Tao, Raymond E Jackson, and Qiong Cheng. Directed evolution of copy number of a broad host range plasmid for metabolic engineering. *Metabolic Engineering*, 7(1):10–17, 2005.

[330] N Itoh, T Kawanami, C Nitta, N Iwata, S Usami, Y Abe, and Y Koide. Characterization of pNI10 plasmid in Pseudomonas, and the construction of an improved Escherichia and Pseudomonas shuttle vector, pNUK73. *Appl Microbiol Biotechnol*, 61:240–246, 2003.

[331] Herbert P. Schweizer and Tung T. Hoang. An improved system for gene replacement and xylE fusion analysis in Pseudomonas aeruginosa. *Gene*, 158(1):15–22, 1995.

[332] Tung T. Hoang, Roxann R. Karkhoff-Schweizer, Alecksandr J. Kutchma, and Herbert P. Schweizer. A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: Application for isolation of unmarked Pseudomonas aeruginosa mutants. *Gene*, 212(1):77–86, 1998.

[333] Christopher J. Marx and Mary E. Lidstrom. Broad-host-range cre-lox system for antibiotic marker recycling in Gram-negative bacteria. *BioTechniques*, 33(5):1062–1067, 2002.

[334] Xi Luo, Yunwen Yang, Wen Ling, Hao Zhuang, Qin Li, and Guangdong Shang. Pseudomonas putida KT2440 markerless gene deletion using a combination of λ Red recombineering and Cre/loxP site-specific recombination. *FEMS microbiology letters*, 363(4), 2016.

[335] J Hashimoto, B Stevenson, and T M Schmidt. Rates and consequences of recombination between ribosomal RNA operons. *Journal of Bacteriology*, 185(3):966–972, 2002.

[336] Andreas Schäfer, Andreas Tauch, Wolfgang Jäger, Jörn Kalinowski, Georg Thierbach, and Alfred Pühler. Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene*, 145(1):69–73, 1994.

[337] Esteban Martínez-García and Víctor de Lorenzo. Engineering multiple genomic deletions in Gram-negative bacteria: Analysis of the multi-resistant antibiotic profile of Pseudomonas putida KT2440. *Environmental Microbiology*, 13(10):2702–2716, 2011.

[338] P Mali, K M Esvelt, and G M Church. Cas9 as a versatile tool for engineering biology. *Nat Methods*, 10(10):957–963, 2013.

[339] Wenyan Jiang, David Bikard, David Cox, Feng Zhang, and Luciano a Marraffini. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*, 31(3):233–239, 2013.

[340] Jee-Hwan Oh and Jan-Peter van Pijkeren. CRISPR-Cas9-assisted recombineering in Lactobacillus reuteri. *Nucleic acids research*, 42(25):1–11, 2014.

[341] Chai Lian Lee, Dave Siak Wei Ow, and Steve Kah Weng Oh. Quantitative real-time polymerase chain reaction for determination of plasmid copy number in bacteria. *Journal of Microbiological Methods*, 65(2):258–267, May 2006.

[342] Jean Paul Meijnen, Johannes H. De Winde, and Harald J. Ruijssenaars. Engineering Pseudomonas putida S12 for efficient utilization of D-xylose and L-arabinose. *Applied and Environmental Microbiology*, 74(16):5031–5037, 2008.

[343] Ross H. Durland, Aresa Toukdarian, Ferric Fang, and Donald R. Helinski. Mutations in the trfA replication gene of the broad-host-range plasmid RK2 result in elevated plasmid copy numbers. *Journal of Bacteriology*, 172(7):3859–3867, 1990.

[344] Christine Persson and Kurt Nordström. Control of replication of the broad host range plasmid RSF1010: The incompatibility determinant consists of directly repeated DNA sequences. *Molecular & General Genetics*, 203(1):189–192, 1986.

[345] An Xiao, Zhenchao Cheng, Lei Kong, Zuoyan Zhu, Shuo Lin, Ge Gao, and Bo Zhang. CasOT : a genome-wide Cas9 / gRNA off-target searching tool. *Bioinformatics*, 30(8):1180–1182, 2014.

[346] Jonathan R Beckwith, David Zipser, and Others. *The lactose operon*, volume 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1970.

[347] L McMurry, R E Petrucci, and S B Levy. Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, 77(7):3974–7, 1980.

[348] Koloi Rvbetzt. Regulation of the L-Arabinose Transport Operons in Escherichia coli. *Journal of molecular biology*, 151(1):215–227, 1981.

[349] Lars H. Hansen, Søren J. Sørensen, and Lars B. Jensen. Chromosomal insertion of the entire Escherichia coli lactose operon, into two strains of Pseudomonas, using a modified mini-Tn5 delivery system. *Gene*, 186(2):167–173, 1997.

[350] D a Siegele and J C Hu. Gene expression from plasmids containing the araBAD promoter at subsaturating inducer concentrations represents mixed populations. *Proceedings of the National Academy of Sciences of the United States of America*, 94(15):8168–8172, 1997.

[351] Jay D. Keasling, Barry L. Wanner, Tove Skaug, Kirill A. Datsenko, and Artem Khlebnikov. Homogeneous expression of the PBAD promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology*, 147(12):3241–3247, December 2001.

[352] Rachael M Morgan-Kiss, Caryn Wadler, and John E Cronan. Long-term and homogeneous regulation of the Escherichia coli araBAD promoter by use of a lactose transporter of relaxed specificity. *Proceedings of the National Academy of Sciences of the United States of America*, 99(11):7373–7, 2002.

[353] Vieira Jeffrey and Messing Joachim. New pUC-derived cloning vectors with different selectable markers and DNA replication origins. *Gene*, 100(C):189–194, 1991.

[354] E-Chiang Lee, Daiguan Yu, J. Martinez de Velasco, Lino Tessarollo, Deborah A. Swing, Donald L. Court, Nancy A. Jenkins, and Neal G. Copeland. A Highly Efficient Escherichia coli-Based Chromosome Engineering System Adapted for Recombinogenic Targeting and Subcloning of BAC DNA. *Genomics*, 73(1):56–65, 2001.

[355] William E. Bentley, Noushin Mirjalili, Dana C. Andersen, Robert H. Davis, and Dhinakar S. Kompala. Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnology and Bioengineering*, 35(7):668–681, 1990.

[356] R. H. Durland and D. R. Helinski. Replication of the broad-host-range plasmid RK2: Direct measurement of intracellular concentrations of the essential TrfA replication proteins and their effect on plasmid copy number. *Journal of Bacteriology*, 172(7):3849–3858, 1990.

[357] Y. Itoh, L. Soldati, T. Leisinger, and D. Haas. Low- and intermediate-copy-number cloning vectors based on the Pseudomonas plasmid pVS1. *Antonie van Leeuwenhoek*, 54(6):567–573, 1988.

[358] Tomás Aparicio, Sheila I. Jensen, Alex T. Nielsen, Victor de Lorenzo, and Esteban Martínez-García. The Ssr protein (T1E_1405) from Pseudomonas putida DOT-T1E enables oligonucleotide-based recombineering in platform strain P. putida EM42. *Biotechnology Journal*, 11(10):1309–1319, 2016.

[359] Biliana Lesic and Laurence G Rahme. Use of the lambda Red recombinase system to rapidly generate mutants in Pseudomonas aeruginosa. *BMC molecular biology*, 9:20, 2008.

[360] Bryan Swingle, Zhongmeng Bao, Eric Markel, Alan Chambers, and Samuel Cartinhour. Recombineering Using RecTE from Pseudomonas syringae. *Applied and Environmental Microbiology*, 76(15):4960–4968, 2010.

[361] Michael E. Pyne, Murray Moo-Young, Duane A. Chung, and C. Perry Chou. Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified Chromosomal Gene Replacement in Escherichia coli. *Applied and Environmental Microbiology*, 81(15):5103–5114, 2015.

[362] Nadja Graf and Josef Altenbuchner. Genetic engineering of Pseudomonas putida KT2440 for rapid and high-yield production of vanillin from ferulic acid. *Applied Microbiology and Biotechnology*, 98(1):137–149, 2014.

[363] Nicole Brendel, Laila P Partida-Martinez, Kirstin Scherlach, and Christian Hertweck. A cryptic PKS-NRPS gene locus in the plant commensal Pseudomonas fluorescens Pf-5 codes for the biosynthesis of an antimitotic rhizoxin complex. *Organic & biomolecular chemistry*, 5(September 2016):2211–2213, 2007.

[364] Yi Ming Chiang, Shu Lin Chang, Berl R. Oakley, and Clay C C Wang. Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Current Opinion in Chemical Biology*, 15(1):137–143, 2011.

[365] David J. Newman and Gordon M. Cragg. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *Journal of Natural Products*, 75(3):311–335, March 2012.

[366] Thomas A. Keating and Christopher T. Walsh. Initiation, elongation, and termination strategies in polyketide and polypeptide antibiotic biosynthesis. *Current Opinion in Chemical Biology*, 3(5):598–606, October 1999.

[367] Stephanie A. Vanner, Xiang Li, Rostyslav Zvanych, Jonathon Torchia, Jing Sang, David W. Andrews, and Nathan A. Magarvey. Chemical and biosynthetic evolution of the antimycin-type depsipeptides. *Molecular BioSystems*, 9(11):2712, 2013.

[368] Guoqing Niu. Genomics-Driven Natural Product Discovery in Actinomycetes. *Trends in Biotechnology*, 36(3):238–241, March 2018.

[369] Peter J Rutledge and Gregory L Challis. Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature Reviews Microbiology*, 13(8):509–523, August 2015.

[370] Stefan M Gaida, Nicholas R Sandoval, Sergios A Nicolaou, Yili Chen, Keerthi P Venkataramanan, and Eleftherios T Papoutsakis. Expression of heterologous sigma factors enables functional screening of metagenomic and heterologous genomic libraries. *Nature Communications*, 6(1):7045, November 2015.

[371] Domen Pogorevc, Ying Tang, Michael Hoffmann, Gregor Zipf, Hubert S Bernauer, Alexander Popoff, Heinrich Steinmetz, and Silke C. Wenzel. Biosynthesis and Heterologous Production of Argyrins. *ACS Synthetic Biology*, 8(5):1121–1133, May 2019.

[372] Ruidong Chen, Qingbo Zhang, Bin Tan, Liujuan Zheng, Huixian Li, Yiguang Zhu, and Changsheng Zhang. Genome Mining and Activation of a Silent PKS/N-RPS Gene Cluster Direct the Production of Totopotensamides. *Organic Letters*, 19(20):5697–5700, October 2017.

[373] Xiangyang Liu, Kangmin Hua, Dongxu Liu, Zhen-Long Wu, Ying Wang, Haoran Zhang, Zixin Deng, Blaine A. Pfeifer, and Ming Jiang. Heterologous Biosynthesis of Type II Polyketide Products Using E. coli. *ACS Chemical Biology*, 15(5):1177–1183, May 2020.

[374] Yanyan Li, Kira J. Weissman, and Rolf Müller. Insights into Multienzyme Docking in Hybrid PKS-NRPS Megasynthetases Revealed by Heterologous Expression and Genetic Engineering. *ChemBioChem*, 11(8):1069–1075, April 2010.

[375] Alexey Dudnik, Laurent Bigler, and Robert Dudler. Heterologous expression of a Photorhabdus luminescens syrbactin-like gene cluster results in production of the potent proteasome inhibitor glidobactin A. *Microbiological Research*, 168(2):73–76, February 2013.

[376] Masahisa Oka, Yuji Nishiyama, Shinichi Ohta, Hideo Kamei, Masataka Konishi, Takeo Miyaki, Toshikazu Oki, and Hiroshi Kawaguchi. Glidobactins A, B and C, new antitumor antibiotics. I. Production, isolation, chemical properties and biological activity. *The Journal of Antibiotics*, 41(10):1331–1337, 1988.

[377] M. L. Stein, P. Beck, M. Kaiser, R. Dudler, C. F. W. Becker, and M. Groll. One-shot NMR analysis of microbial secretions identifies highly potent proteasome inhibitor. *Proceedings of the National Academy of Sciences*, 109(45):18367–18371, November 2012.

[378] Neil R Williamson, Peter C Fineran, Finian J Leeper, and George P C Salmond. The biosynthesis and regulation of bacterial prodiginines. *Nature Reviews Microbiology*, 4(12):887–899, December 2006.

[379] Barbara Schellenberg, Laurent Bigler, and Robert Dudler. Identification of genes involved in the biosynthesis of the cytotoxic compound glidobactin from a soil bacterium. *Environmental Microbiology*, 9(7):1640–1650, July 2007.

[380] Qiang Tu, Jennifer Herrmann, Shengbiao Hu, Ritesh Raju, Xiaoying Bian, Youming Zhang, and Rolf Müller. Genetic engineering and heterologous expression of the disorazol biosynthetic gene cluster via Red/ET recombineering. *Scientific Reports*, 6(1):21066, February 2016.

[381] D. Riesenberg, V. Schulz, W.A. Knorre, H.-D. Pohl, D. Korz, E.A. Sanders, A. Roß, and W.-D. Deckwer. High cell density cultivation of Escherichia coli at controlled specific growth rate. *Journal of Biotechnology*, 20(1):17–27, August 1991.

[382] DG Gibson, Lei Young, and RY Chuang. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature ...*, 6(5):12–17, 2009.

[383] Kyoung-Hee Choi and Herbert P Schweizer. mini-Tn7 insertion in bacteria with single attTn7 sites: example Pseudomonas aeruginosa. *Nature Protocols*, 1(1):153–161, June 2006.

[384] Sankar Das, Jody C Noe, Sehmi Paik, and Todd Kitten. An improved arbitrary primed PCR method for rapid characterization of transposon insertion sites. *Journal of Microbiological Methods*, 63(1):89–94, October 2005.

[385] Michelle F Clasquin, Eugene Melamud, and Joshua D Rabinowitz. LC-MS Data Processing with MAVEN: A Metabolomic Analysis and Visualization Engine. In *Current Protocols in Bioinformatics*, volume 37, pages 14.11.1–14.11.23. John Wiley & Sons, Inc., Hoboken, NJ, USA, March 2012.

[386] Eugene Melamud, Livia Vastag, and Joshua D. Rabinowitz. Metabolomic Analysis and Visualization Engine for LC-MS Data. *Analytical Chemistry*, 82(23):9818–9826, December 2010.

[387] Xiaoying Bian, Fan Huang, Hailong Wang, Thorsten Klefisch, Rolf Müller, and Youming Zhang. Heterologous Production of Glidobactins/Luminmycins in Escherichia coli Nissle Containing the Glidobactin Biosynthetic Gene Cluster from Burkholderia DSM7029. *ChemBioChem*, 15(15):2221–2224, October 2014.

[388] Jia Mi, Anne Sydow, Florence Schempp, Daniela Becher, Hendrik Schewe, Jens Schrader, and Markus Buchhaupt. Investigation of plasmid-induced growth defect in Pseudomonas putida. *Journal of Biotechnology*, 231:167–173, August 2016.

[389] Juan Pablo Gomez-Escribano and Mervyn J Bibb. Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. *Microbial Biotechnology*, 4(2):207–215, March 2011.

[390] Kaushik Raj, Naveen Venayak, and Radhakrishnan Mahadevan. Novel two-stage processes for optimal chemical production in microbes. *Metabolic Engineering*, 62:186–197, November 2020.

[391] Sang Yup Lee, Heng Ho Wong, Jong-il Choi, Seung Hwan Lee, Sang Cheol Lee, and Chul Soo Han. Production of medium-chain-length polyhydroxyalkanoates by high-cell-density cultivation ofPseudomonas putida under phosphorus limitation. *Biotechnology and Bioengineering*, 68(4):466–470, May 2000.

[392] Thomas Eng, Deepanwita Banerjee, Andrew K. Lau, Emily Bowden, Robin A. Herbert, Jessica Trinh, Jan-Philip Prahl, Adam Deutschbauer, Deepti Tanjore, and Aindrila Mukhopadhyay. Engineering Pseudomonas putida for efficient aromatic conversion to bioproduct using high throughput screening in a bioreactor. *Metabolic Engineering*, 66:229–238, July 2021.

[393] Ana Ugidos, Gracia Morales, Eduardo Rial, Huw D. Williams, and Fernando Rojo. The coordinate regulation of multiple terminal oxidases by the Pseudomonas putida ANR global regulator. *Environmental Microbiology*, 10(7):1690–1702, July 2008.

[394] Pablo I Nikel, Juhyun Kim, and Víctor de Lorenzo. Metabolic and regulatory rearrangements underlying glycerol metabolism in Pseudomonas putida KT2440. *Environmental Microbiology*, 16(1):239–254, January 2014.

[395] Gracia Morales, Ana Ugidos, and Fernando Rojo. Inactivation of the Pseudomonas putida cytochrome o ubiquinol oxidase leads to a significant change in the transcriptome and to increased expression of the CIO and cbb3-1 terminal oxidases. *Environmental Microbiology*, 8(10):1764–1774, October 2006.

[396] Xiaoying Bian, Alberto Plaza, Youming Zhang, and Rolf Müller. Luminmycins A–C, Cryptic Natural Products from Photorhabdus luminescens Identified by Heterologous Expression in Escherichia coli. *Journal of Natural Products*, 75(9):1652–1655, September 2012.

[397] Heidi J Imker, Daniel Krahn, Jérôme Clerc, Markus Kaiser, and Christopher T Walsh. N-Acylation during Glidobactin Biosynthesis by the Tridomain Nonribosomal Peptide Synthetase Module GlbF. *Chemistry & Biology*, 17(10):1077–1083, October 2010.

[398] Rufeng Wang, Brady F Cress, Zheng Yang, John C Hordines, Gyoo Yeol Jung, Zhengtao Wang, and Mattheos A G Koffas. Design and characterization of biosensors for the screening of modular assembled naringenin biosynthetic library in Saccharomyces cerevisiae. *ACS Synthetic Biology*, 2019.

[399] Iman Farasat, Manish Kushwaha, Jason Collens, Michael Easterbrook, Matthew Guido, and Howard M Salis. Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Molecular Systems Biology*, 10(6):731, June 2014.

[400] Ariel Hecht, Jeff Glasgow, Paul R. Jaschke, Lukmaan A. Bawazer, Matthew S. Munson, Jennifer R. Cochran, Drew Endy, and Marc Salit. Measurements of translation initiation from all 64 codons in E. coli. *Nucleic Acids Research*, 45(7):3615–3626, April 2017.

[401] Joshua R. Elmore, Gara N. Dexter, Davinia Salvachúa, Jessica Martinez-Baird, E. Anne Hatmaker, Jay D. Huenemann, Dawn M. Klingeman, George L. Peabody, Darren J. Peterson, Christine Singer, Gregg T. Beckham, and Adam M. Guss. Production of itaconic acid from alkali pretreated lignin by dynamic two stage bioconversion. *Nature Communications*, 12(1):2261, December 2021.

[402] Auxiliadora Prieto, Isabel F Escapa, Virginia Martínez, Nina Dinjaski, Cristina Herencias, Fernando de la Peña, Natalia Tarazona, and Olga Revelles. A holistic view of polyhydroxyalkanoate metabolism in Pseudomonas putida. *Environmental Microbiology*, 18(2):341–357, February 2016.

[403] I. F. Escapa, C. del Cerro, J. L. García, and M A Prieto. The role of GlpR repressor in Pseudomonas putida KT2440 growth and PHA production from glycerol. *Environmental Microbiology*, 15(1):93–110, January 2013.

[404] Yasuhiro Igarashi, Martha E. Trujillo, Eustoquio Martínez-Molina, Saeko Yanase, Satoshi Miyanaga, Takamasa Obata, Hiroaki Sakurai, Ikuo Saiki, Tsuyoshi Fujita, and Tamotsu Furumai. Antitumor anthraquinones from an endophytic actinomycete Micromonospora lupini sp. nov. *Bioorganic & Medicinal Chemistry Letters*, 17(13):3702–3705, July 2007.

[405] Alexander O. Brachmann, Susan A. Joyce, Holger Jenke-Kodama, Gertrud Schwär, David J. Clarke, and Helge B. Bode. A type II polyketide synthase is responsible for anthraquinone biosynthesis in Photorhabdus luminescens. *ChemBioChem*, 8(14):1721–1728, 2007.

[406] JUDITH SCHIMANA, KLAUS GEBHARDT, ALEXANDRA HÖLTZEL, DIETMAR G. SCHMID, RODERICH SÜSSMUTH, JOHANNES MÜLLER, RÜDIGER PUKALL, and HANS-PETER FIEDLER. Arylomycins A and B, New Biaryl-bridged Lipopeptide Antibiotics Produced by Streptomyces sp. Tue 6075. I. Taxonomy, Fermentation, Isolation and Biological Activities. *The Journal of Antibiotics*, 55(6):565–570, 2002.

[407] Peter A Smith, Michael F T Koehler, Hany S Girgis, Donghong Yan, Yongsheng Chen, Yuan Chen, James J Crawford, Matthew R Durk, Robert I Higuchi, Jing Kang, Jeremy Murray, Prasuna Paraselli, Summer Park, Wilson Phung, John G Quinn, Tucker C Roberts, Lionel Rougé, Jacob B Schwarz, Elizabeth Skippington, John Wai, Min Xu, Zhiyong Yu, Hua Zhang, Man-Wah Tan, and Christopher E. Heise. Optimized arylomycins are a new class of Gram-negative antibiotics. *Nature*, 561(7722):189–194, September 2018.

[408] Jr. Earl B. Herr, Robert L. Hammill, and James M. McGuire. Capreomycin And Its Preparation, 1964.

[409] Haitian Fan, Adam B. Conn, Preston B. Williams, Stephen Diggs, Joseph Hahm, Howard B. Gamper, Ya-Ming Hou, Seán E. O'Leary, Yinsheng Wang, and Gregor M. Blaha. Transcription–translation coupling: direct interactions of RNA polymerase with ribosomes and ribosomal subunits. *Nucleic Acids Research*, 45(19):11043–11055, November 2017.

[410] I. Iost, J. Guillerez, and M. Dreyfus. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *Journal of Bacteriology*, 174(2):619–622, 1992.

[411] I. Iost and M. Dreyfus. The stability of Escherichia coli lacZ mRNA depends upon the simultaneity of its synthesis and translation. *The EMBO Journal*, 14(13):3252–3261, July 1995.

[412] Sarah C. Mutka, John R. Carney, Yaoquan Liu, and Jonathan Kennedy. Heterologous production of epothilone C and D in Escherichia coli. *Biochemistry*, 45(4):1321–1330, 2006.

[413] Alexander C Reis and Howard M Salis. An Automated Model Test System for Systematic Development and Improvement of Gene Expression Models. *ACS Synthetic Biology*, 9(11):3145–3156, November 2020.

[414] Yuriy Rebets, Elke Brötz, Bogdan Tokovenko, and Andriy Luzhetskyy. Actinomycetes biosynthetic potential: How to bridge in silico and in vivo? *Journal of Industrial Microbiology and Biotechnology*, 41(2):387–402, 2014.

[415] Yuyang Zhang, Hongping Chen, Yao Zhang, Huifang Yin, Chenyan Zhou, and Yan Wang. Direct RBS Engineering of the biosynthetic gene cluster for efficient productivity of violaceins in E. coli. *Microbial Cell Factories*, 20(1):38, December 2021.

[416] John W. Blunt, Anthony R. Carroll, Brent R. Copp, Rohan A. Davis, Robert A. Keyzers, and Michèle R. Prinsep. Marine natural products. *Natural Product Reports*, 35(1):8–53, 2018.

[417] Qinying Peng, Guixi Gao, Jin Lü, Qingshan Long, Xuefei Chen, Fei Zhang, Min Xu, Kai Liu, Yemin Wang, Zixin Deng, Zhiyong Li, and Meifeng Tao. Engineered Streptomyces lividans Strains for Optimal Identification and Expression of Cryptic Biosynthetic Gene Clusters. *Frontiers in Microbiology*, 9(December):1–15, December 2018.

[418] Juhyun Kim, Juan Carlos Oliveros, Pablo I. Nikel, Víctor de Lorenzo, and Rafael Silva-Rocha. Transcriptomic fingerprinting of Pseudomonas putida under alternative physiological regimes. *Environmental Microbiology Reports*, 5(6):883–891, December 2013.

[419] Jörg Stülke and Wolfgang Hillen. Carbon catabolite repression in bacteria. *Current Opinion in Microbiology*, 2(2):195–201, April 1999.

[420] John Villadsen, Jens Nielsen, and Gunnar Lidén. *Bioreaction engineering principles*. Springer Science & Business Media, 2011.

[421] G Eggink, P. Waard, and G.N.M. Huijberts. The role of fatty acid biosynthesis and degradation in the supply of substrates for poly(3-hydroxyalkanoate) formation in Pseudomonas putida. *FEMS Microbiology Letters*, 103(2-4):159–163, December 1992.

[422] M. Alejandro Dinamarca, Ana Ruiz-manzano, and Fernando Rojo. Inactivation of Cytochrome o Ubiquinol Oxidase Relieves Catabolic Repression of the Pseudomonas putida GPo1 Alkane Degradation Pathway. *Journal of Bacteriology*, 184(14):3785–3793, July 2002.

[423] L. Petruschka, G. Burchhardt, C. Müller, C. Weihe, and H. Herrmann. The cyo operon of Pseudomonas putida is involved in carbon catabolite repression of phenol degradation. *Molecular Genetics and Genomics*, 266(2):199–206, October 2001.

[424] Birgitta E Ebert, Felix Kurth, Marcel Grund, Lars M Blank, and Andreas Schmid. Response of Pseudomonas putida KT2440 to Increased NADH and ATP Demand. *Applied and Environmental Microbiology*, 77(18):6597–6605, September 2011.

[425] Chengjun Zhu, Christopher T. Nomura, Joseph A. Perrotta, Arthur J. Stipanovic, and James P. Nakas. Production and characterization of poly-3-hydroxybutyrate from biodiesel-glycerol by Burkholderia cepacia ATCC 17759. *Biotechnology Progress*, 26(2):NA–NA, 2009.

[426] Anca-Irina Galaction, Dan Cascaval, Corneliu Oniscu, and Marius Turnea. Enhancement of oxygen mass transfer in stirred bioreactors using oxygen-vectors. 1. Simulated fermentation broths. *Bioprocess and Biosystems Engineering*, 26(4):231–238, July 2004.

[427] I. F. Escapa, J. L. García, B. Bühler, L. M. Blank, and M. A. Prieto. The polyhydroxyalkanoate metabolism controls carbon and energy spillage in Pseudomonas putida. *Environmental Microbiology*, 14(4):1049–1063, April 2012.

[428] Adam M Guss, Joshua R Elmore, and Jay D Huenemann. Engineered microbes for conversion of organic compounds to medium chain length alcohols and methods of use, 2021.

[429] Kurt Throckmorton, Vladimir Vinnik, Ratul Chowdhury, Taylor Cook, Marc G. Chevrette, Costas Maranas, Brian Pfleger, and Michael George Thomas. Directed Evolution Reveals the Functional Sequence Space of an Adenylation Domain Specificity Code. *ACS Chemical Biology*, 14(9):2044–2054, September 2019.

[430] Vladimir Vinnik, Fan Zhang, Hyunjun Park, Taylor B. Cook, Kurt Throckmorton, Brian F. Pfleger, Tim S. Bugni, and Michael G. Thomas. Structural and Biosynthetic Analysis of the Fabrubactins, Unusual Siderophores from Agrobacterium fabrum Strain C58. *ACS Chemical Biology*, 16(1):125–135, January 2021.

[431] Iain L Lamont, Paul A Beare, Urs Ochsner, Adriana I Vasil, and Michael L Vasil. Siderophore-mediated signaling regulates virulence factor production in Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences of the United States of America*, 99(10):7072–7, 2002.

[432] Eric J Drake, Bradley R Miller, Ce Shi, Jeffrey T Tarrasch, Jesse A Sundlov, C. Leigh Allen, Georgios Skiniotis, Courtney C Aldrich, and Andrew M Gulick. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature*, 529(7585):235–238, 2016.

[433] Gavin E. Crooks. WebLogo: A Sequence Logo Generator. *Genome Research*, 14(6):1188–1190, May 2004.

[434] Jos Kramer, Özhan Özkaya, and Rolf Kümmerli. Bacterial siderophores in community and host interactions. *Nature Reviews*, 2019.

[435] Heath E. Klock, Eric J. Koesema, Mark W. Knuth, and Scott A. Lesley. Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. *Proteins: Structure, Function and Genetics*, 71(2):982–994, 2008.

[436] Rebecca A Schomer, Hyunjun Park, John J Barkei, and Michael G Thomas. Alanine Scanning of YbdZ, an MbtH-like Protein, Reveals Essential Residues for Functional Interactions with Its Nonribosomal Peptide Synthetase Partner EntF. *Biochemistry*, 57(28):4125–4134, July 2018.

[437] Elizabeth A. Felnagle, Angela M. Podevels, John J. Barkei, and Michael G. Thomas. Mechanistically Distinct Nonribosomal Peptide Synthetases Assemble the Structurally Related Antibiotics Viomycin and Capreomycin. *ChemBioChem*, 12(12):1859–1867, 2011.

[438] Eric Guenzi, Giuliano Galli, Ingeborg Grgurina, Dennis C Gross, and Guido Grandi. Characterization of the Syringomycin Synthetase Gene Cluster. *Journal of Biological Chemistry*, 273(49):32857–32863, December 1998.

[439] Wen Li, Hassan Rokni-Zadeh, Matthias De Vleeschouwer, Maarten G K Ghequire, Davy Sinnaeve, Guan-Lin Xie, Jef Rozenski, Annemieke Madder, José C. Martins, and René De Mot. The Antimicrobial Compound Xantholysin Defines a New Group of Pseudomonas Cyclic Lipopeptides. *PLoS ONE*, 8(5):e62946, May 2013.

[440] Shinya Kodani, Joanna Bicz, Lijiang Song, Robert J Deeth, Mayumi Ohnishi-Kameyama, Mitsuru Yoshida, Kozo Ochi, and Gregory L Challis. Structure and biosynthesis of scabichelin, a novel tris-hydroxamate siderophore produced by the plant pathogen Streptomyces scabies 87.22. *Organic & Biomolecular Chemistry*, 11(28):4686, 2013.

[441] Jürgen J. May, Thomas M. Wendrich, and Mohamed A. Marahiel. The dhb Operon of Bacillus subtilisEncodes the Biosynthetic Template for the Catecholic Siderophore 2,3-Dihydroxybenzoate-Glycine-Threonine Trimeric Ester Bacillibactin. *Journal of Biological Chemistry*, 276(10):7209–7217, March 2001.

[442] Pierre Cornelis and Sandra Matthijs. Diversity of siderophore-mediated iron uptake systems in fluorescent pseudomonads: not only pyoverdines. *Environmental Microbiology*, 4(12):787–798, December 2002.

[443] Regine Fuchs, Mathias Schafer, Valerie Geoffroy, and Jean-Marie Meyer. Siderotyping A Powerful Tool for the Characterization of Pyoverdines. *Current Topics in Medicinal Chemistry*, 1(1):31–57, May 2001.

[444] Gudrun Koch, Pol Nadal Jimenez, Remco Muntendam, Yixi Chen, Evelina Papaioannou, Stephan Heeb, Miguel Cámara, Paul Williams, Robbert H. Cool, and Wim J. Quax. The acylase PvdQ has a conserved function among fluorescent Pseudomonas spp. *Environmental Microbiology Reports*, 2(3):433–439, March 2010.

[445] Pol Nadal-Jimenez, Gudrun Koch, Carlos R. Reis, Remco Muntendam, Hans Raj, C. Margot Jeronimus-Stratingh, Robbert H. Cool, and Wim J. Quax. PvdP is a tyrosinase that drives maturation of the pyoverdine chromophore in Pseudomonas aeruginosa. *Journal of Bacteriology*, 196(14):2681–2690, 2014.

[446] Tania Henríquez, Nicola Victoria Stein, and Heinrich Jung. PvdRT-OpmQ and MdtABC-OpmB efflux systems are involved in pyoverdine secretion in Pseudomonas putida KT2440. *Environmental Microbiology Reports*, 11(2):98–106, April 2019.

[447] Mélissa Hannauer, Emilie Yeterian, Lois W. Martin, Iain L. Lamont, and Isabelle J. Schalk. An efflux pump is involved in secretion of newly synthesized siderophore by Pseudomonas aeruginosa. *FEBS Letters*, 584(23):4751–4755, December 2010.

[448] Carl Maximilian Hommelsheim, Lamprinos Frantzeskakis, Mengmeng Huang, and Bekir Ülker. PCR amplification of repetitive DNA: a limitation to genome editing technologies and many other applications. *Scientific Reports*, 4(1):5052, May 2015.

[449] Youming Zhang, Joep P.P. Muyrers, Giuseppe Testa, and A Francis Stewart. DNA cloning by homologous recombination in Escherichia coli. *Nature Biotechnology*, 18(12):1314–1317, December 2000.

[450] Jeffrey B Endelman, Jonathan J Silberg, Z.-G. Wang, and Frances H Arnold. Site-directed protein recombination as a shortest-path problem. *Protein Engineering Design and Selection*, 17(7):589–594, August 2004.

[451] Dominik A. Herbst, Björn Boll, Georg Zocher, Thilo Stehle, and Lutz Heide. Structural Basis of the Interaction of MbtH-like Proteins, Putative Regulators of Nonribosomal Peptide Biosynthesis, with Adenylating Enzymes. *Journal of Biological Chemistry*, 288(3):1991–2003, January 2013.

[452] Michelle R. Rondon, Katie S. Ballering, and Michael G. Thomas. Identification and analysis of a siderophore biosynthetic gene cluster from Agrobacterium tumefaciens C58. *Microbiology*, 150(11):3857–3866, 2004.

[453] Hans Beiderbeck, Kambiz Taraz, Herbert Budzikiewicz, and Anthony E. Walsby. Anachelin, the Siderophore of the Cyanobacterium Anabaena cylindrica CCAP 1403/2A. *Zeitschrift für Naturforschung C*, 55(9-10):681–687, October 2000.

[454] Man-Cheng Tang, Cheng-Yu Fu, and Gong-Li Tang. Characterization of SfmD as a Heme Peroxidase That Catalyzes the Regioselective Hydroxylation of 3-Methyltyrosine to 3-Hydroxy-5-methyltyrosine in. *The Journal of Biological Chemistry*, 287(7):5112–5121, 2012.

[455] María Elisa Pavan, Nancy I López, and M Julia Pettinari. Melanin biosynthesis in bacteria, regulation and production perspectives. *Applied Microbiology and Biotechnology*, 104(4):1357–1370, February 2020.

[456] Judy Savitskaya, Ryan J. Protzko, Francesca-Zhoufan Li, Adam P. Arkin, and John E. Dueber. Iterative screening methodology enables isolation of strains with improved properties for a FACS-based screen and increased L-DOPA production. *Scientific Reports*, 9(1):5815, December 2019.

[457] Ave Tooming-Klunderud, David P Fewer, Thomas Rohrlack, Jouni Jokela, Leo Rouhiainen, Kaarina Sivonen, Tom Kristensen, and Kjetill S Jakobsen. Evidence for positive selection acting on microcystin synthetase adenylation domains in three cyanobacterial genera. *BMC evolutionary biology*, 8:256, 2008.

[458] Aleksandra Wlodek, Steve G. Kendrew, Nigel J. Coates, Adam Hold, Joanna Pogwizd, Steven Rudder, Lesley S. Sheehan, Sarah J. Higginbotham, Anna E. Stanley-Smith, Tony Warneck, Mohammad Nur-E-Alam, Markus Radzom, Christine J. Martin, Lois Overvoorde, Markiyan Samborskyy, Silke Alt, Daniel Heine, Guy T. Carter, Edmund I. Graziani, Frank E. Koehn, Leonard McDonald, Alexander Alanine, Rosa María Rodríguez Sarmiento, Suzan Keen Chao, Hasane Ratni, Lucinda Steward, Isobel H. Norville, Mitali Sarkar-Tyson, Steven J. Moss, Peter F. Leadlay, Barrie Wilkinson, and Matthew A. Gregory. Diversity oriented biosynthesis via accelerated evolution of modular gene clusters. *Nature Communications*, 8(1):1206, December 2017.

[459] Diane M Retallack, Hongfan Jin, and Lawrence Chew. Reliable protein production in a Pseudomonas fluorescens expression system. *Protein Expression and Purification*, 81(2):157–165, 2012.

[460] Kyoung-Hee Choi, Jared B Gaynor, Kimberly G White, Carolina Lopez, Catharine M Bosio, RoxAnn R Karkhoff-Schweizer, and Herbert P Schweizer. A Tn7-based broad-range bacterial cloning and expression system. *Nature Methods*, 2(6):443–448, June 2005.

[461] Esteban Martínez-garcía and Víctor De Lorenzo. *Synthetic Gene Networks*, volume 813 of *Methods in Molecular Biology*. Humana Press, Totowa, NJ, 2012.

[462] Arjan de Groot, Alain Filloux, and Jan Tommassen. Conservation of xcp genes, involved in the two-step protein secretion process, in different Pseudomonas species and other gram-negative bacteria. *Molecular and General Genetics MGG*, 229(2):278–284, October 1991.

[463] Souvik Naskar, Michael Hohl, Matteo Tassinari, and Harry H. Low. The structure and mechanism of the bacterial type II secretion system. *Molecular Microbiology*, 115(3):412–424, March 2021.

[464] Florian Putker, Ria Tommassen-van Boxtel, Michiel Stork, José J. Rodríguez-Herva, Margot Koster, and Jan Tommassen. The type II secretion system (Xcp) of Pseudomonas putida is active and involved in the secretion of phosphatases. *Environmental Microbiology*, 15(10):2658–2671, 2013.

[465] Esteban Martínez-García, Pablo I Nikel, Tomás Aparicio, and Víctor de Lorenzo. Pseudomonas 2.0: genetic upgrading of P. putida KT2440 as an enhanced host for heterologous gene expression. *Microbial Cell Factories*, 13(1):159, 2014.

[466] Joseph Sambrook and David W Russell. Transformation of E. coli by Electroporation. *Cold Spring Harbor Protocols*, 2006(1):pdb.prot3933, June 2006.

[467] Francesco Imperi, Fabiola Ciccosanti, Ariel Basulto Perdomo, Federica Tiburzi, Carmine Mancone, Tonino Alonzi, Paolo Ascenzi, Mauro Piacentini, Paolo Visca, and Gian Maria Fimia. Analysis of the periplasmic proteome of Pseudomonas aeruginosa, a metabolically versatile opportunistic pathogen. *PROTEOMICS*, 9(7):1901–1915, April 2009.

[468] Susan M. Egan and Robert F. Schleif. A Regulatory Cascade in the Induction of rhaBAD. *Journal of Molecular Biology*, 234(1):87–98, November 1993.

[469] Ciarán L. Kelly, Zilei Liu, Akihide Yoshihara, Sarah F. Jenkinson, Mark R. Wormald, Jose Otero, Amalia Estévez, Atsushi Kato, Mikkel H.S. Marqvorsen, George W.J. Fleet, Ramón J. Estévez, Ken Izumori, and John T. Heap. Synthetic Chemical Inducers and Genetic Decoupling Enable Orthogonal Control of the rhaBAD Promoter. *ACS Synthetic Biology*, 5(10):1136–1145, 2016.

[470] J. Shine and L. Dalgarno. The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proceedings of the National Academy of Sciences*, 71(4):1342–1346, April 1974.

[471] Pablo I. Nikel, Max Chavarr??a, Tobias Fuhrer, Uwe Sauer, and V??ctor De Lorenzo. Pseudomonas putida KT2440 strain metabolizes glucose through a cycle formed by enzymes of the Entner-Doudoroff, embden-meyerhof-parnas, and pentose phosphate pathways. *Journal of Biological Chemistry*, 290(43):25920–25932, 2015.

[472] Filoklis D. Pileidis and Maria-Magdalena Titirici. Levulinic Acid Biorefineries: New Challenges for Efficient Utilization of Biomass. *ChemSusChem*, 9(6):562–582, March 2016.

[473] Mohammad G Al-Shaal, Adam Dzierbinski, and Regina Palkovits. Solvent-free γ-valerolactone hydrogenation to 2-methyltetrahydrofuran catalysed by Ru/C: a reaction network analysis. *Green Chem.*, 16(3):1358–1364, 2014.

[474] Wenxiang Xuan, Minna Hakkarainen, and Karin Odelius. Levulinic Acid as a Versatile Building Block for Plasticizer Design. *ACS Sustainable Chemistry and Engineering*, 7(14):12552–12562, 2019.

[475] Joseph J. Bozell, L. Moens, D.C Elliott, Y. Wang, G.G Neuenscwander, S.W Fitzpatrick, R.J Bilski, and J.L Jarnefeld. Production of levulinic acid and use as a platform chemical for derived products. *Resources, Conservation and Recycling*, 28(3-4):227–239, February 2000.

[476] Christopher R. Mehrer, Jacqueline M. Rand, Matthew R. Incha, Taylor B. Cook, Benginur Demir, Ali Hussain Motagamwala, Daniel Kim, James A. Dumesic, and Brian F. Pfleger. Growth-coupled bioconversion of levulinic acid to butanone. *Metabolic Engineering*, 55(April):92–101, September 2019.

[477] R.G. Tischer, C.R. Fellers, and B.J. Doyle. The Nontoxicity of Levulinic Acid. *Journal of the American Pharmaceutical Association (Scientific ed.)*, 31(7):217–220, July 1942.

[478] Stephanie R. Harris, Guo-Fang Zhang, Sushabhan Sadhukhan, Anne M. Murphy, Kristyen A. Tomcik, Edwin J. Vazquez, Vernon E. Anderson, Gregory P. Tochtrop, and Henri Brunengraber. Metabolism of Levulinate in Perfused Rat Livers and Live Rats. *Journal of Biological Chemistry*, 286(7):5895–5904, February 2011.

[479] Lawrence P Carter, Huifang Wu, Weibin Chen, Marilyn M Matthews, Ashok K Mehta, R Jason Hernandez, Jennifer A Thomson, Maharaj K Ticku, Andrew Coop, Wouter Koek, and Charles P. France. Novel γ-Hydroxybutyric Acid (GHB) Analogs Share Some, but Not All, of the Behavioral Effects of GHB and GABA B Receptor Agonists. *Journal of Pharmacology and Experimental Therapeutics*, 313(3):1314–1323, June 2005.

[480] Steen Kølvraa, Niels Gregersen, Ernst Christensen, and Ida Grøn. Calcium levulinate medication. A pitfall in the diagnosis of organic acidurias. *Clinica Chimica Acta*, 77(2):197–201, June 1977.

[481] Jerry Vockley. Defects of mitochondrial β-oxidation: a growing group of disorders. *Neuromuscular Disorders*, 12(3):235–246, March 2002.

[482] Miao He, Zhengtong Pei, Al Walid Mohsen, Paul Watkins, Geoffrey Murdoch, Paul P. Van Veldhoven, Regina Ensenauer, and Jerry Vockley. Identification and characterization of new long chain Acyl-CoA dehydrogenases. *Molecular Genetics and Metabolism*, 102(4):418–429, 2011.

[483] L. Bian, R. L. Hanson, Y. L. Muller, L. Ma, S. Kobes, W. C. Knowler, C. Bogardus, and L. J. Baier. Variants in ACAD10 are associated with type 2 diabetes, insulin resistance and lipid oxidation in Pima Indians. *Diabetologia*, 53(7):1349–1353, July 2010.

[484] Dadi Jiang, Edward L. LaGory, Daniela Kenzelmann Brož, Kathryn T. Bieging, Colleen A. Brady, Nichole Link, John M. Abrams, Amato J. Giaccia, and Laura D. Attardi. Analysis of p53 Transactivation Domain Mutants Reveals Acad11 as a Metabolic Target Important for p53 Pro-Survival Function. *Cell Reports*, 10(7):1096–1109, February 2015.

[485] Kristine M. Briedis, Ayelet Starr, and Philip E. Bourne. Analysis of the Human Kinome Using Methods Including Fold Recognition Reveals Two Novel Kinases. *PLoS ONE*, 3(2):e1597, February 2008.

[486] Wai-Ching Hon, Geoffrey A. McKay, Paul R. Thompson, Robert M. Sweet, Daniel S.C. Yang, Gerard D. Wright, and Albert M. Berghuis. Structure of an Enzyme Required for Aminoglycoside Antibiotic Resistance Reveals Homology to Eukaryotic Protein Kinases. *Cell*, 89(6):887–895, June 1997.

[487] Jose Juan Almagro Armenteros, Marco Salvatore, Olof Emanuelsson, Ole Winther, Gunnar Von Heijne, Arne Elofsson, and Henrik Nielsen. Detecting sequence signals in targeting peptides using deep learning. *Life Science Alliance*, 2(5):1–14, 2019.

[488] Jacqueline M Rand. *Elucidating Metabolic Pathways for Catabolizing and Upgrading Levulinic Acid in Bacteria*. PhD thesis, The University of Wisconsin - Madison, Madison, 2017.

[489] Shuichi FURUTA, Shoko MAYAZAWA, and Takeshi HASHIMOTO. Purification and Properties of Rat Liver Acyl-CoA Dehydrogenases and Electron Transfer Flavoprotein1. *The Journal of Biochemistry*, 90(6):1739–1750, October 1981.

[490] Eric S. Goetzman, Yudong Wang, Miao He, Al-Walid Mohsen, Brittani K. Ninness, and Jerry Vockley. Expression and characterization of mutations in human very long-chain acyl-CoA dehydrogenase using a prokaryotic system. *Molecular Genetics and Metabolism*, 91(2):138–147, June 2007.

[491] Edna Bode, Antje K. Heinrich, Merle Hirschmann, Desalegne Abebew, Yan-Ni Shi, Tien Duy Vo, Frank Wesche, Yi-Ming Shi, Peter Grün, Svenja Simonyi, Nadine Keller, Yvonne Engel, Sebastian Wenski, Reuel Bennet, Sophie Beyer, Iris Bischoff, Anthony Buaya, Sophie Brandt, Ibrahim Cakmak, Harun Çimen, Simone Eckstein, Denia Frank, Robert Fürst, Martin Gand, Gerd Geisslinger, Selcuk Hazir, Marina Henke, Ralf Heermann, Virginie Lecaudey, Wilhelm Schäfer, Susanne Schiffmann, Anja Schüffler, Rebecca Schwenk, Marisa Skaljac, Eckhard Thines, Marco Thines, Thomas Ulshöfer, Andreas Vilcinskas, Thomas A. Wichelhaus, and Helge B. Bode. Promoter Activation in Δhfq Mutants as an Efficient Tool for Specialized Metabolite Production Enabling Direct Bioactivity Testing. *Angewandte Chemie*, 131(52):19133–19139, December 2019.

[492] Melinda L Micallef, Paul M D Agostino, Deepti Sharma, Rajesh Viswanathan, and Michelle C Moffitt. Genome mining for natural product biosynthetic gene clusters in the Subsection V cyanobacteria. *BMC Genomics*, 16:669, 2015.

[493] Kai Blin, Simon Shaw, Katharina Steinke, Rasmus Villebro, Nadine Ziemert, Sang Yup Lee, Marnix H. Medema, and Tilmann Weber. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Research*, 47(W1):W81–W87, July 2019.

[494] Sarah E Ongley, Xiaoying Bian, Brett a Neilan, and Rolf Müller. Recent advances in the heterologous expression of microbial natural product biosynthetic pathways. *Natural Product Reports*, 30(8):1121, 2013.

[495] Seong-Hee Park, Kyusung Lee, Jae Woo Jang, and Ji-Sook Hahn. Metabolic Engineering of Saccharomyces cerevisiae for Production of Shinorine, a Sunscreen Material, from Xylose. *ACS Synthetic Biology*, 8(2):346–357, February 2019.

[496] Julie E. Chaves, Rosemarie Wilton, Yuqian Gao, Nathalie Munoz Munoz, Meagan C. Burnet, Zachary Schmitz, John Rowan, Leah H. Burdick, Joshua Elmore, Adam Guss, Dan Close, Jon K. Magnuson, Kristin E. Burnum-Johnson, and Joshua K. Michener. Evaluation of chromosomal insertion loci in the Pseudomonas putida KT2440 genome for predictable biosystems design. *Metabolic Engineering Communications*, 11(April):e00139, December 2020.

[497] Scott A Scholz, Rucheng Diao, Michael B Wolfe, Elayne M Fivenson, Xiaoxia Nina Lin, Peter L Freddolino, Scott A Scholz, Rucheng Diao, Michael B Wolfe, Elayne M Fivenson, and Xiaoxia Nina Lin. High-Resolution Mapping of the Escherichia coli Chromosome Reveals Positions of High and Low Transcription. *Cell Systems*, 8(3):212–225, 2019.

[498] Hualan Liu and Adam M Deutschbauer. Rapidly moving new bacteria to model-organism status. *Current Opinion in Biotechnology*, 51:116–122, June 2018.

[499] Kou-San Ju, Jiangtao Gao, James R. Doroghazi, Kwo-Kwang A. Wang, Christopher J. Thibodeaux, Steven Li, Emily Metzger, John Fudala, Joleen Su, Jun Kai Zhang, Jaeheon Lee, Joel P. Cioni, Bradley S. Evans, Ryuichi Hirota, David P. Labeda, Wilfred A. van der Donk, and William W. Metcalf. Discovery of phosphonic acid natural products by mining the genomes of 10,000 actinomycetes. *Proceedings of the National Academy of Sciences*, 112(39):12175–12180, 2015.

[500] Giovanna Pirri, Andrea Giuliani, Silvia Fabiole Nicoletto, Lorena Pizzuto, and Andrea C. Rinaldi. Lipopeptides as anti-infectives: a practical perspective. *Central European Journal of Biology*, 4(3):258–273, 2009.

[501] Ruiqi Huang, Frank Hippauf, Diana Rohrbeck, Maria Haustein, Katrin Wenke, Janie Feike, Noah Sorrelle, Birgit Piechulla, and Todd J. Barkman. Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates. *Proceedings of the National Academy of Sciences of the United States of America*, 109(8):2966–2971, 2012.

[502] Alexander Amatuni and Hans Renata. Identification of a lysine 4-hydroxylase from the glidobactin biosynthesis and evaluation of its biocatalytic potential. *Organic & Biomolecular Chemistry*, 17(7):1736–1739, 2019.