12-2020

Advancing a systems cell-free metabolic engineering approach to natural product synthesis and discovery

Benjamin Mohr
University of Tennessee

To the Graduate Council:

I am submitting herewith a dissertation written by Benjamin Mohr entitled "Advancing a systems cell-free metabolic engineering approach to natural product synthesis and discovery." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Energy Science and Engineering.

Mitchel Doktycz, Major Professor

We have read this dissertation and recommend its acceptance:

Jennifer Morrell-Falvey, Dale Pelletier, Michael Simpson, Robert Hettich

Accepted for the Council:

Dixie L. Thompson

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

Advancing a systems cell-free metabolic engineering approach to natural product synthesis and discovery

A Dissertation Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Benjamin Pintz Mohr

December 2019

I am dedicating this thesis to the incredible people who have supported me.

To my dad for encouraging me to follow my passions, to set high goals and for believing I could achieve them.

To my brother and the smartest person I know, Chris; your critical thinking and patient help have shaped this thesis and my approach to science.

And to my wife, Chelsea, who has only been my wife for 7 months, but has been my best friend for 10 years. Thank you for your constant encouragement and support. I love you.

Acknowledgments

I am extremely grateful and lucky to have had the opportunity to perform this work at the cutting edge facilities at Oak Ridge National Laboratory and with the helpful and talented scientists that work there. I would like to thank my doctoral committee for their continued support of my work and the field. I would also especially like to Dr. Amber B. Webb, Dr. Greg B. Hurst, and Dr. Richard J. Giannone for sharing their wisdom, technical expertise and patience. Finally, I'd like to thank my friends near and far for their understanding and their ability to nod along even when they don't.

Abstract

Next generation DNA sequencing has led to an accumulation of a putative biosynthetic gene clusters for many natural product classes of interest. *In vivo* extraction and heterologous expression do not have sufficient throughput to validate predicted enzyme functions and inform future annotations. Further, engineering the production of new natural products is laborious and limited by the trade-offs between cell growth and product synthesis. Conversely, cell-free platforms, particularly those capable of cell-free protein synthesis (CFPS), facilitate rapid screening of enzyme function and prototyping of metabolic pathways. However, the protein content and metabolic activity of many cell-free systems are poorly defined, increasing variability between lysates and impeding systematic engineering. Here, the strength of untargeted peptidomics as an enabling tool for the engineering of cell-free systems is established based upon its ability to measure both global protein abundances and newly synthesized peptides. Synthesis of peptide natural products was found to be more robust in purified enzyme CFPS systems compared to crude lysates; however, non-specific peptide degradation, detected through peptidomics, remains a concern. Crude cell-free systems were determined be better suited to small molecule production, due to the extensive metabolic networks they were found to possess. Perturbations of these networks, carried out through changes to growth media, were observed through shotgun proteomics and informed engineering of phenol biosynthesis in a crude *Escherichia coli* lysate. Implementing shotgun proteomics as an analytical tool for cell-free systems will increase reproducibility and further the development of a platform for high-throughput functional genomics and metabolic engineering.

Table of Contents

List of Tables

List of Figures

Supplemental Files

Supplemental Table A-1 (Supplemental Table A-1.xlsx). 21

Supplemental Table A-2 (Supplemental Table A-2.xlsx). 21

Supplemental Table A-3 (Supplemental Table A-3.xlsx). 25, 31

Supplemental Table B-1 (Supplemental Table B-1.xlsx). 41

Supplemental Table B-2 (Supplemental Table B-2.xlsx). 41

Supplemental Table B-3 (Supplemental Table B-3.csv). xiii, 46, 47

Supplemental Table B-4 (Supplemental Table B-4.csv). xiv, 48, 49

Supplemental Table B-5 (Supplemental Table B-5.csv). xv, 51, 52

Supplemental Table C-1 (Supplemental Table C-1.xlsx). xviii, 70, 85

Supplemental Table C-2 (Supplemental Table C-2.xlsx). 68, 75

Supplemental Table C-3 (Supplemental Table C-3.xlsx). 68

Chapter 1

Introduction

Small molecules extracted from plants, fungi and bacteria have long been valued for their medicinal and biochemical properties. Natural products (NPs) are the compounds derived from living sources and are among our most important natural resources. NPs account for or have inspired between 50-70% of all pharmaceuticals and are promising biofuels, bioplastics and pest control agents [1]. NPs were traditionally identified by activity screening, purification and structural determination through mass spectroscopy (MS) or nuclear magnetic resonance imaging (NMR) [2]. While low-throughput, this methodology successfully identified numerous NPs. However, genome sequencing has revealed that encoded in the DNA of plants and microbes are a wealth of biosynthetic blueprints to undiscovered medicines, fuels, foods and polymers.

It has been suggested that microbes produce a wide array of natural products so they can more readily respond to diverse stimuli [3, 4]. By changing growth conditions or providing stressors such as competing species, microbes can be coaxed to produce natural products [5, 6, 7]. Microbes often produce a cocktail of compounds in response to single stimuli, some of which may already be known. Rediscovery of NPs is a hinderance to prospecting efforts because previously discovered NPs are often produced in higher quantities and obfuscate the detection of more novel metabolites due to their similar physiochemical properties [8]. Enabled by the development of next-generation sequencing platforms, high quality genomes are becoming available with more microbes sequenced each year [9]. Combining this wealth of sequence data with *in vivo* and *in vitro* characterizations of enzyme activity has allowed

for structure and function prediction tools based on nucleotide and amino acid sequence homology like BLAST, pFAM, and HMMER3 to become commonplace [10, 11]. Genome mining and prediction are high-throughput but must be accompanied by experimental validation.

The development of flexible methods for genetic manipulation of plasmid and genomic DNA has aided the discovery of NPs and functional characterization of the enzyme which produce them. The rapidly decreasing cost of DNA synthesis provides access to the biosynthetic genes of difficult to culture organisms [12]. Assembly of these synthetic gene fragments by Gibson and Golden Gate assembly allow for optimization of codon usage and removal of inhibitory mRNA secondary structures [13, 14, 15, 16, 17]. Tools for transformation and integration of DNA into diverse organisms have increased the range of NPs accessible by heterologous expression [18, 19]. Furthermore, engineering of transcription factors has facilitated activation of cryptic biosynthetic gene clusters (BGCs); this is particularly of interest for expression of extremely large BGCs (>100 kB) that remain prohibitively expensive to synthesize [20, 21]. Through genetic engineering previously undescribed and "silent" BGCs can be expressed in both endogenous and heterologous systems [22].

In addition to its use as a tool for the exploration of BGCs, microbial metabolism can be engineered to sustainably produce NPs and biochemicals of importance [23]. Rational engineering strategies including enzyme engineering to remove feedback inhibition, and improvement of substrate supply through balancing gene expression and gene knockouts of competing pathways have grown the yield and titer of NP biosynthesis in microbes [24, 25]. Tools for prediction of optimal genetic manipulations have accelerated these efforts [26]. Additionally, stress-tolerant microbial hosts have been created by leveraging evolution for metabolic engineering [27, 28, 29]. Recent achievements include production of anti-malarial terpenoids in yeast, increasing yields of sustainable jet-fuel alternatives in bacteria and mixed culture fermentation of alkaloids [30, 31, 32].

Nevertheless, *in vivo* validation cannot keep pace with in silico predictions as genetic engineering and heterologous expression can be time-consuming even in model hosts. Furthermore, NP biosynthesis can be limited by lack of NP precursor availability, toxic

products and metabolic burden of heterologous pathways [33]. While, native NP producers can be engineered to activate silent BGCs, non-model genetic engineering tools lack robustness and many organisms remain recalcitrant to modification [34].

Cell-free systems are an alternative to the laborious engineering of organisms. The protein machinery from inside living cells, extracted by a combination of lysis, fractionation and purification, can be harnessed to carry out complex biosynthetic processes *in vitro*. Removal of the cell wall and membrane and endogenous transcription and translation renders the cell lysate as a pseudo-chemical reactor, a predominantly static network of chemical reactions able to be added to, sampled from, and modified without genetic manipulation. Traditionally referred to as crude lysates or cell extracts, the extracted machinery of animal and microbial cells were important tools in the study of enzyme function prior to genome sequencing.

Rennet, the bovine stomach extract used to curdle milk, is likely one of the earliest cell-free systems, dating back more than 6000 years [35]; rennet is now known to be primarily comprised of chymosin, which cleaves the highly charged kappa-caesin chain causing aggregation of uncharged micelles [36]. Early cell-free systems were uncharacterized mixtures of proteins and metabolites extracted from a living source, that performed chemical transformations; the identity of the numerous enzymes which collectively performed the observed biochemical functions were not know. However by analyzing the function of these lysates, through measurement of the consumption and production of small molecules, the enzymes of the Kreb's cycle, the mechanism of ethanolic fermentation in yeast and the nucleic acid code were discovered [37, 38, 39].

Cell-free systems proved valuable for capturing the native metabolic and biochemical functions of cellular machinery. However, without heterologous expression, the capability of discovering and characterizing new peptides, proteins and enzymes in crude lysates was limited. Activation of *in vitro* transcription and translation changed the landscape of cell-free systems. Developed by Spirin *et al.* more than two decades prior to the ideation of synthetic biology, the S100 cellular extract of *Escherichia coli* was capable of *in vitro* translation from an mRNA template [40]. Shortly thereafter, coupled *in vitro* transcription and translation was carried out from a heterologous DNA template in the crude S30 extract of *E. coli* [41].

Transcription in the S30 extract is generally carried out by the viral T7 polymerase; however, native *E. coli* polymerases and sigma factors have been activated as well [42, 43].

Over the following 30 years, through the tireless work of numerous researchers, the capabilities of CFPS have grown considerably. In 2001, the Ueda lab constructed a CFPS system from exclusively purified bacterial enzymes, which has been commercialized as the PURE system; this system recycles tRNAs and can produce protein at a rate of $160\,\mu g\,mL^{-1}\,h^{-1}$ [44]. Others have iterated upon this system increasing yield and decreasing the costs of purification [45, 46, 47]. While the S30 lysate has the highest batch reaction protein yield, CFPS has also been achieved across the kingdoms of life, including from wheat germ cells [48], tobacco BY-2 cells [49], Chinese hamster ovary cells [50] and insect cells [51]. One of the greatest advantages of performing CFPS in eukaryotes is the ability to perform post-translational modifications inaccessible to bacteria [52, 53]. With these advances, the modern CFPS toolbox can be used to synthesize membrane proteins, perform most post-translational modifications and reach yields of more than $2\,mg\,mL^{-1}$ in less than 8 hours [54]. Further, with the aid of enabling technologies in microfluidics and lyophilization, cell-free systems, in particular CFPS, are well-suited for remote and on-demand bioproduction as remarked upon in Chapter 2.

Development of cell-free systems has not been limited to *in vitro* protein synthesis, many cell-free systems have been engineered for small molecule production as well. The crude S30 cell lysate is not only the most robust bacterial CFPS system but has also been deployed as a platform for metabolic engineering. Cell-free metabolic engineering (CFME) leverages the *in vitro* metabolic activity of enzymes and crude lysates alongside the benefits of cell-free systems for small molecule production. The energy regeneration of the S30 CFPS system forms the basis of many CFME efforts. Early S30 systems were energized by regenerating ATP through consumption of a sacrificial creatine phosphate or phosphoenolpyruvate molecule [55, 56], but greater improvements in CFPS yield were achieved through activation of endogenous glycolytic activity producing ATP from glucose [57]. Coupling glycolysis in crude *E. coli* lysate with heterologous enzymes enriched prior to lysis or expressed by CFPS has enabled production of terpenoids, plastic precursors and biofuels [58, 59, 60]. Compared to living systems, removal of the genome shuts down much

of the cell's instructional programming and eliminates the need to cope with a continually growing and changing system. This permits biosynthesis pathways to be engineered *in vitro*, minimizing carbon and energy lost to growth.

Enabled by inexpensive synthetic DNA and heterologous expression, defined cell-free systems for small molecule production have been constructed from purified heterologous enzymes. Purified multi-enzyme reaction cascades have been used to synthesized terpenoids and biohydrogen [61, 62] with high turnover rates and conversion efficiencies. Furthermore, purified enzymes have been used to create new biochemical functions, that would be toxic *in vivo*, such as a molecular purge value that balances concentration of NAD^+ and $NADP^+$ [63]. In contrast to crude lysates, all metabolic activity in defined systems must be assembled from purified enzymes. While they have many advantages, the size and number of metabolic pathways that are feasible to simultaneously activate in a defined system is strictly limited by the expensive and time-consuming purification of its protein components [64].

That lack of this requirement is one of the greatest strengths of crude cell-free systems; however, this trade-off is not without costs as most crude cell-free systems are a black box - a system with largely undefined components. Numerous variables impact the metabolic potential of crude lysates and few have been empirically optimized (**Figure 1.1**. Notably, while the S30 crude CFPS system is among the most commonly used lysates, the impact of the numerous improvements to its preparation on the protein complement of the lysate are unknown [65, 66]. For example, inorganic phosphate is included its preparation media to reduce kinase activity; however, downregulation of kinases has not been observed experimentally [67]. Furthermore, there are an abundance of published preparation protocols, but standards for benchmarking are poor - frequently consisting solely of expression of fluorescent proteins and mRNA aptamers.

A promising alternative method for characterization of crude cell-free systems is shotgun proteomics; in this process, protein mixtures are enzymatically digested, and the resulting peptides are separated through reverse-phase liquid chromatography and sequenced with tandem mass spectromtery [68]. By measuring the abundance of the peptides which make up each protein, this un-targetted method provides comparative abundance measures of both the protein machinery of the cell-free system and of newly synthesized proteins. This

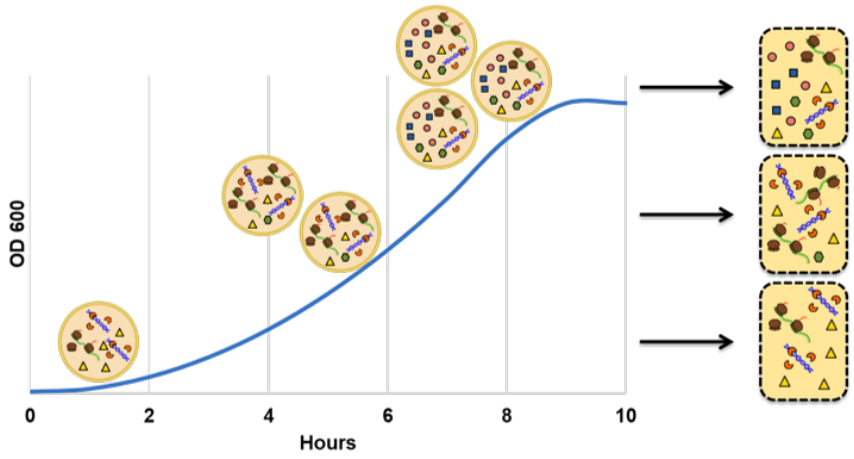


Figure 1.1: Cellular growth stage and harvest timing impact both cellular protein content and the proteome of the resulting cell lysate. The S30 lysate is harvested at early exponential phase to maximize ribosome concentration and *in vitro* protein synthesis capacity

method is made possible through advances in bioinformatic data analysis tools for predicting peptide MS/MS spectra from genomic data and matching them to observed spectra [69]. Additionally, coupling the high scan rate of quadrupole mass analyzers with high mass accuracy analyzers in quadrupole-Time-Of-Flight and quadrupole-orbitrap instruments has increased the number of peptides that can be identified in a single run. Nevertheless, it is important to consider that an observed increase or decrease in abundance of a protein is not equivocal with up- and down-regulation of that protein; many non-regulatory factors including degradation, digestion efficiency and sample composition can impact observed protein abundance.

Shotgun proteomics excels at identification of highly abundant proteins; however the presence of peptides of highly abundant proteins can interfere with identification of lower abundance proteins and limits the dynamic range of this method. These impacts are less prevalent in samples with lower complexity and can be offset by improved fractionation and chromatography methods [70]. This makes shotgun proteomics particularly robust for measuring protein abundances in bacteria, which have small genomes with few peptides shared among proteins. Its application is well-suited to crude bacterial cell-free systems as lysis has already been uniformly performed and most of the dynamic regulation present in a

cell is eliminated. With the application of omics tools such as shotgun proteomics, iterative and rational engineering of crude cell-free systems is possible.

1.0.1 Research Aims

The unique ability of cell-free systems to carry out complex biochemistries, including some inaccessible to living cells, while mimicking the cellular environment makes them an important tool to synthetic biology efforts to explore, understand and engineer metabolic processes. The high-throughput *in vitro* expression of enzymes made possible by CFPS enables cell-free systems to be used to discover new natural products, engineer new metabolisms and to produce biocommodities at scale [71]. However, improved analytical tools must be employed to systematically engineer lysates and monitor reaction progress. The goal of this dissertation is to detail the development and use of a systems metabolic engineering approach to the discovery, production, and characterization of small molecules in crude cell-free systems.

CFPS is the most robust and well-researched function in cell-free systems. The recent successes of cell-free protein production as a research tool and as a platform for synthesis of biologics at the point-of-care are reviewed in Chapter 2. The many advantages to cell-free bioproduction have driven interest in cell-free platforms for small molecule production; however, these systems require additional development and optimization to reach the productivity of CFPS. The remainder of this work focuses on expanding cell-free bioproduction to two small molecule classes – products derived from endogenous *E. coli* metabolism and peptidic NPs.

Chapter 3 and chapter 4 focus on the investigation and engineering of the biosynthetic potential of endogenous metabolism in crude *E. coli* cell lysates. These works develop shotgun proteomics as an enabling technology for engineering cell-free systems with an emphasis on small molecule production. In chapter 3 the impacts of growth condition variables on the enzymes involved in pyruvate synthesis are revealed through shotgun proteomics. Chapter 4 expands upon this approach by employing shotgun proteomics to guide modifications to a defined crude lysate preparation media for increased production of

aromatic amino acids *in vitro*. These works demonstrate the power of omics-scale analytical techniques for elucidating the metabolic black box of cell-free systems.

Chapter 5 focuses on the expansion of CFPS to the production of peptidic NPs in cell-free systems. For the investigation of lasso peptide BGCs in the Populus root microbiome, a methodology was developed pairing cell-free peptide synthesis with untargeted peptide sequencing by LC MS/MS to monitor the production and tailoring of ribosomally synthesized and post-translational modified natural products. This work highlights the need for more comprehensive measures of protein and polypeptide synthesis in cell-free systems, which would enable better benchmarking and consistency for CFPS.

Chapter 2

While-you-wait proteins? Producing biomolecules at the point of need

A version of this chapter was originally published by Benjamin Mohr, Scott Retterer and Mitchel Doktycz:

Mohr, B. P., Retterer, S. T., & Doktycz, M. J. (2016). While-you-wait proteins? Producing biomolecules at the point of need. Expert Review of Proteomics, 13(8), 707–709.

BPM wrote the manuscript with assistance for STR and MJD. We would like to thank Dr. Jennifer L Morrell-Falvey for editting this article.

2.1 Abstract

Protein drugs, or biologics, are a growing emphasis for the pharmaceutical industry owing to their unparalleled potency and specificity. Due to their temperature sensitivity, however, storage and distribution remain challenges. Furthermore, industrial scale production of biologics is not well matched to the needs of personalized medicine. To take full advantage of the potential of biologics, new methodologies for their synthesis and purification are required to produce them when and where they are needed most – at the point of care. Cell-free protein synthesis possesses the requisite portability and versatility to meet these needs. Recent innovations that are leading to distributed production techniques, enhanced product yields, and custom modifications will enable the rapid, on-demand production of personalized medicines.

2.2 Article

The functional diversity of proteins is astounding. Proteins can selectively organize, bind or catalyze a vast array of substrates. Living systems are enabled by these capabilities and various applications benefit from their engineered production. For example, protein-based pharmaceuticals, or biologics, have grown in popularity since the development of biosynthetic human insulin in 1978. Currently, biologics include some of the most highly prescribed drugs, including Humira for treating autoimmune diseases and Epogen/Procrit for addressing anemia resulting from renal failure and chemotherapy. Biologics are the fastest

growing sector of the pharmaceutical industry and their advantages over small molecule drugs ensure continued development [72]. These advantages include; i) unmatched potency, ii) potentially lower toxicity and iii) a broader diversity of targets for which they are selective [73].

Advances in genetically engineered organisms allow economical production of biologics. Often, microbial systems are employed due to their fast growth and genetic tractability. Currently, over a third of biologics are produced from genetically-modified microbes such as *S. cerevisiae* and *E. coli* [74]. Further, modern genetic construction and cloning techniques facilitate the discovery and development of strategies for the biosynthesis of novel biologics [73]. These genetically engineered microbial processes can be scaled-up to large batch cultures capable of manufacturing biologics with yields up to 7 g/L [75]. As optimization tools for genetic and metabolic engineering progress, the time and costs associated with developing new biologics are reduced.

Producing and distributing biologics comes with challenges. Unlike many small molecule drugs, biologics can be expensive to manufacture and temperature sensitive, causing a reduced shelf-life. While storage and distribution rarely present problems for advanced medical facilities, use in under-developed communities and remote environments is challenged by the need for uninterrupted cold storage. Maintaining protein activity until use may demand alternative approaches for preparing biologics. For example, if the material can be manufactured at its point of need, storage and distribution requirements can be minimized. Living cells are well suited for large-scale production. However, maintaining cell viability until use is not practical. In contrast, cell-free approaches for preparing biomolecules can overcome this limitation. This technique utilizes the protein machinery extracted from cells to carry out complex biosynthesis processes. For example, the transcription and translation machinery can be used to synthesize a chosen protein, or metabolic capabilities can be activated, or created, to prepare a metabolic compound. Potentially, cell extracts can be stored as universal reagents capable of transforming simpler reagents into a diverse range of biologics. Through selection and addition of a DNA sequence, and a mixture of shelf-stable reagents, potentially any desired protein can be created. While approaches for maintaining optimal activity of cell extracts require further investigation, transitioning

to a universal reagent set would maintain versatility and dramatically reduce needs for cold storage. Further, use of lyophilized cell extracts has been demonstrated, which could facilitate long-term reagent stability [76].

The usefulness of cell free methods matches other emerging needs in medicine. As DNA sequencing costs drop, a paradigm shift in medicine is occurring with the goal of tailoring treatments to a patient's genetic make-up. Individualized biologics, personalized to the unique needs of the patient, have been proposed for treating disease [77]. While large-batch industrial production of biologics has been sufficient for generally applicable pharmaceuticals, a flexible production strategy will be necessary to accommodate the rapid personalization of medicine. Feasibly, the quantity of a custom biologic needed for a single dose can be prepared by cell free protein synthesis (CFPS) in a portable format allowing remote, personal production of pharmaceuticals for specialized and emergency situations.

To advance CFPS for rapid, on-demand production of biologics, the product yield of the reaction must be improved. Reaction performance depends on the cell extract that is chosen. Cell extracts can be isolated in numerous ways, and from a variety of organisms. At the highest level, these extract preparation methods can be separated into two categories: purified and crude extracts. Purified extracts include one-pot reaction cascades carried out by combinations of purified enzymes and cellular machinery [44, 78]. These well-defined extracts reduce the number of potential side products and can simplify product clean-up steps; however, the upfront enzyme purification steps may render these extracts too costly for purposes other than research and development. In contrast, crude extracts are made from the lysis of whole cells. Native nucleic acids, membranes and metabolites are removed and the remaining cellular machinery is resuspended in a solution of salts, buffers and cofactors necessary to carry out bioprocesses. Extracts from both eukaryotic and prokaryotic cell lines have been used for carrying out *in vitro* bioprocesses [79, 80, 81]. Prokaryotic systems, namely *E. coli*, are the most advanced, and ongoing efforts in strain engineering and energy regeneration systems have improved protein synthesis yields [82, 83]. Typically, protein synthesis reactions are complete within a few hours. Bacterial extracts, however, cannot carry out all of the steps necessary to synthesize functional eukaryote-derived and post-translationally modified proteins without additional enzymatic processing. In contrast,

common eukaryotic wheat-germ, rabbit reticulocyte and insect cell free expression systems are capable of particular protein modifications; however, they lack the strain engineering tools and high protein yields offered by *E. coli* extracts. Continued efforts in genome wide engineering of cell extract source strains through CRISPR, MAGE and TRMR have increased mRNA and protein stability and enabled new functions [47]. The development of robust cell free protein expression in other cell extracts, or appropriate modifications of bacterial extracts, is needed for effective production of biologics.

The potential for producing therapeutically useful amounts of proteins in a portable format is becoming increasingly feasible. Typical requirements for therapeutic proteins dosages range between 0.01 to several milligrams per day [84]. For comparison, highly productive *E. coli* strains lead to cell free protein synthesis yields, in batch mode, of more than 2 mg/mL [54]. Further, it has been demonstrated that protein synthesis in microfluidic formats can improve yields by more than 6 fold compared to traditional batch reactors [84]. These efforts are critical to enabling portability, and when combined with high-yield extracts, can make remote production of useful amounts of biologics possible. However, similar to cell-based bioprocesses, proteins derived from cell-free systems need additional purification in order to be usable in a therapeutic context. Cell free synthesis approaches alleviate the need for cell lysis and related processing steps allowing the reaction products to be directly applied to chromatography columns. Microfluidic versions of common protein separation techniques have also been demonstrated [85]. Modular combinations of purification procedures, such as affinity binding, anion exchange and protein desalting, can potentially be configured to isolate a variety of desired products in a miniature format. Complementary tools for assessing protein quantity, activity, and purity at the microscale are needed. Development of these tools is promising as they share many of the same challenges as point-of-care diagnostics, which have enjoyed steady improvements in the analysis of proteins, nucleic acids and complex biological samples [86]. Nevertheless, the tools and technology for rapid, portable production of proteins are bringing on-demand biologics closer to reality.

With advances in on-demand biologics, other technical applications become feasible. Cell free approaches can also produce enzymatically active proteins. Functional metabolic pathways can be reconstructed from purified enzyme products, or possibly produced in

situ by CFPS, to prepare valuable commercial and therapeutic small molecules (e.g., antibiotics, vitamins, dyes, and pest control compounds) [62]. Beyond use in remote or resource challenged settings, cell-free biosynthesis is capable of facilitating various avenues of basic research; miniaturized and high-throughput proteomics for diagnostics, prototyping of biosynthetic pathways for eventual scale-up or analysis of genetic networks for synthetic biology can be considered [87]. Further, preparation of any of the large number of unknown proteins identified by genome sequencing efforts can facilitate understanding of their functional roles and can benefit from microfluidic-based approaches to determine their binding partners [88, 89].

The ease of implementing protein synthesis and purification at reduced scales, coupled with the constantly decreasing costs of gene synthesis, will expand the reach of cell free technologies and facilitate its use by small businesses, schools, and enterprising individuals, similar to the now-widespread adoption of additive manufacturing in the Maker community. Comparable to 3D printing, cell free techniques can accommodate niche and diverse markets due to low start-up costs and decreasing barriers to entry compared to traditional large-scale bioprocesses. As with most democratized technology, there are issues that will need to be addressed regarding regulations, intellectual property and biosafety; but, many of these issues are already under consideration by the scientific and regulatory communities with regards to synthetic biology.

Cell-free protein synthesis exploits the efficiency and versatility of biological processes without the limitations that come with managing living organisms. With the aid of micro- and nano-fluidic technologies, this technique can produce diverse biomolecules at therapeutically relevant scales. Furthermore, the absence of requirements for cell viability makes this technique ideal for the on-demand and customized synthesis of biologics at the point-of-care. Continued development and integration of the underlying technologies can enable rapid, convenient access to the functional diversity of proteins.

Chapter 3

Elucidating the potential of crude cell extracts for producing pyruvate from glucose

A version of this chapter was originally published by David C. Garcia, Benjamin P. Mohr, Jakob T. Dovgan, Gregory B. Hurst, Robert F. Standaert, and Mitchel J. Doktycz:
Garcia, D. C., Mohr, B. P., Dovgan, J. T., Hurst, G. B., Standaert, R. F., & Doktycz, M. J. (2018). Elucidating the potential of crude cell extracts for producing pyruvate from glucose. Synthetic Biology, 3(1).

BPM, DCG and MJD conceived the presented idea. DCG and JTD prepared cell-free systems and collected metabolic data. BPM and GBH performed proteomics experiments and analysis. BPM and DCG wrote the manuscript with assistance from RFS and MJD.

3.1 Abstract

Living systems possess a rich biochemistry that can be harnessed through metabolic engineering to produce valuable therapeutics, fuels and fine chemicals. In spite of the tools created for this purpose, many organisms tend to be recalcitrant to modification or difficult to optimize. Crude cellular extracts, made by lysis of cells, possess much of the same biochemical capability, but in an easier to manipulate context. Metabolic engineering in crude extracts, or cell-free metabolic engineering, can harness these capabilities to feed heterologous pathways for metabolite production and serve as a platform for pathway optimization. However, the inherent biochemical potential of a crude extract remains ill-defined, and consequently, the use of such extracts can result in inefficient processes and unintended side products. Here, we show that changes in cell growth conditions lead to changes in the enzymatic activity of crude cell extracts and result in different abilities to produce the central biochemical precursor pyruvate when fed glucose. Proteomic analyses coupled with metabolite measurements uncover the diverse biochemical capabilities of these different crude extract preparations and provide a framework for how analytical measurements can be used to inform and improve crude extract performance. Such informed developments can allow enrichment of crude extracts with pathways that promote or deplete particular metabolic processes and aid in the metabolic engineering of defined products.

3.2 Introduction

Synthetic biology aims to manipulate and exploit the existing biochemical functions of livings organisms for desired purposes. Unfortunately, efforts to engineer these systems to unlock their diverse metabolic potential require developing clever methodologies to overcome aspects of the machinery that the organism uses for survival. *In vitro* synthetic biology offers an alternative way to harness an organism's rich metabolism; it is driven by the prospect of easy to manipulate, static systems (Smith et al., 2014). Livings cells require membranes, energy and building blocks for growth, and dynamic regulation of their biochemical processes. By removing the requirement to sustain life, *in vitro* systems can sidestep many of the barriers to manipulation and present an ideal system for metabolic engineering.

In their most basic form, *in vitro* systems for metabolic engineering lack the genetic material and membranes inherent to a living system. Such *in vitro*, or cell-free metabolic engineering (CFME), approaches enable the use of techniques usually reserved for chemical engineering approaches such as continuous reaction monitoring, allowing for greater control over enzymes and metabolite concentrations [80, 90]. Coupled with systems biology tools for flux balance analysis and elementary mode analysis, *in vitro* systems present a potent platform for bioproduction [91]. The absence of a cell wall and membrane facilitates the exchange of substrate to, and product from, the system and simplifies reaction work up. Removal of the genome shuts down much of the cell's instructional programming and eliminates the need to cope with a continually growing and changing system. This enables biosynthesis pathways to be engineered *in vitro*, minimizing carbon and energy lost to growth. Additionally, this minimizes the management of feedback regulation and allows for the production of metabolites that would be toxic to intact cells [92].

Ideally, a cell-free metabolic engineering system would contain only the components necessary to carry out the desired biochemical process. One promising approach for complex chemical conversion uses a defined set of purified enzymes. This methodology has been successfully demonstrated for hydrogen production and protein synthesis among others[44, 93]. While recent efforts in co-purification of full reaction cascades have reduced costs, any process utilizing bulk purified proteins remains expensive [94, 47]. To date, the use

of purified components for CFME has resulted in long running systems capable of catalyzing reactions for several days, but with the drawback of slow catalysis rates. Novel work by Korman *et al.* on the production of limonene showcases the strengths and limitations of this approach [61]. Additionally, optimization of purified systems depends on ample information about the pathway and the involved proteins. These methods may fail to include accessory proteins which can improve pathway yield.

Crude cell extracts are finding increasing applications as alternatives to purified enzyme systems for metabolic engineering. Cell growth, followed by lysis and minimal fractionation can rapidly create robust biochemical systems for a fraction of the cost of purified enzymes. These systems contain the same enzymes and much of the same biochemistry as living systems and can serve as a proxy for the engineering of metabolite production by conventional, *in vivo* metabolic engineering. Recent work has demonstrated crude extracts as a platform for bioproduction as well, due to reduced costs of scale up and their compatibility with traditional chemical reactors [95, 96]. Further, early work in the optimization of bacterial cell free protein synthesis (CFPS) systems demonstrated the ability of crude cell extracts to energize translation *in vitro* through the consumption of glucose or other glycolytic intermediates [57, 97]. Glucose conversion is accomplished through the 10-step enzymatic process of glycolysis starting with the phosphorylation of glucose to glucose-6-phosphate and producing ATP through a series of substrate level phosphorylations. As shown in the aforementioned works, crude *E. coli* extracts can metabolize low-cost feedstocks like glucose to provide key intermediates and energy that can be drawn upon for myriad applications. The limits of the flexibility of crude extracts, granted by their inherently diverse biochemistry, remain uncertain. The proteome that enables these capabilities is only beginning to be explored and the extract preparation variables that influence this proteome require illumination [98]. Proteomic analyses coupled with metabolite measurements can be used to identify and characterized the biochemical pathways capable of being supported by crude extracts.

3.3 Materials and methods

3.3.1 Cell-free extract preparation

Cell extracts were prepared from *E. coli* BL21 Star (DE3) grown at 37 °C in one of three media: M9-fructose (11.1 mg l^{-1} $CaCl_2$, 0.120 g l^{-1} $MgSO_4$, 4.0 g l^{-1} fructose, 0.15 g l^{-1} KH_2PO_4, 3.39 g l^{-1} Na_2HPO_4, 0.25 g l^{-1} NaCl, 0.5 g l^{-1} NH_4Cl); lysogeny broth (LB: 10 g l^{-1} tryptone, 5 g l^{-1} yeast extract, 10 g l^{-1} NaCl); or 2xYPTG (16 g l^{-1} tryptone, 10 g l^{-1} yeast extract, 5 g l^{-1} NaCl,7 g l^{-1} KH_2PO_4, 3 g l^{-1} K_2HPO_4, 18 g l^{-1} glucose). The extracts prepared from these media are referred to as, DF, LB, and YT, respectively. Cell extracts were prepared by harvesting 50-mL cultures grown in baffled Erlenmeyer flasks to an OD600 of 1.0 for DF, 2.0 for LB, or 4.0 for mid-log phase YT (YT-M). The DF cells were additionally transferred to M9 salt solution containing no fructose for 24 hours before harvesting. A second-type of YT extract, YT-E, was prepared by growing cells to an OD600 of 2.8 and harvesting. Cells were harvested by centrifugation at 5000x g for 10 min in 50 mL volumes and washed twice with S30 buffer (14 mM magnesium acetate, 60 mM potassium glutamate, 1 mM dithiothreitol (DTT) and 10 mM Tris-acetate, pH 8.2) by resuspension and centrifugation. After the final centrifugation, pellets were weighed, flash-frozen in liquid nitrogen and stored at –80 °C. For extract preparation, cells were thawed and resuspended in 0.8 mL of S30 buffer per mg of cell wet weight before sonicating using 530 joules per mL of suspension at 50% tip amplitude with ice water cooling. After sonication, the cell-slurry was centrifuged twice for 10 minutes at 21,100 x g at 4 °C, aliquoted, flash-frozen and stored at –80 °C.

3.3.2 Cell-free reactions

Cell free glucose conversion reactions were carried out at 37 °C for 24 hours in 25 µL volumes with a final concentration 250 mM glucose, 18 mM magnesium glutamate, 15 mM ammonium glutamate, 195 mM potassium glutamate, 1 mM ATP, 150 mM Bis-Tris, 1 mM NAD^+, 10 mM dipotassium phosphate. Pyruvate consumption reactions were carried out using the same conditions and reagents, with the exception of glucose being replaced with 25 mM

pyruvate. Growth enriched extracts were added to a final protein concentration of 4 mg mL^{-1}. The reactions were quenched by the addition of an equal volume of 5% trichloroacetic acid. The supernatant after centrifugation at 11,000 x g for 15 minutes was used for analytical measurements.

3.3.3 Analytical measurements

High-performance liquid chromatography (HPLC) was used to measure pyruvate and glucose in the cell-free reactions. An Agilent 1260 series HPLC system equipped with a diode array UV-visible detector (Agilent, Santa Clara, CA) reading at 191 nm, with an Aminex HPX 87-H column (Bio-Rad, Hercules, CA) was used for the quantifications. Analytes were eluted with isocratic 5 mM sulfuric acid at a flow rate of 0.55 mL/ min at 35 °C for 25 min.

3.3.4 Proteomics

CFME extracts were denatured with 6 M guanidinium chloride for 1 h at 60 °C and allowed to cool to room temperature. Cysteines were reduced by incubation in 2 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP) for 20 min at room temperature and carboxamidomethylated by incubation in 10 mM iodoacetamide in the dark for 15 min. Samples were diluted with 5 volumes of digestion buffer (50 mM Tris-HCl, 10 mM $CaCl_2$, pH 7.6), and the proteins were digested by adding trypsin at a 1:50 weight ratio (based on Bradford assay) with overnight incubation at 37 °C. An additional identical amount of trypsin was then added, with an additional 4 h incubation at 37 °C. Trypsin was inactivated by addition of formic acid to a final concentration of 0.1%. Tryptic peptides were obtained by centrifugation through a 10 kDa molecular weight cutoff filter (Microcon YM-10, Millipore, Billerica MA) for 20 min at 14,000 x g. 50 µg of tryptic digests were loaded onto a strong cation exchange resin (SCX) (Luna, Phenomenex, Torrance, CA) and desalted. Digests were analyzed by two-dimensional liquid chromatography-tandem mass spectrometry [99]. Briefly, peptides were eluted from SCX resin with an eleven-step gradient of aqueous ammonium acetate (50 mM to 500 mM) onto reverse-phase C18 resin (Aqua, Phenomenex, Torrance, CA). Peptides were eluted from reverse-phase over two hours with

a gradient from 100% solvent A (5% CH_3CN, 0.1% formic acid in water) to 50:50 solvent A: solvent B (70% CH_3CN, 0.1% formic acid in water). Peptides eluted from the column were introduced into the linear ion trap mass spectrometer (LTQ-XL, ThermoScientific) by nanoelectrospray. Peptide identifications were obtained from MS/MS spectra using the program Myrimatch (version 2.1.138) and compared against the *Escherichia coli* BL21 (DE3) proteome (UP000002032), and protein identifications were assembled from peptide identifications using IDPicker, version 3.1.599 [100, 69]. KEGG Orthologies and Enzyme Commission numbers were assigned by BlastKOALA [101]. Full tables of detected proteins, tryptic peptides and KEGG orthology assignments are deposited in **Supplemental Table A-1** and **Supplemental Table A-2**. Descriptions of these tables and their legends are supplied in **Supporting Information**.

3.3.5 Statistical analysis

Three biological replicates were used for all HPLC measurements. Error bars in figures represent $\pm$ 1σ. Proteome analyses were likewise performed on three biological replicates. Significant changes in protein abundance for a given pair of treatments were identified using T-tests (2-tailed, unpaired, equal variances) on log10-transformed normalized spectral abundance factor (NSAF) value, with Benjamini-Hochberg correction for multiple hypothesis testing. Differential abundance was determined by ANOVA and direction of regulation by comparisons of prevalence value as previously described [98]. Gene set enrichment analysis (GSEA) was performed using the tools designed by Subramanian et al. [102]. In brief GSEA was performed using gene ontology (GO) terms and Uniprot pathway and superpathway annotations as pairwise comparisons of log10 transformed NSAF values between each set of extracts. Gene sets enriched with a false discovery rate $<$ 25% were retained.

3.4 Results

Given that the metabolic capabilities of a cell-free extract result from the active proteome, we hypothesized that changes to growth conditions prior to preparation of cell extracts would create shifts in the protein content and the resulting metabolic abilities of the crude

extracts. With the goal of investigating the protein elements of crude cell extracts that influence precursor supply, pyruvate biosynthesis in crude extracts was assessed. Pyruvate is both an important compound central to carbon metabolism, linking glycolysis and Krebs cycle and a launching point for numerous biotechnological targets [103]. Proteome profiles were obtained for the resulting crude extracts and validated by measuring the extracts' ability to produce pyruvate after the addition of glucose.

The effects of four growth conditions on the protein content and metabolic ability of *E. coli* crude extracts to produce pyruvate from glucose, were assessed. Three different growth media were used: standard rich broth (lysogeny broth, LB), M9 minimal medium with fructose (deprived fructose, DF), and extra-rich broth (2xYTPG, YT) with cells collected at mid-log phase. Cell growth in the 2xYPTG medium saturates at an OD600 of 8-10. Cells grown in this media were collected at both early, (OD600 2.8) and mid-log phase (OD600 4.0), and are referred to as YT-E and YT-M, respectively. The 2xYPTG condition, collected in early-log phase growth, is commonly used for CFPS [104]. Cells collected early in log phase growth have the greatest specific growth rate, a parameter that is suggested to influence CFPS capabilities and may affect the abundance of glycolytic enzymes [105, 106]. These growth conditions were chosen based on variables with the potential to enrich for glycolytic enzymes and for their frequent use for bacterial growth and related experiments that employ crude cell extracts. The DF condition employed M9 medium with fructose as the carbon source and a starvation regimen, which is reported to increase expression of glycolytic enzymes [107].

3.4.1 Effects of Growth Conditions on Proteomes of Extracts

Three biological replicates of the chosen cell extracts were digested with trypsin, and tryptic peptides were analyzed using multidimensional protein identification technology (MudPIT) as previously described [99]. Proteins were identified and assigned functions by matching peptides against an *E. coli* BL21 (DE3) genome that had previously been annotated by KEGG's BlastKOALA software [101]. After removal of low abundance proteins and data filtering, 1170 unique proteins were identified across all four conditions, with a core set of 796 proteins present in all four conditions. Despite overlap in media components between

several of the growth conditions, there were measurable differences in the proteomes of the four resulting cell-extracts. Excluding YT-M, each growth condition contained at least 10 unique proteins (**Figure 3.1**). As the same 2xYTPG media is used for both the YT-M and YT-E conditions the conditions overlap strongly, a larger number of proteins (40) are shared and are distinct from the LB and DF conditions. Further, the rich media conditions (LB, YT-E, YT-M) share a large number of proteins (132) that are distinct from the minimal media based DF condition.

With the goal of characterizing the proteins contributing to pyruvate production, the 10 enzymes comprising glycolysis were analyzed. In *E. coli*, four of the glycolytic enzymes exist as isoforms; 6-phosphofructokinase (Enzyme Commission (EC) number 2.7.1.11), glycerol-3-phosphate dehydrogenase (1.2.1.12), phosphoglycerate mutase (5.4.2.12), and pyruvate kinase (2.7.1.40), which may be regulated and function differently in a living cell. As the origin of these differences is beyond the scope of this work, each enzyme group is henceforth represented functionally as an EC number shared within isoforms. At least one isoform of each of the 10 glycolytic enzymes is present in all 4 conditions (**Figure 3.2**). While glycolysis

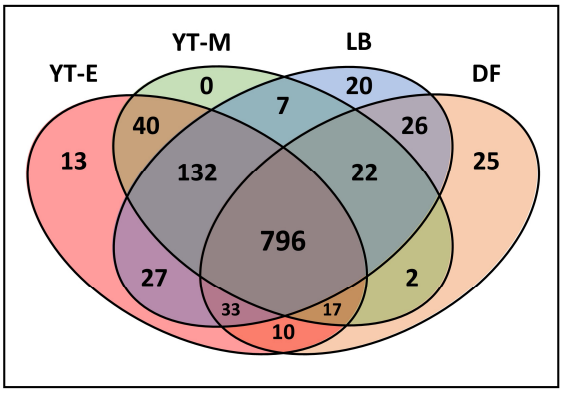


Figure 3.1: Venn diagram representing proteins found in each of the four growth conditions after filtering out low abundance proteins. There is a core set of 796 proteins present in all four conditions. The different conditions contain a small number of unique proteins, with the exception of YT-M, which may be a subset of the YT-E conditions. The cell extracts grown in rich media (YT-E, YT-M, LB) share a greater number of proteins (132) than any other subset of these conditions.

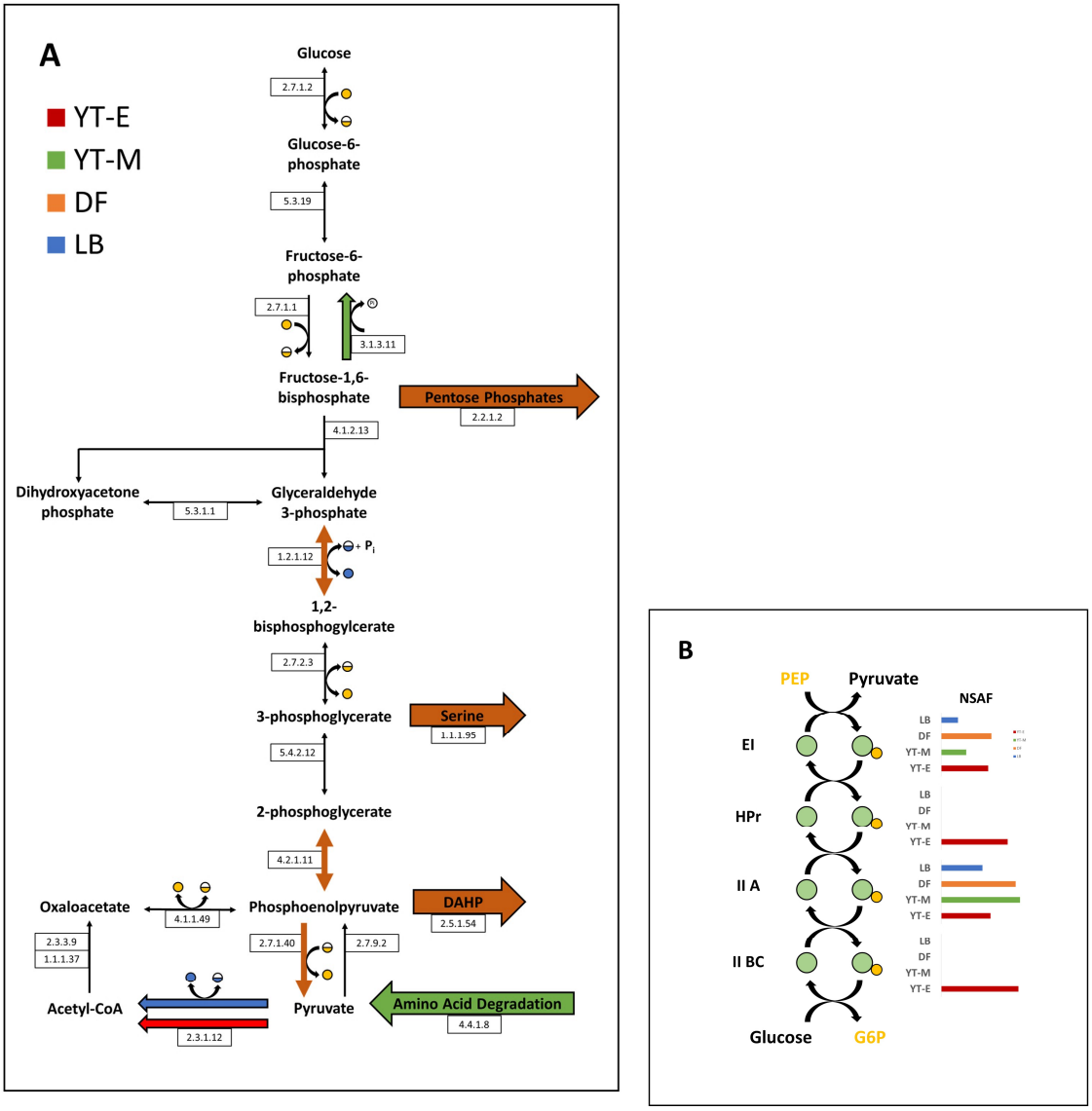


Figure 3.2: A.) Proteomic assessment of the potential *in vitro* metabolic fates of glycolytic intermediates. Enzymatic steps are represented by Enzyme Commission (EC) numbers. Each EC number and reaction arrow corresponds to at least one protein detected in all growth conditions. Upregulated steps are represented by red, green, orange and blue colored arrows for YT-E, YT-M, DF, and LB growth conditions, respectively. Only differentially abundant sinks of glycolytic intermediates are shown. Cofactors are depicted as colored circles, full yellow circle, ATP; half yellow circle, ADP; full blue circle, NADH; half blue circle, NAD+; "Pi" within the empty circle, inorganic phosphate. B.) Phosphoenolpyruvate transferase system reaction schematic. General PEP phosphotransferase enzyme, EI, is upregulated in both the YT-E and DF cell extracts. HPr phosphotransferase protein and glucose-specific IIA protein are only present in YT-E condition. Enrichment is represented by NSAFs.

is predominantly regulated through allosteric interactions, differences in abundance of the participating enzymes may also affect flux through the pathway [108]. As shown in (**Figure 3.2**), three enzymes within glycolysis are upregulated in the DF condition (1.2.1.12, 4.2.1.11, and 2.7.1.40).

In addition to energy, glycolysis provides the metabolic feedstocks for many important pathways, including the pentose phosphate pathway and the Krebs cycle, which in turn are the source of the carbon backbone for most primary and secondary metabolites in bacteria. The many carbon sinks leading out of glycolysis represent a significant draw away from pyruvate production. To account for the effect of these pathways, abundances of every enzyme in the BL21 (DE3) genome known to interact with an intermediate or product of glycolysis were analyzed for differences. These results are compiled in **Supplemental Table A-3**. **Figure 3.2** depicts the key differentially abundant enzymes in the cell extract proteomes that can act on glycolytic molecules. In particular, the pathway to the aromatic amino acid precursor shikimate was differentially represented via the pentose phosphate pathway enzyme transaldolase (2.2.1.2) and 3-deoxy-D-arabino-2-heptulosonic acid 7-phosphate (DAHP) synthetase (2.5.1.54).

The 10-enzyme glycolytic pathway from glucose to pyruvate begins with a phosphorylation that can be performed by hexokinase (2.7.1.2). This enzyme is present in all four extract conditions, but the phosphoenolpyruvate phosphotransferase system (PTS) provides an alternative entry point into glycolysis. PTS is a multi-protein phosphorylation cascade that *in vivo* results in a phosphorylated sugar moiety using PEP as an energy source. Previous work *in vitro* has demonstrated activity of the glucokinase and PEP phosphatase enzymes in crude extracts [109]). PTS specificity is dictated by the non-membrane bound IIA enzyme and membrane-bound IIBC enzyme, which often will only phosphorylate a single sugar, allowing for selective import of dedicated sugars. The EI PTS protein is upregulated in both the YT-E and DF extract (**Figure 3.2**). While the DF extract condition contains the fructose/mannitol-specific IIBC protein, the YT-E extract proteome uniquely contains the glucose-specific IIBC enzyme.

Analysis of individual enzymes helps to predict the flow of carbon through a cell extract. However analysis of individual abundances may fail to detect systematic differences between

cell extracts. Sets of phenotypically related genes can be co-regulated, but individually fail pairwise tests of significance. GSEA was performed using GO terms and Uniprot pathway designations annotated with the genome to account for these differences [102]. GSEA was performed as pairs of comparisons but some enrichments were shared amongst different comparisons and were consolidated (**Table 3.1**).

3.4.2 Pyruvate Production

We first investigated the glycolytic activity of the differentially prepared crude extracts by introducing them to a standard reaction mixture of the necessary co-factors NAD^+ and ATP as well as a set of buffering reagents and salts in order to confirm their ability to consume glucose and drive glycolysis towards pyruvate production. Over the course of a 24-hour incubation, aliquots of each reaction were halted using TCA, and quantified for glucose by HPLC analyses. Each extract broke down different amounts of glucose with YT-E consuming the largest amount at 147 mM at an average rate of 6.75 mM l^{-1} over a 24 hour period (**Figure 3.3**).

The concentration of pyruvate was simultaneously analyzed along with glucose consumption for each of the prepared extracts over a 24-hour time period. As would be expected, the final concentrations of pyruvate complemented the consumption rate of glucose with the YT-E extract producing the greatest amount of pyruvate at 21.29 mM. The DF extract produced the least amount of pyruvate at 0.73 mM and the values for LB and YT-M fell in between. However, the conversion of glucose to pyruvate was not quantitative. The differences in the extract's ability to both consume glucose and produce pyruvate, implies that CFME extracts can have a variety of metabolic capabilities based on their different protein content that results from changes in the cell growth conditions.

Due to the breadth of potential metabolic pathways present in the crude extract, we next sought to understand if the presence of the targeted metabolite, pyruvate, was maintained at a sufficient level to be an adequate feedstock for subsequent metabolic conversion. To test the activity of the extract's downstream pyruvate consumption pathways, we exogenously added pyruvate and co-factors to each extract and measured total pyruvate consumption after a 24-hour period (**Figure 3.4**). As suggested from the proteomic analyses, sink pathways for

Table 3.1: Summary of gene set enrichment analysis results based on biological process and molecular function GO terms and Uniprot Superpathway annotations. Table represents all enrichments found with a false discovery rate $< 25\%$ in pairwise comparisons. Enrichments found in more than one comparison have been combined.

Condition	Enrichment	Description	Direction of regulation	Comparison
YT-E	Carbohydrate Metabolism		Down	LB,DF
YT-M	Purine Metabolism		Up	YT-E
YT-M	Cofactor Biosynthesis		Up	YT-E
DF	Amino Acid Biosynthesis		Up	YT-E,LB
DF	Carbohydrate Degradation		Up	LB
DF	GO:0000049	tRNA Binding	Down	YT-M, YT-E
DF	GO:0003676	Nucleic acid binding	Down	YT-M, YT-E
DF	GO:0003723	RNA binding	Down	YT-M, YT-E
DF	GO:0003735	Ribosome	Down	YT-M, YT-E
DF	GO:0005506	Iron ion binding	Down	YT-M, YT-E
DF	GO:0006260	DNA Replication	Down	YT-M, YT-E, LB
DF	GO:0006281	DNA Repair	Down	YT-M
DF	GO:0006412	Translation	Down	YT-M, YT-E
DF	GO:0006457	Protein Folding	Down	YT-M, YT-E, LB
DF	GO:0007049	Cell Cycle	Down	YT-M, YT-E, LB
DF	GO:0019843	rRNA binding	Down	YT-M, YT-E
DF	GO:0051301	Cell Division	Down	YT-M, LB
DF	GO:0043565	Sequence-specific DNA binding	Down	YT-E
DF	GO:0003700	Transcription factor	Down	YT-E
DF	GO:0016301	Kinase activity	Up	YT-M
DF	GO:0016491	Oxidoreductase activity	Up	YT-M, YT-E, LB
DF	GO:0050660	Flavin adenine dinucleotide binding	Up	YT-M
DF	GO:0050661	NADP Binding	Up	YT-M
DF	GO:0051287	NAD binding	Up	YT-M
LB	GO:0030170	pyridoxal phosphate binding	Up	YT-E
LB	GO:0050660	Flavin adenine dinucleotide binding	Up	YT-E
LB	GO:0006099	Tricarboxylic acid cycle	Up	YT-E, YT-M

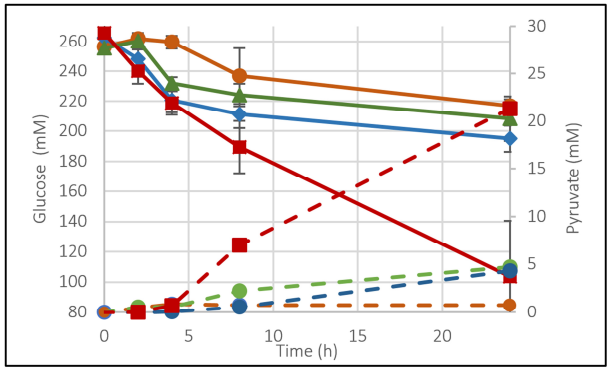


Figure 3.3: Extracts are shown here as orange, blue, green and red colored symbols and lines for DF, LB, YT-M and YT-E growth conditions, respectively. Data and standard deviation for the time course reactions were acquired using n=3 biological replicates. Glucose and pyruvate were simultaneously measured at various time points over a 24-hour period. Solid lines indicate glucose time courses and dashed lines indicate those for pyruvate.

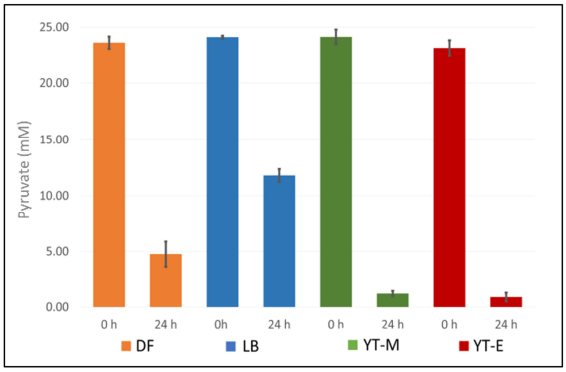


Figure 3.4: Extracts are shown here as orange, blue, green and red colored symbols and lines for DF, LB, YT-M and YT-E growth conditions, respectively. Data and standard deviation for the pyruvate substrate reactions were acquired using n=3 biological replicates. The extracts' ability to consume exogeneous pyruvate was measured. Time 0 h indicates initial pyruvate concentration. Final concentration is measured after 24 h.

glycolytic intermediates are well represented in the crude extracts. Calculation of the glucose to pooled pyruvate conversion rates and the pyruvate consumption rates indicate significant differences in the fraction of glucose that passes through pyruvate (**Table 3.2**). Each of the extracts was capable of consuming a large portion of the pyruvate provided regardless of the extract preparation condition. The YT-M extract was capable of consuming pyruvate at nearly the same rate as the YT-E extract, 0.96mM l^{-1} and 0.93mM l^{-1} respectively, despite the YT-E extract producing a larger pool of pyruvate after 24-hours.

The DF condition was capable of breaking down pyruvate nearly 26 times faster than it could be produced indicating that the extract is relatively enriched for pathways downstream of pyruvate, in addition to those that deplete glycolytic intermediates. In the YT-M and the YT-E conditions, the consumption of pyruvate was greater than the DF condition, however, the potential consumption from downstream pathways was not enough to deplete the pyruvate reservoirs created by the extract. The YT-E extract, in particular, was able to maintain a reservoir of pyruvate that nearly matched the total pyruvate that it could consume within a 24-hour time frame.

3.5 Discussion

The metabolic potential of crude extract preparations and their use for metabolite production can be assessed through exploration of their proteomic and metabolic profiles. Despite progress in the use of crude extracts for protein expression, the actual content of a crude extract and its metabolic potential is poorly understood. We sought to address this deficiency by exploring the protein profiles of different cell extract preparations and assess their ability to produce pyruvate from glucose in CFME. As a central player in a variety of cellular processes, pyruvate is not only a key indicator of a crude extract's glycolytic potential, but also an important proxy for the extract's ability to produce small molecules of commercial interest [103]. To explore the optimization of precursor production in cell-free extracts, we modulated cell growth conditions in order to create global changes in an extract's protein content. Given the static nature of the protein content in crude extracts, understanding the

Table 3.2: The percent glucose consumed and the percent glucose converted to pooled pyruvate were determined after 24 hours of feeding the reactions 250 mM glucose and measuring the remaining glucose concentration and the amount of pyruvate produced, respectively. The percent of glucose converted to pyruvate and downstream metabolism was determined after measuring the consumption rate of 25 mM pyruvate after 24 hours in the absence of glucose. Conversion amounts were determined using n=3 biological replicates.

Extract	% Glu Consumed	% of Consumed Glucose Converted to Pooled Pyruvate[a]	% of Consumed Glucose Converted to Pyruvate and Downstream Metabolism[b]
DF	15.55%	0.92%	24.55%
LB	25.74%	3.50%	12.60%
YT-M	18.52%	4.58%	28.83%
YT-E	61.00%	6.57%	13.45%

[a] Glucose conversion was calculated by measuring pooled pyruvate after 24 h and converting to glucose.

[b] The expected glucose used to make the pyruvate consumed by downstream metabolism was combined with the glucose consumed in order to produce the pooled pyruvate to account for the breakdown of pyruvate due to downstream metabolism and show the extract's ability to synthesize glucose from pyruvate without the sink of downstream metabolism.

proteomic and metabolic potential of these systems can provide an effective platform onto which heterologous pathways could be engineered with predictable effects and high yields.

The ability of each crude extract to potentially break down glucose is evident from proteomic analyses. All growth conditions resulted in extracts with the presence of at least a minimal set of the ten enzymes required for converting glucose to pyruvate. Confirmation of glycolytic activity was supported by metabolite analyses. The different extracts all converted glucose to pyruvate in the presence of the appropriate cofactors. Further, a nearly thirty-fold difference in the amount of pooled pyruvate, after 24-hours, is observed when comparing the different crude extracts; ranging from 21.29 mM in YT-E and 0.73 mM in DF. These comparisons showcase the importance of growth conditions when preparing an extract. Proteomic analyses aid in interpreting the effect of these changes. For example, the differences in methodology to create the YT-E and the YT-M extracts are seemingly minor, with the YT-E extract being harvested at an earlier point in the growth phase when compared to YT-M. However, the effects of this change significantly impact the metabolic

pathways present in the extract. While both the YT-M and the YT-E extracts are able to break down pyruvate at a similar rate, the production of pooled pyruvate in the YT-M extract is only 1/5th that of the YT-E extracts. This difference can be attributed to the prevalence of the glucose-specific PTS system, which may serve as both an entry and exit for glycolysis and account for both the increased glucose consumption and the increased production of pyruvate in the YT-E extract. The absence of the HPr protein in the DF condition removes this alternative glucose consumption pathway despite a relative abundance of the glucose-specific proteins. Glucose consumption in cell-free systems can be a robust process and has been shown in previous studies [57]. As confirmed here, 15-60% of added glucose can be metabolized by crude extract preparations. Biosynthesis and degradation pathways drawing from central metabolism, such as those for nucleotides, lipids and amino acids, can affect the flow of carbon to pyruvate in the DF cell extract. Proteome analyses indicated that the DF cell extract, which was grown on a minimal medium, is enriched in amino acid and nucleotide biosynthesis pathways that are not prevalent in the other extract preparations (**Figure 3.2, Supplemental Table A-3**). These pathways rely upon intermediates from glycolysis for their carbon backbones and decrease overall flux towards pyruvate. The upregulation of the glycolytic enzymes combined with the presence of shunting pathways show a clear path by which the DF extract could produce pyruvate, but not accumulate pyruvate, in the same fashion as the YT-E extract where the overwhelming amount of of the produced pyruvate was shunted downstream.

Each tested growth condition resulted in an extract capable of breaking down pyruvate, which depletes the pool of this metabolite. The different crude extracts were capable of consuming up to 90% of added pyruvate. After accounting for this consumption, the overall production rates of pyruvate for the DF, LB, YT-M, and YT-E extracts are 0.82mM l^{-1}, 0.71mM l^{-1}, 1.14mM l^{-1}, and 1.82mM l^{-1}, respectively. None of the consumed pyruvate appears to be converted back to glucose. As previously noted, the production of pyruvate from PEP is effectively irreversible [110]. The pyruvate is likely funneled into downstream metabolic pathways, and analyses of proteomic information provided insights. GSEA reveals that carbohydrate metabolism, specifically the Krebs cycle was up-regulated in the LB extract. Conversely, the YT-E and YT-M extracts were relatively depleted in the Krebs

cycle as is common for cells in exponential growth [111]. Interestingly, a component of the pyruvate dehydrogenase complex (2.3.1.12) was upregulated in both the YT-E and YT-M extracts, potentially indicating the channeling of pyruvate to acetyl-CoA. Rapid growth, which might be expected under conditions with abundant resources, results in a need for biomass components, the biosynthesis of which both starts from intermediates in glycolysis, and heavily deplete the central precursors therein [112]. Extracts derived from rapidly reproducing cells can result in active biomass accumulation pathways and can result in a significant drain on both feed metabolites and cofactors in metabolic engineering endeavors. These pathways likely lead to the increased consumption of pyruvate observed in the YT-E and YT-M extracts. While sink pathways draw carbon away from central metabolism, their effect may be mitigated by the prevalence of upstream pathways providing a balancing effect. The LB extract is upregulated in both anaplerotic pathways and gluconeogenesis (4.1.3.1, 4.1.1.49, 3.1.3.11) and consumed comparatively less pyruvate than the other extracts.

The use of proteomic and biochemical analyses to describe the metabolic capabilities of a crude extract provides a useful framework for realizing an extract's potential applications and optimization. Changes, either genetic or to growth conditions, can be made to further tailor a crude extract for desired function. Here, it is evident that growth on a minimal medium results in the expression of many sink pathways for glycolytic intermediates. Moreover, it appears specific growth rate, which has been previously examined as a key variable in CFPS extract preparation, plays a role in reducing sinks due to the Krebs cycle, but at the price of directing flux towards undetermined biomass accumulation pathways [105]. Proteomic analysis is a robust technique for determining candidates for genetic manipulation and can guide *in vivo* protein overexpression or knockdowns in source strains that will affect the flux of small molecules after extract preparation. Alternative strategies, such as targeted protein degradation and pull downs, have been described for the removal of deleterious proteins from crude extracts after cell lysis to avoid negatively impacting cell growth and survival [113]. Crude extracts made from high-yielding *in vivo* pyruvate production strains represent another opportunity to use *in vitro* synthetic biology to enable metabolic engineering [114, 115]. Some of the highest producing strains are limited in their genetic tractability, but

omics data can provide strong candidates for modification and minimize the amount of genetic engineering needed.

3.6 Conclusion

Critically analyzing the central precursors of cell-free systems as well as how the conditions in which these extracts are grown can impact the metabolic potential of a cell free system due to changes in the underlying proteomic content. Notably, we demonstrate that simple changes in cell-free extract preparation can result in profound differences in metabolite pooling. Further, these changes in extract preparation have the potential to deplete important precursors that could be used for synthesis of a final product. These different metabolic characteristics can be interpreted through the combined use proteomics and metabolomics techniques. These analytical measurements further our understanding of the composition of cell-free extracts and provide a rich dataset from which to engineer improved solutions for metabolite production. These tools can guide genetic manipulations and strain optimization conditions for maximizing the production of pyruvate, as well as other important biosynthetic precursors. Feasibly, effective development of crude extracts can lead to a general platform suitable for testing biochemical pathways and for production of useful metabolites.

3.7 Acknowledgements

Research supported by the U. S. Department of Energy (DOE) Office of Biological and Environmental Research, Genomic Science Program. This research was supported in part by an appointment to the Higher Education Research Experiences Program at Oak Ridge National Laboratory for JTD. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. DOE under Contract no. DEAC05-00OR22725.

Chapter 4

A systems-cell-free metabolic engineering approach to optimizing growth media for *in vitro* phenol biosynthesis in crude *Escherichia coli* cell-free systems

A version of this chapter by Benjamin P. Mohr, Richard J. Giannone, Robert L. Hettich, and Mitchel J. Doktycz is submitted to Metabolic Engineering Communications

BPM designed and carried out the study. RJG and RLH assisted with proteomics measurements and analysis. BPM and MJD wrote the manuscript with the assistance of RJG and RLH. We would like to thank Dr. Brian Sanders for editting this article.

4.1 Abstract

Progress in cell-free protein synthesis (CFPS) has spurred resurgent interest in engineering complex biological metabolism outside of the cell. Unlike purified enzyme systems, crude cell-free systems can be prepared for a fraction of the cost and contain endogenous cellular pathways that can be activated for biosynthesis. However, efforts to use crude cell-free systems as a bioproduction platform have been hampered by the under-described complexity of the metabolic networks inherent to a crude lysate. A systems cell-free metabolic engineering (S-CFME) approach incorporates tools such as shotgun proteomics to comprehensively describe the biochemical networks in crude lysates for subsequence refinement and augmentation of complex lysate phenotypes for improved yield and titer. Here a S-CFME approach has been developed and implemented for *in vitro* phenol biosynthesis using crude *E. coli* lysates. Crude cell lysate metabolic activity was focused towards the limiting precursor tyrosine by optimization of growth conditions guided by shotgun proteomics. The result is the activation of a 25-step enzymatic reaction cascade involving at least three endogenous *E. coli* metabolic pathways including glycolysis, the pentose phosphate pathway and the shikimate pathway. Additional modification of this system, through CFPS of feedback intolerant AroG and phenol-tyrosine lyase from *Pasteurella multicloida*, improves yield. This effort demonstrates the ability to activate a long, highly complex pathway *in vitro* and provides a framework for harnessing the metabolic potential of diverse organisms for cell-free metabolic engineering. The more than six-fold increase in phenol yield with limited genetic manipulation demonstrates the benefits of optimizing

growth media for crude cell-free extract production and illustrates the advantages of a systems approach to cell-free metabolic engineering.

4.2 Introduction

Crude cellular lysates were essential reagents in the initial exploration of cellular biochemistry. The identity of each amino acid, the mechanisms of the Krebs cycle, and transcription and translation were discovered through work performed with crude cell lysates [37, 38, 39]. These discoveries were possible because crude lysates retain many of the biochemical functions of living cells, outside of the cell. Isolation of these functions through various lysate preparation methodologies allows for their study and engineering in a cell-like context. Recently, engineered crude cell lysates have gained new popularity, reborn in the age of synthetic biology as cell-free systems[116].

Cell-free systems are complex mixtures of enzymes and small molecules, derived from the controlled lysis of living cells, that are capable of biochemical reaction cascades. Cell-free systems have several advantages compared to live cellular systems. Upon lysis, many of the cellular properties associated with life which impede engineering efforts are eliminated. Without concerns for cell viability, cytotoxic products can be prepared and resources can be directed away from growth-associated processes to improve yield. The absence of an outer membrane removes challenges in product recovery, system handling, and supply of cofactors and reagents. The destruction of endogenous genetic material through sonication, bead beating, or enzymatic degradation removes transcriptional and translational regulation within the system. Additionally, lysate fractionation strategies can be used to isolate, combine or remove specific biochemical functions to create novel pathways [117]. Purification of entire metabolic pathways from crude lysates has been demonstrated, but scalability remains a concern for cell-free systems containing purified components [61, 64]. In contrast, crude cell-free systems, commonly referred to as cell extracts or lysates, scale similarly to living cells but can achieve higher turnover rates for complex biochemical functions including protein synthesis [71, 54]. While allosteric interactions are still present, crude cell-free systems are a predominantly static biochemical snapshot of the moment the cell

was lysed. Removal of the outer membrane, endogenous genetic material and many of the related regulatory systems streamlines engineering efforts.

Crude bacterial cell-free systems have been engineered to carry out *in vitro* transcription and translation from an exogeneous genetic template. Early systems consumed small molecules containing high energy phosphates such as phosphoenolpyruvate and creatine phosphate to regenerate ATP and fuel cell-free protein synthesis (CFPS) [118]. More recently, crude *Escherichia coli* CFPS has been improved by the recycling of inorganic phosphate and the regeneration of ATP by endogenous lysate metabolism to reduce cost and improve yield [119]. When provided with sugars or glycolytic intermediates, lysates produce ATP from central carbon metabolism *in vitro* [57]. The small molecules produced at the terminal end of *in vitro* glycolysis serve as launching points for synthesis of secondary metabolites and form the basis of most cell-free metabolic engineering (CFME) efforts in crude lysates to date [58, 95, 103].

Despite the reliance upon native lysate metabolism for substrate processing, energy generation, cofactor cycling, and protein synthesis, endogenous activities are rarely engineered and optimized in CFME. Though crude cell-free systems are less complex than living cells, they can still contain thousands of different proteins capable of interacting and producing complex, sometimes off-target effects [98, 120]. Further, lysate preparation can result in co-isolation of deleterious enzymes and metabolites, which can complicate downstream engineering efforts.

To address these issues, it is pertinent to interrogate the system holistically rather than myopically focusing on discrete pathways where underlying, systemic issues may complicate desired CFME outcomes. As follows, systems metabolic engineering utilizing "-omics" measurements and computational tools to interrogate and engineer the complexity of living cells can be important to the success of CFME. By incorporating tools designed to characterize and model complex biochemical interactions in vivo, the preparation of crude cell-free systems can be guided to increase product yield by harnessing the intrinsic metabolic activity of the lysate. This systems cell-free metabolic engineering (S-CFME) approach will facilitate creation of more efficient cell-free bioproduction platforms. Building upon recent efforts using shotgun proteomics to describe the malleability of *E. coli* cell-free system endogenous metabolism [121], we present a S-CFME approach to direct metabolic

activity towards defined intermediates to create and enhance the yield of a non-endogenous metabolite, phenol. By optimizing both strain culture conditions, informed by proteomics data, as well as *in vitro* expression of enzyme variants, cell-free phenol production is greatly improved (**Figure 4.1**). As increasingly complex pathways are engineered in crude cell-free systems, this generally applicable systems approach will become essential to unlocking the untapped potential of endogenous metabolic activity in crude cell-free systems.

4.3 Materials and methods

4.3.1 Strains, plasmid and reagents

All plasmids used in this study were synthesized by Genscript (Piscataway, New Jersey, USA). Reagents were purchased from Sigma Aldrich (St. Louis, Missouri, USA) unless otherwise noted. *E. coli* BL21 Star (DE3) was purchased from New England Biosciences (Ipswich, Massachusetts, USA).

4.3.2 Cell extract preparation

Cell extracts were prepared from *E. coli* BL21 Star (DE3) grown at 37 °C in variants of YPTG (16 g L^{-1} tryptone, 10 g L^{-1} yeast extract, 5 g L^{-1} NaCl, 7 g L^{-1} KH_2PO_4, 3 g L^{-1} K_2HPO_4, 18 g L^{-1} glucose) and EzRich media. EzRich defined rich media kit was supplied by Teknova and purchased from VWR (Radnor, Pennsylvania, USA). 5x EZ Supplement

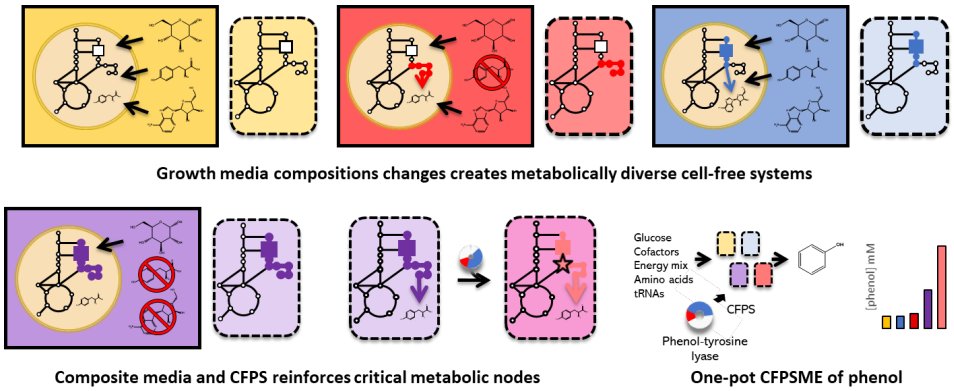


Figure 4.1: Graphical Abstract

without tyrosine, tryptophan and phenylalanine was purchased from BioWorld (Atlanta, Georgia, USA). Variants of EZRich media and their designations are summarized in **Table 1**. In brief, for preparation of cell extracts, 50 mL cultures were grown in baffled Erlenmeyer flasks to an OD600 of ≈0.8 and induced to 1 mM isopropyl-β-d-thiogalactopyranoside. Cells were harvested at early exponential phase, about 2.5 hours after induction, corresponding to OD600 of 2.8 and 4.0 for media YTPG and EzRich variants, respectively. Cells were harvested by centrifugation at 5000x g for 10 min and washed with S30 buffer (2x, 25 mL, 14 mM magnesium acetate, 60 mM potassium glutamate, 1 mM dithiothreitol and 10 mM Tris-acetate, pH 8.2). Cell pellets were weighed, flash-frozen in liquid nitrogen, and stored at –80 °C. For extract preparation, cells were thawed and resuspended in 0.8 mL of S30 buffer per mg of cell wet weight before lysis with a Branson Ultrasonics Sonifier SFX250 equipped with a microprobe. Cells were lysed with 530 joules per mL of suspension at 50% tip amplitude in a 0 °C water bath. Post-lysis the cell-slurry was centrifuged twice for 10 minutes at 21,100 x g at 4 °C, the supernatant was aliquoted, flash-frozen and stored at –80 °C.

4.3.3 Cell-free reactions

Cell-free reactions for protein synthesis or phenol production were carried out at 30 °C in 25 µL volumes with the following components: 40 mM $^{13}C_6$ glucose, 1.2 mM ATP; 0.85 mM each of GTP, UTP and CTP; 34 µg mL^{-1} folinic acid; 67.7 mM creatine phosphate, 3 µg mL^{-1} creatine kinase, 0.4 mM pyridoxal 5'-phosphate, 2 mM each of the 20 translatable amino acids, 0.33 mM nicotinamide adenine dinucleotide (NAD), 0.26 mM coenzyme A (CoA), 33 mM PEP, 18 mM magnesium glutamate, 15 mM ammonium glutamate, 195 mM potassium glutamate, 1.5 mM spermidine, 1 mM putrescine, 57 mM Bis-Tris pH 7, 100 ng µL^{-1} plasmid DNA and 15 µL cell extract adjusted to 10 mg mL^{-1} by Bradford assay. Cell-free reactions were overlaid with 100 µL of tributyrin to prevent evaporation. Cell-free protein synthesis of sfGFP was performed in a 96 well plate in a Perkin Elmer EnSpire 2300 for 8 hours, with fluorescent measurements (excitation 488 nm, emission 509 nm) every 20 minutes. Phenol production reactions were run for 48 hours in 1.5 mL microcentrifuge tubes. After 48 hours, phenol production reactions were vortexed and centrifuged for 10 minutes at 21,100 x g at

4 °C. 50 µL of tributyrin overlay was removed, added to 0.5 mL of dicholoromethane and subjected to analysis by GCMS.

4.3.4 Phenol quantitation

In vitro synthesized phenol was quantified on an Agilent 7890A gas chromatograph equipped with a 5975C mass spectrometer. Tributyrin overlays diluted with dicholormethane were injected onto a HP-5MS column at 40 °C. Initial oven temperature was held for 3 minutes, ramped to 120 °C at 22 °C/ min and held for 1 additional minute. The oven was then heated to 325 °C and maintained for 3 minutes. $^{13}C_6$, $^{13}C_4$, and non-labeled phenol were monitored at m/z 100.1, 98.1, and 96.1 respectively. Phenol was quantified by peak integration and comparison to a standard curve in Thermo Xcalibur. Two technical replicates were measured for every sample.

4.3.5 Proteomics

To prepare cell lysates for proteomic analysis, each lysate was precipitated with trichoroacetic acid 20% w/v. Precipitated proteins were pelleted by centrifugation at 21000 x g for 10 min at 4 °C and washed with acetone (2x, 1 mL, -20 °C). Acetone was decanted, and pellets were air dried. Protein pellets were resolubilized in 8 M urea, 10 mM DTT, 100 mM Tris–HCl, pH 8.0, alkylated with 20 mM iodoacetamide, and digested with trypsin as previously described [121]. Digestion was halted by addition of 200 mM NaCl and 0.1% formic acid. Tryptic peptides were then passed through a 10 kDa molecular weight cutoff size exclusion filter (Microcon YM-10, Millipore, Billerica MA). Peptides were quantified by BCA assay and 50 µg were loaded via a pressure cell onto the reverse phase (Aqua C18, 5 micron, Phenomenex, Torrance, CA) of a biphasic strong cation exchange (Luna SCX, 5 micron, Phenomenex, Torrance, CA) and C18 column, desalted, and eluted onto the strong cation phase. Digests were analyzed by two-dimensional liquid chromatography in line with nanospray tandem MS using an LTQ-Orbitrap XL mass spectrometer (Thermo Scientific). Peptides were eluted with 4 salt cuts of ammonium acetate (10, 25, 50, and 100 mM), each followed by a 100-minute organic gradient to separate peptides on C18 [99]. The mass

spectrometer was operated in data dependent acquisition (one full scan at 30K resolution followed by 5 MS/MS scans in the LTQ, 3 m/z isolation window, dynamic exclusion: window = 1 m/z, duration = 60 s). Using MyriMatch (version 2.1.138), protein identifications were obtained by matching observed MS/MS spectra to predicted peptide fragmentation spectra from a database of 4156 *Escherichia coli* BL21 (DE3) protein sequences (proteome UP000002032) appended with common contaminants and decoy sequences [69]. Protein identifications were assembled from filtered peptide spectrum matches (PSM) using IDPicker, version 3.1.599 (2 peptide minimum per protein; protein false-discovery rate <0.8%; peptide false-discovery rate < 0.3%, spectrum false-discovery rate < 0.1%) [100]. PSMs were assigned matched-ion intensities by integration of observed peptide fragment peaks; unique PSMs matching individual proteins were summed to derive protein intensity. Protein intensities were log2 transformed, normalized (LOESS) and standardized across all samples by median absolute deviation using InfernoRDN [122]. To simulate the lower bounds of detection, intensities of protein not observed in a condition, but described globally were imputed with a normal distribution (downshift = 1.8, width = 3) using Persues [123]. Enzyme Commission numbers were assigned by BlastKOALA [101]. Full tables of detected proteins, tryptic peptides are deposited in **Supplemental Table B-1** and **Supplemental Table B-2**. Descriptions of these tables and their legends are supplied in **Supporting Information**.

4.3.6 Statistical analysis

Three biological replicates were used for all proteomics measurements. Differences in protein abundance, based upon average log2 protein intensity, were determined by Student's t-test (2-tailed, unpaired, equal variance). Statistics were performed, and plots were generated in R (version 3.5.3) with packages Tidyverse and ggpubr [124].

4.4 Results and Discussion

4.4.1 Enabling phenol production in *E. coli* cell-free systems

Aromatic compounds are valuable chemicals with uses as industrial solvents, fuels, and substrates for chemical synthesis. Largely derived from petroleum, manufacturing of aromatic compounds by microbial fermentation of a low-cost sugar substrate would present an environmentally friendly alternative. As aromatic rings are present in nucleotide bases and in three of the proteinogenic amino acids, many organisms have biosynthetic pathways to produce aromatic compounds. The building blocks for the aromatic amino acids phenylalanine, tryptophan, and tyrosine result from the shikimate pathway. Additionally, the shikimate pathway is the metabolic launching point for biosynthesis of phenylpropanoids, a diverse class of secondary metabolites synthesized from iterative additions of malonyl- and coumaroyl-CoAs, that include medicinally valuable compounds such as flavonoids and stilbenoids. Others have succeeded in developing *in vitro* biosynthetic pathways for highly conjugated compounds including acyl-CoAs, but production of aromatic compounds by the shikimate pathway *in vitro* has not been explored [125].

Phenol is one of the simplest aromatic compounds, consisting of a six-carbon aromatic ring appended with a single hydroxyl group. Phenol-tyrosine lyases (PTL, 4.1.99.2) from various enterobacteria have been found to catalyze the synthesis of phenol from the amino acid tyrosine, and recently, Kim *et al.* successfully produced microbially derived phenol by heterologous expression of PTL. Improving substrate availability by engineering tyrosine biosynthesis increased phenol yield, but cytotoxicity limited productivity [126]. The reduced impact of highly cytotoxic products on cell-free bioproduction platforms provides an attractive alternative for phenol biosynthesis [127].

While many microorganisms, including *E. coli*, can make their own tyrosine, high-yield tyrosine biosynthesis is a complex phenotype. Tyrosine biosynthesis requires not only the four and three carbon building blocks, erythrose 4-phosphate (E4P) and phosphoenolpyruvate (PEP), which are condensed to form 3-deoxy-D-arabino-heptuloseonate 7-phosphate (DAHP), but an additional PEP, ATP, and NADPH are also required. NADPH can be

regenerated through the prephenate dehydrogenase activity of TyrA (5.4.99.5/1.3.1.12), however PEP and ATP must be generated outside of the shikimate pathway (**Figure 4.2**).

In this work, the one-pot *in vitro* biosynthesis of phenol was achieved by coupling endogenous production of tyrosine from glucose with CFPS of PTL from *Pasteurella multocida*. Fully-labeled $^{13}C_6$ glucose was used as the carbon source to distinguish between phenol synthesized from amino acids added as a substrate for CFPS and the desired full pathway. CFPS and phenol production both require exogenous ATP; as oxidative phosphorylation is not expected to be active in these systems, creatine phosphate and creatine kinase were added to energize these reactions [128]. Simultaneous addition of PTL template DNA, labeled glucose, and creatine kinase initiated *in vitro* phenol production, which proceeded over the course of 48 hours and was quantified by GC/MS.

4.4.2 Characterization of crude cell-free systems prepared from defined media

Previously we demonstrated that composition of growth media for cell extract preparation can alter lysate metabolic activity [121]. Small changes in available nutrients and growth conditions result in large compensatory shifts in protein abundance which can be observed with shotgun proteomics. To provide fine control over media conditions, a cell-free system based upon growth on defined media was developed. Using this system, variables potentially impacting tyrosine production including carbon source and presence of aromatic compounds in the media were investigated. In particular, the effect of aromatic amino acids and nucleotide bases in the media was explored. Impacts of each change to the growth media were evaluated by shotgun proteomics and used to inform subsequent modifications. All media compositions are detailed in **Table 4.1**.

E. coli cell-free systems for protein production are generally grown using the rich, complex media YTPG, which consists of five components: yeast extract, tryptone, NaCl, potassium phosphate and glucose. Yeast extract and tryptone contain many different complex biomolecules with significant batch to batch variations; this presents limited opportunity for modification and optimization. The rich, defined media described by Neidhardt *et*

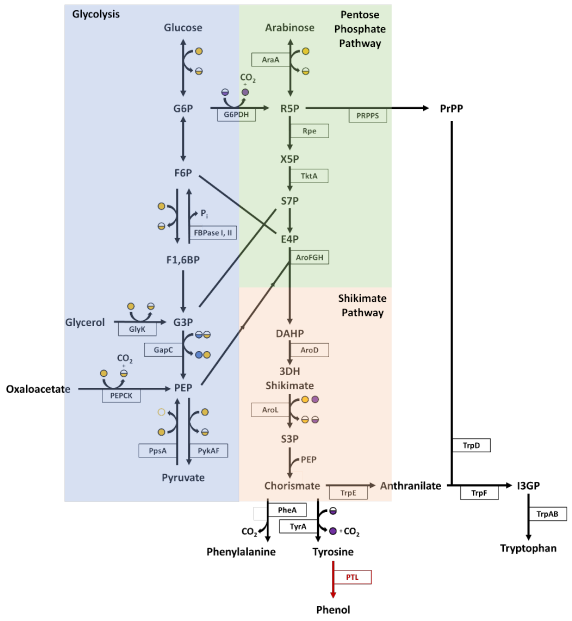


Figure 4.2: Abbreviated metabolic map of phenol biosynthesis by heterologous expression of phenol-tyrosine lyase in *E. coli*. Only enzymatic transformation found significant in this study are represented. Heterologous enzymes expressed by cell-free protein synthesis are colored red. Symbols: Full yellow circles, ATP; half yellow circles, ADP; empty yellow circles, AMP; full purple circles, NADPH; half purple circles, $NADP^+$; full blue circles, NADH; half blue circles, NAD^+. Metabolite abbreviations: G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; F1,6BP, fructose 1,6-bisphosphate; G3P, glyceraldehyde-3-phosphate; PEP, phosphoenolpyruvate; R5P, ribose 5-phosphate; X5P, xylulose 5-phosphate; S7P, sedoheptulose 7-phosphate; E4P, erythrose 4-phosphate; PrPP, phosphoribosyl pyrophosphate; DAHP, 3-deoxy-D-arabinoheptulosonate 7-phosphate; 3DHS, 3-dehydroshikimate; S3P, shikimate 3-phosphate; I3GP, indole-3-glycerol phosphate. Enzyme abbreviations with Enzyme Commission numbers: G6PDH, glucose 6-phosphate dehydrogenase (1.1.1.49); AraA, arabinose isomerase (5.3.1.4); PRPPS, phosphoribosyl pyrophosphate synthase (2.7.6.1); Rpe, ribulose 5-phosphate 3-epimerase (5.1.3.1); TktA, transketolase 1 (2.2.1.1); FBPase I, fructose 1,6-bisphosphotase class I (3.1.3.11); FBPase II, fructose 1,6-bisphosphotase class II (3.1.3.11); GlyK, glycerol kinase (2.7.1.30); GapC, glyceraldehyde-3-phosphoate dehydrogenase (1.2.1.12); PEPCK, phosphoenolpyruvate carboxykinase (4.1.1.49); PpsA, phosphoenolpyruvate synthase (2.7.9.2); PykAF, pyruvate kinase (2.7.1.40); AroFGH, deoxy-D-arabinoheptulosonate 7-phosphate synthase (2.5.1.54); AroD, 3-dehydroquinate dehydratase (4.2.1.10); AroL, Shikimate kinase II (2.7.1.71); PheA, chorismate mutase/ prephenate dehydratase (5.4.99.5/4.2.1.51); TyrA, chorismate mutase/prephenate dehydrogenase (5.4.99.5/1.3.1.12); TrpD, anthranilate phosphoribosyltransferase (2.4.2.18); TrpE, anthranilate synthase component 1 (4.1.3.27); TrpCF, multifunctional fusion protein (4.1.1.48/5.3.1.24); TrpAB, tryptophan synthase (4.2.1.20); PTL, phenol-tyrosine lyase from *Pasteurella multocida* (4.1.99.2).

Table 4.1: Composition of each EzRich derived media.

Growth Condition	Supplement EZ	ACGU mix	Carbon Source
EzRich	1x	+	11 mM Glucose
EzGlc	1x	+	100 mM Glucose
XtraGlc	2.5x	+	100 mM Glucose
AAA	-Trp, Tyr, Phe	+	100 mM Glucose
ACGU	1x	-	100 mM Glucose
EzAra	1x	+	100 mM Arabinose
EzGly	1x	+	100 mM Glycerol
DDGlc	-Trp, Tyr, Phe	-	100 mM Glucose

al. and commercially available as "EzRich" by Teknova provides greater flexibility as each component can be individually changed [129]. A modified CFPS extract preparation protocol was developed based upon EZRich media.

Maintaining CFPS capabilities was a priority in the development of this system as *in vitro* protein expression can shorten design-build-test cycles and allow synthesis of different end products. Further, as has been demonstrated in the engineering of isoprenoid biosynthesis, tuning of expression levels of terminal synthases is an important step to optimize product yield [130]. In development of a crude cell-free system grown from defined media, several growth variables were modified and CFPS capabilities of the resulting systems were compared to the standard, YTPG CFPS system. Optimal OD600 at harvest was adjusted to compensate for a higher terminal OD600 compared to YTPG (**Figure 4.7**). As EzRich contains different molar quantities of amino acids and glucose when compared to YTPG, (**Supplemental Table 4.2**) several variants of EzRich were assessed to account for different nutrient concentrations.

CFPS yield of sfGFP from plasmid pJL1 was used to benchmark all cell-free systems generated for this study. Lower amino acid concentrations did not significantly impact CFPS capabilities (**Figure 4.8**). However, increasing the proportion of the Ez Supplement in the media beyond 1x was deleterious to CFPS yield. With minimal optimization, the EzGlc media resulted in the most comparable CFPS to YTPG and was thus selected for further characterization. Further, none of the other variant cell-free systems grown on defined media exhibited significant differences *in vitro* protein synthesis yields (**Figure 4.9**). To assess the impact of the growth conditions on the proteins involved in CFPS, the 87 proteins in the

minimal PURE system were identified in all cell-free systems and statistical differences in their abundances were measured [98]. Across cell-free systems generated for this study, 26 protein elements of the PURE system [44] were identified to be differentially abundant with a fold change of greater than two compared to YTPG in at least one condition (**Figure 4.10**). It remains unclear, which individual proteins have the largest impact on *in vitro* protein synthesis yield. However, others suggest that some variation in concentration of ribosome subunits is permissible [105], which is corroborated by these data.

Cell-free phenol yield was assessed in both YTPG and EzGlc cell-free systems (**Figure 4.3**). Additionally, the protein content of each system was measured and compared with a focus on changes within the 25 enzymes directly involved in tyrosine biosynthesis (**Figure 4.3A**). Global changes in protein content are represented in **Figure 4.3C** and in **Supplemental Table B-3**. Notably, there was a large increase in nearly all amino acid biosynthesis pathways when cell-free systems are prepared from EzGlc media compared to the YTPG extracts. These include tyrosine biosynthesis enzymes DAHP synthase (AroF, 2.5.1.54), 3-dehydroquinate dehydratase (AroD, 4.2.1.10), and TyrA, which were upregulated by 98-fold, 2.5-fold and 66-fold, respectively. However, despite these large increases in protein abundance, phenol yield only increased from $10.9\,\mathrm{mg\,L^{-1}}$ in the YTPG condition to $12.4\,\mathrm{mg\,L^{-1}}$ in the EzGlc condition (p=0.048, **Figure 4.3B**). This comparatively small increase in yield is likely caused by the addition of new carbon sinks in the form of other upregulated amino acid biosynthesis pathways.

4.4.3 Impact of carbon source on *in vitro* phenol biosynthesis

In *E. coli*, all three aromatic amino acids are derived from chorismate, the nine-carbon product of the shikimate pathway. Metabolic flux to each amino acid is regulated primarily by transcriptional control [105]. While endogenous transcription, and the associated regulation, are not expected to be present in cell-free systems, tyrosine biosynthesis is also limited by the availability of shikimate pathway precursors PEP and E4P derived from glycolysis and the pentose phosphate pathway, respectively [131].

With the goal of increasing precursor supply, two media with alternative carbon sources were prepared. The EzAra media contains the pentose sugar arabinose, which

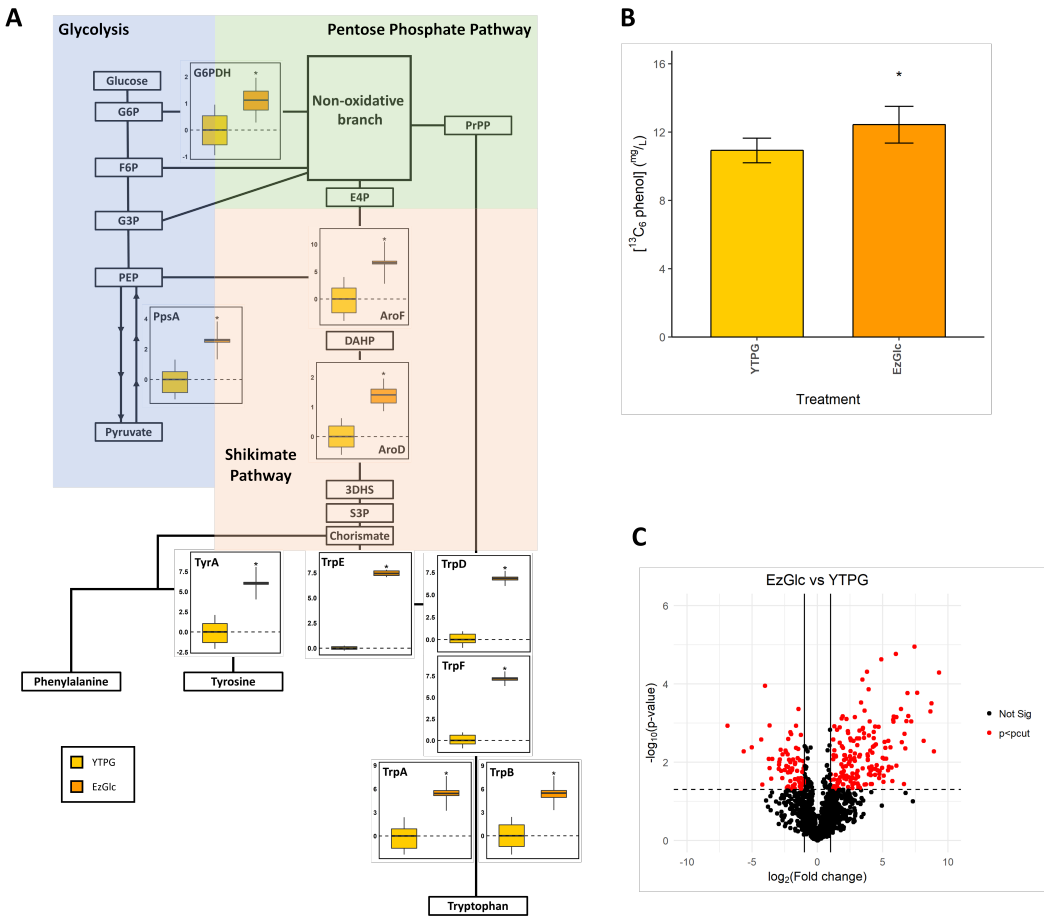


Figure 4.3: A) Comparison of protein abundance in tyrosine metabolism (including abbreviated glycolysis, pentose phosphate pathway, and shikimate pathway) between complex media YTPG and defined media EzGlc. Each bar represents the mean log2(fold change) in protein abundance in a variant growth media compared to mean protein abundance in the YTPG cell-free system. Significance was determined by a two-tailed Student's t-test compared to the YTPG cell-free system:*, $p<0.05$. Pathway enzymes not depicted can be assumed to have undergone no significant change in abundance. B) *in vitro* phenol biosynthesis from $^{13}C_6$ glucose in a one-pot CFPS-ME reaction measured at 48 hours. Only $^{13}C_6$ phenol is depicted (m/z=100.1). Values represent averages of technical replicates (n=3) and errors bars represent 1 SD. Significance was determined by a two-tailed Student's t-test compared to the YTPG cell-free system:*, $p<0.05$; ns, $p\geq0.05$. C) Volcano plot of proteomic data. Volcano plots are depicted with the log2(fold change) in abundance of each protein and the -log10(p-value) derived from performing a Student's T-test. The average abundance of each protein in the EzGlc cell-free system (n=3) was compared against the average abundance of each protein in the YTPG cell-free system (n=3). Red points show proteins which have been significantly up or down regulated at least twofold and $p<0.01$. Black points are not significantly changed. Significantly changed proteins are listed in **Supplemental Table B-3**.

was hypothesized to upregulate transketolase and transaldolase as arabinose enters *E. coli* metabolism through the pentose phosphate pathway. Media EzGly contains glycerol which is converted into the glycolytic intermediate 3-phosphoglycerate and was added to upregulate gluconeogenesis and stabilize the pool of PEP.

Changing carbon sources resulted in large increases in several proteins (**Figure 4.4**). Glycerol kinase (GlyK, 2.7.1.30) was upregulated 128-fold in the EzGly condition and arabinose isomerase (AraA, 5.3.1.4) was upregulated by three orders of magnitude in the EzAra condition. Changes within central carbon metabolism were less dramatic, but nonetheless significant. Downregulation of glyceraldehyde-3-phosphate dehydrogenase (GapC, 1.2.1.12) and upregulation of both fructose 1,6, bisphosphatase I and II (FBPase I and II, 3.1.3.11) were observed in the EzGly conditions and may result in an increased gluconeogenic potential. Further, both phosphoenolpyruvate carboxykinase (PEPCK, 4.1.1.49) and phosphoenolpyruvate synthase (PpsA, 2.7.9.2) were upregulated by growth on EzGly (**Figure 4.4A**). This suggests that growth on a triose has the potential to stabilize the pool of PEP in a cell lysate. The EzAra growth media did not result in any other substantial changes within tyrosine biosynthesis; global changes in protein content are summarized in **Supplemental Table B-4**.

Unfortunately, growth on media EzAra and EzGly resulted in decreasing two DAHP synthase isozymes (AroHF, 2.5.1.54), which would limit tyrosine production. Further, both conditions reduced abundance of TyrA, which *in vivo* engineering efforts have shown is critical to tyrosine production [132]. Although both conditions also downregulated the competing bifunctional phenylalanine biosynthesis enzyme PheA (5.4.99.5/4.1.1.51), it does not appear as though this compensated for the deleterious changes. The EzAra and EzGly cell-free systems both underperformed the EzGlc and base YTPG cell-free systems producing $8.8\,\mathrm{mg\,L^{-1}}$ and $5.8\,\mathrm{mg\,L^{-1}}$ phenol, respectively (**Figure 4.4B**). Due to their reduced phenol yield and the downregulation of key enzymes, both the EzAra and EzGly media were not studied further.

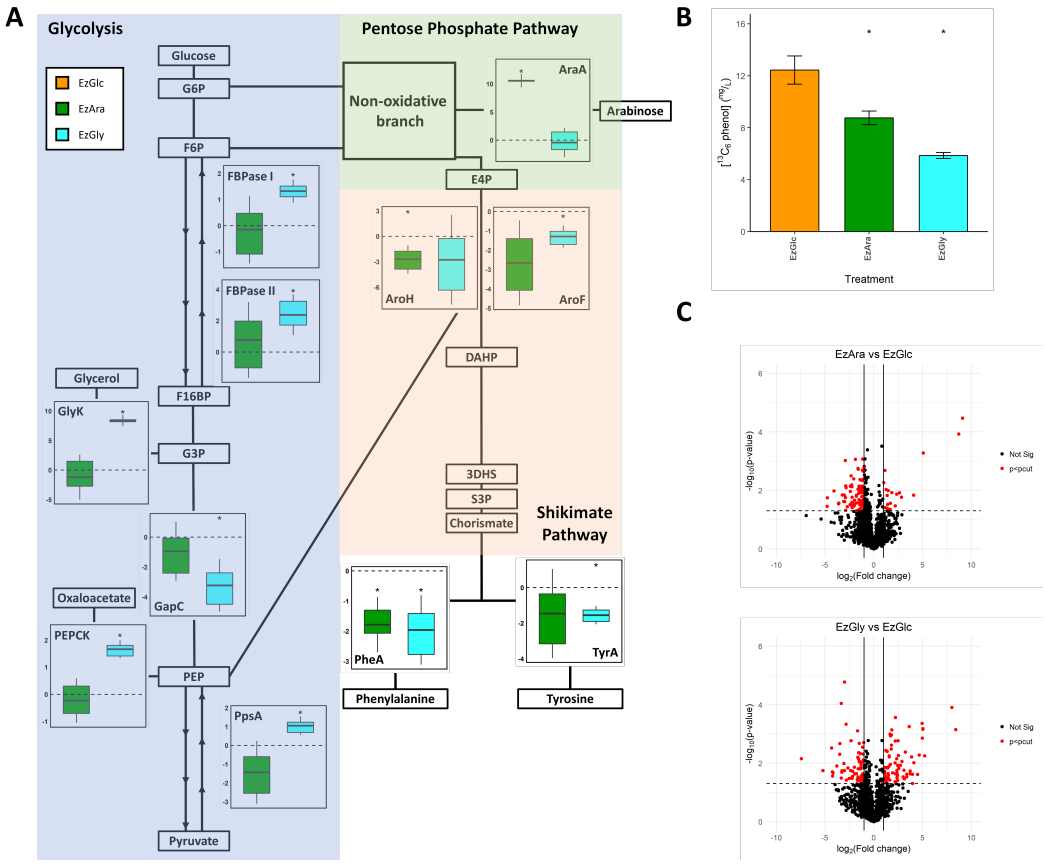


Figure 4.4: A) Comparison of protein abundance in tyrosine metabolism (including abbreviated glycolysis, pentose phosphate pathway, shikimate pathway, arabinose uptake and glycerol uptake) between EzGlc and media with variant carbon source EzAra and EzGly. Each bar represents the mean log2(fold change) in protein abundance in a variant growth media compared to mean protein abundance in the EzGlc cell-free system. Significance was determined by a two-tailed Student's t-test compared to the EzGlc cell-free system:*, $p<0.05$. Pathway enzymes not depicted can be assumed to have undergone no significant change in abundance. B) *in vitro* phenol biosynthesis from $^{13}C_6$ glucose in a one-pot CFPS-ME reaction measured at 48 hours. Only $^{13}C_6$ phenol is depicted (m/z=100.1). Values represent averages of technical replicates (n=3) and errors bars represent 1 SD. Significance was determined by a two-tailed Student's t-test compared to the EzGlc cell-free system:*, $p<0.05$; ns, $p\geq0.05$. C) Volcano plot of proteomic data. Volcano plots are depicted with the log2(fold change) in abundance of each protein and the -log10(p-value) derived from performing a Student's T-test. The average abundance of each protein in the EzGlc cell-free system (n=3) was compared against the average abundance of each protein in the EzAra (n=3) and EzGly (n=3) cell-free system. Red points show proteins which have been significantly up or down regulated at least twofold and $p<0.01$. Black points are not significantly changed. Significantly changed proteins are listed in **Supplemental Table B-4**.

4.4.4 Removing media components during growth activates biosynthetic pathways in cell lysates

Our prior work showed that abundances of glycolytic enzymes were relatively unchanged across several growth conditions [121]. However, larger shifts in protein abundance were observed outside of central carbon metabolism. With the goal of increasing the activity of aromatic compound biosynthesis *in vitro*, several dropout media were created. Media AAA is a tyrosine, tryptophan and phenylalanine dropout that was hypothesized to increase flux towards the aromatic amino acids. Media ACGU is a dropout of the EzRich nucleotide mixture. As nucleotides are synthesized from ribose-5-phosphate, this dropout was expected to increase flux to the pentose phosphate pathway and increase yield of aromatic compounds in vitro.

Media AAA performed as predicted with upregulation of rate-limiting DAHP synthases AroH and AroF as well as tyrosine-forming dehydrogenase TyrA. However, 3-dehydroquinate synthetase (AroD, 4.2.3.4) was downregulated by nearly two-fold and enzymes known to impact E4P supply were not affected (**Figure 4.5A**). The impact of media ACGU was less predictable. An absence of nucleotides reduced the abundance of both oxidative pentose phosphate pathway enzyme glucose-6-phosphate dehydrogenase (G6PDH, 1.1.1.49) and transketolase (TktA, 2.2.1.1). Intriguingly, growth on media ACGU upregulated shikimate kinase 2 (AroL, 2.7.1.71) four-fold. This effect may be elicited by the increased demand for tetrahydrofolic acid, a chorismate derivative and an essential cofactor in nucleic acid biosynthesis.

Though decreases in AroD in the AAA condition were observed, the increases in rate-limiting enzymes resulted in a 41.5% increase in phenol yield to 16.4 $\mathrm{mg\,L^{-1}}$ (**Figure 4.5B**). Further, cell extracts prepared without nucleotides synthesized phenol at amounts similar to EzGlc despite downregulation of the pentose phosphate pathway. Through evaluation of individual changes to growth media composition by both their impact on phenol yield and the changes they elicit in the lysate proteome, new composite growth conditions can be designed to target specific metabolic pathways.

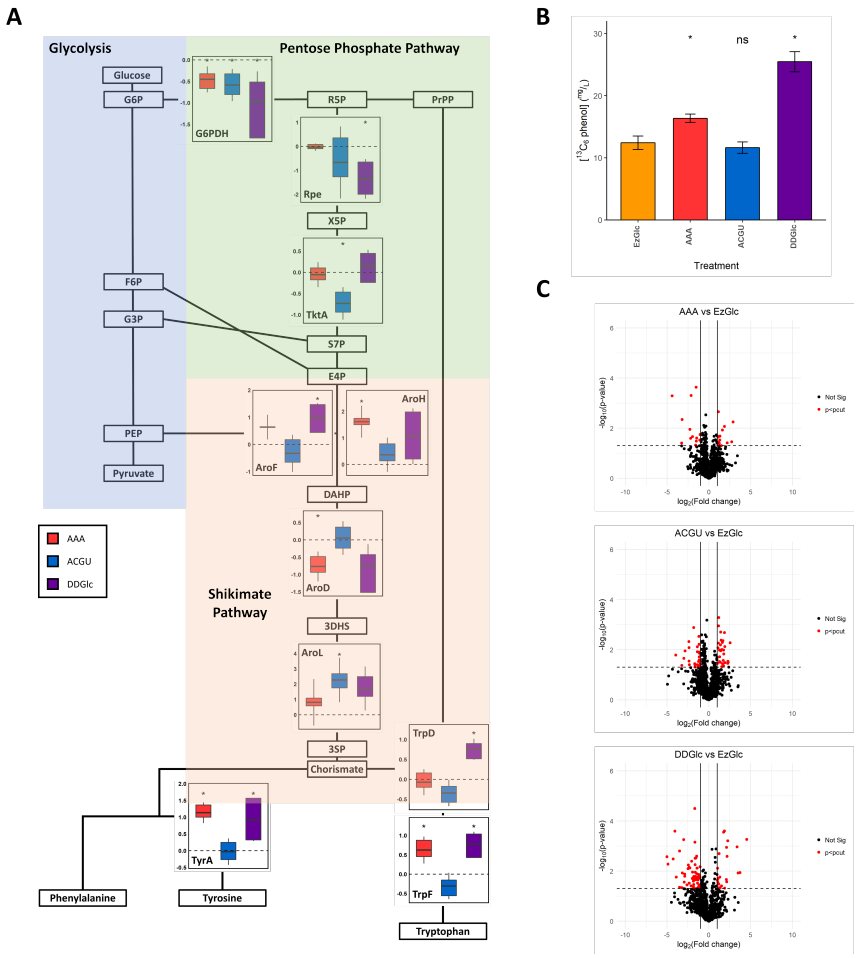


Figure 4.5: A) Comparison of protein abundance in tyrosine metabolism (including abbreviated glycolysis, pentose phosphate pathway, shikimate pathway, and aromatic amino acid biosynthesis) between EzGlc and defined media with aromatic compound dropouts AAA, ACGU and GlcDD. Each bar represents the mean log2(fold change) in protein abundance in a variant growth media compared to mean protein abundance in the EzGlc cell-free system. Significance was determined by a two-tailed Student's t-test compared to the EzGlc cell-free system:*, $p<0.05$. Pathway enzymes not depicted can be assumed to have undergone no significant change in abundance. B) *in vitro* phenol biosynthesis from $^{13}C_6$ glucose in a one-pot CFPS-ME reaction measured at 48 hours. Only $^{13}C_6$ phenol is depicted (m/z=100.1). Values represent averages of technical replicates (n=3) and errors bars represent 1 SD. Significance was determined by a two-tailed Student's t-test compared to the EzGlc cell-free system:*, $p<0.05$; ns, $p\geq0.05$. C) Volcano plot of proteomic data. Volcano plots are depicted with the log2(fold change) in abundance of each protein and the -log10(p-value) derived from performing a Student's T-test. The average abundance of each protein in the EzGlc cell-free system (n=3) was compared against the average abundance of each protein in the AAA (n=3), ACGU (n=3), and GlcDD (n=3) cell-free system. Red points show proteins which have been significantly up or down regulated at least twofold and $p<0.01$. Black points are not significantly changed. Significantly changed proteins are listed in **Supplemental Table B-5**.

While the AAA media was the only one able to increase *in vitro* phenol yield, the ACGU media upregulated unexpected enzymes within the shikimate pathway, which provoked further investigation. A media dropping out both aromatic amino acids and nucleotides with glucose as the carbon source was explored to combine the positive effects of these two sets of changes to the growth media composition. This media, dubbed double dropout glucose (DDGlc), was used to prepare a cell-free system and characterized as previously described. This new system further improved phenol biosynthesis to 25.8 mg L^{-1} and contained several uniquely upregulated enzymes (**Figure 4.5C, Supplemental Table B-5**).

The DDGlc media shares many of the upregulated proteins of its parent cell-free systems, AAA and ACGU. TyrA, AroH and AroL are all upregulated compared to the EzGlc cell-free system. While 3-dehydroquinate synthase is still downregulated in the GlcDD cell-free system, the downregulation of transketolase in the ACGU condition is not maintained in the double dropout. As there are many potential sinks of PEP, determination of the metabolic fate of PEP in the various cell-free systems will likely be necessary to further increase phenol yield.

The double dropout media results in the unique downregulation of ribulose 5-phopshate epimerase (Rpe, 5.1.3.1), which was not observed in any of the parent conditions. This change has the potential to impact E4P supply by limiting the amount of glucose which enters the pentose phosphate pathway *in vitro*. Further, the DDGlc media upregulates anthranilate PrPP transferase (TrpD, 2.4.2.18), a key enzyme in tryptophan biosynthesis which utilizes resources from both the shikimate and pentose phosphate pathway. It is possible that the observed increased flux to tyrosine is a consequence of a greater increase in flux to tryptophan. Eliminating the conversion of chorismate to anthranilate would help channel shikimate pathway products towards tyrosine.

4.4.5 CFPS of rate-limiting enzyme, AroG

Post-lysis addition of enzymes by cell-free protein synthesis not only enables synthesis of heterologous products but can also alter endogenous metabolism. Expression of additional copies of endogenous rate-limiting enzymes can improve flux towards specific pathways [133]. However, expression of multiple constructs in a single cell-free can reduce *in vitro* protein

yield. In this work, tyrosine availability was assumed to limit phenol production as PTL activity is not expected to be limiting below tyrosine concentrations approaching 1 mM [134]. When only PTL was expressed by CFPS in each cell-free system, unlabeled phenol production was relatively unchanged (**Figure S5**). This supports the assertation that the metabolic bottleneck lies upstream of the phenol-tyrosine lyase step. However when sfGFP was expressed by CFPS, in addition to PTL, both fully-labelled and unlabeled phenol yield decreased.

In addition to synthesizing additional copies of endogenous enzymes, mutants can be expressed to overcome allosteric feedback regulation. Three isozymes of DAHP synthase, AroGHF, carry out the rate-limiting condensation of E4P and PEP in aromatic amino acid biosynthesis; each isozyme is allosterically inhibited by one of the aromatic amino acids. AroG, which is sensitive to feedback inhibition by phenylalanine and makes up 80% of endogenous DAHP synthase activity, was expressed along with PTL in the DDGlc cell-free system. Because the feedback regulation on AroG is allosteric, overexpression alone did not increase phenol yield (**Figure 4.6**). However, a single amino acid mutation (146D→N) in AroG abolishes feedback inhibition [135]. CFPS of this feedback insensitive mutant along with PTL resulted in an improved phenol yield of 67.1 mg L^{-1}; this is a more than two-fold increase compared to DDGlc and a more than six-fold increase compared with the initial YTPG cell-free system. While simultaneous expression of feedback-insensitive AroG and PTL resulted in the greatest phenol yield, further optimization of CFPS yield, particularly from multiple templates, could enable further increases in productivity.

4.5 Conclusion

Crude cell-free systems offer the prospect of building and testing unique combination of cellular machinery that are freed from the constraints of cellular survival. Optimization of crude cell lysates for cell-free protein synthesis over the past 30 years has been highly successful and has increased protein yield by 25,000-fold [136]. While it remains to be seen what role cell-free systems will play in the growing bioeconomy; their unique features suggest their largest impact may be in the fields of personalized medicine and remote bioproduction

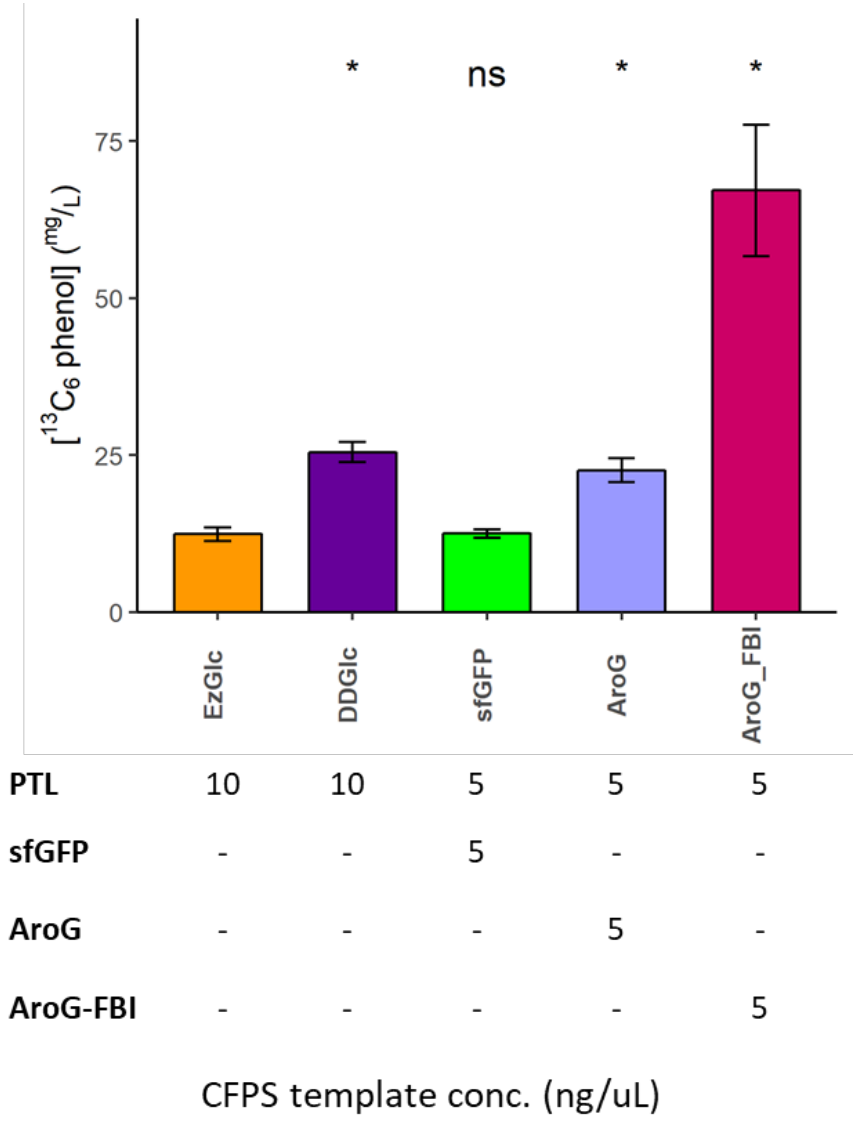


Figure 4.6: *In vitro* phenol biosynthesis from $^{13}C_6$ glucose in a one-pot CFPS-ME reaction in GlcDD cell-free system measured at 48 hours. PTL (4.1.99.2) from *P. multocida*, sfGFP, AroG (2.5.1.54) and feedback-insensitive AroG are expressed by CFPS at the provided concentrations. Only $^{13}C_6$ phenol is depicted (m/z=100.1). Values represent averages of technical replicates (n=3) and errors bars represent 1 SD. Significance was determined by a two-tailed Student's t-test compared to the GlcDD-PTL 10 ng μL^{-1} cell-free reaction:*, $p<0.05$; ns, $p\geq0.05$.

[137]. To reach this goal, the characterization of the metabolic and biochemical activities of crude cell-free systems will be paramount, especially for benchmarking complex biochemical processes like protein synthesis. This work demonstrates the potential of coupling systems-wide omics measurements to inform the engineering and design of unique biological pathways that capitalize on the underlying metabolic potential of a crude cell-free system. The proteome-wide impact of small changes to lysate preparation, which resulted in a more than six-fold increase in *in vitro* phenol biosynthesis, underscores the importance of global analytical measurements, such as untargeted proteomics, to understanding these complex, cell-like systems. Furthermore, as the demand for commercial-scale, low variability, quality-controlled cell-free systems grows, measures of lysate content and activity that are more descriptive than total protein concentration will be required [138]. Nevertheless, cell-free platforms have greater flexibility and are easier to manipulate relative to cellular platforms. With the application of burgeoning systems metabolic engineering tools, including rapid metabolomics and predictions of metabolic fluxes, bioproduction in crude cell-free systems can achieve similar improvements as those seen with protein synthesis and compete with living cell platforms [139, 140].

4.6 Acknowledgements

Research supported by the U. S. Department of Energy (DOE) Office of Biological and Environmental Research, Genomic Science Program. This research was supported in part by an appointment to the Higher Education Research Experiences Program at Oak Ridge National Laboratory for JTD. Oak Ridge National Laboratory is managed by UT-Battelle, LLC, for the U.S. DOE under Contract no. DEAC05-00OR22725.

4.7 Appendix

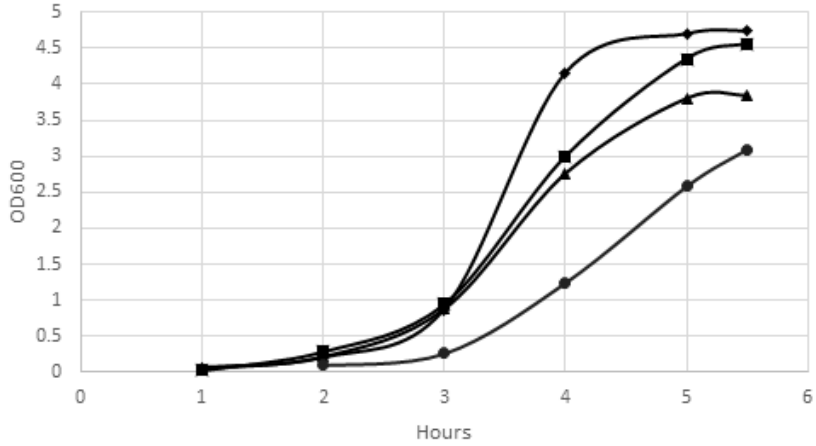


Figure 4.7: Growth profile of *E. coli* BL21De3(star) grown on preliminary cell-free system preparation media. All samples were induced to 1 mM IPTG at OD 0.6. Symbols: circles, YTPG; triangles, EzRich; squares, EzGlc; diamonds, XtrRich.

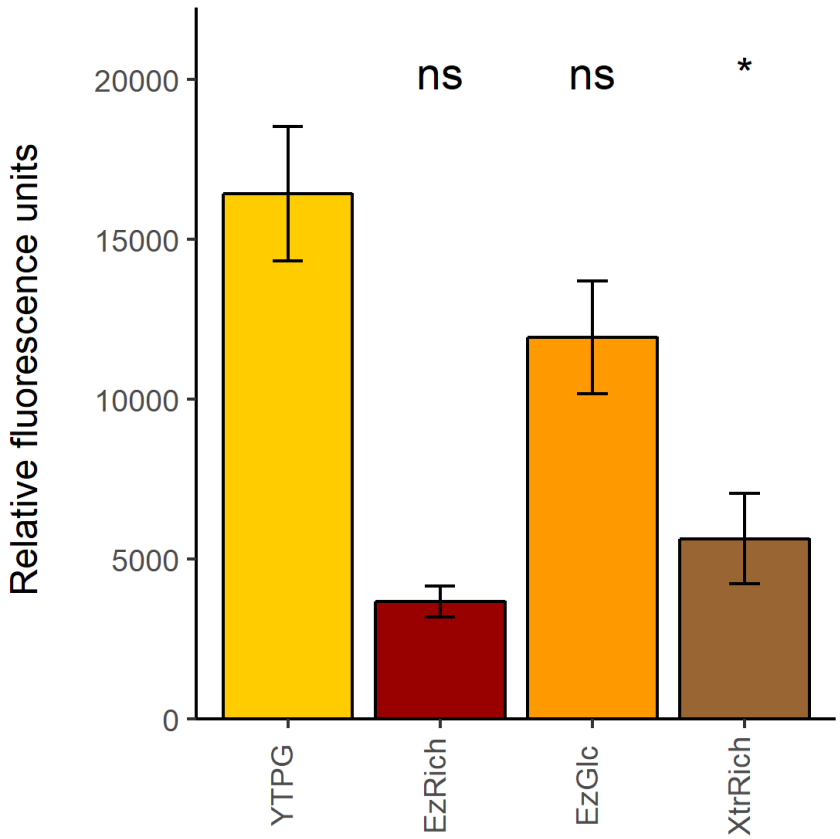


Figure 4.8: CFPS of sfGFP in cell-free systems from preliminary growth media measured at t=8 hours. Glucose (40 mM), creatine kinase (2.5 U/mL) and creatine phosphate (67.7 mM) were provided to energize the reaction. Values represent averages of technical replicates (at least n=2). Significance was determined by a two-tailed Student's t-test compared to the YTPG cell-free system:*, $p<0.05$; ns, $p\geq0.05$.

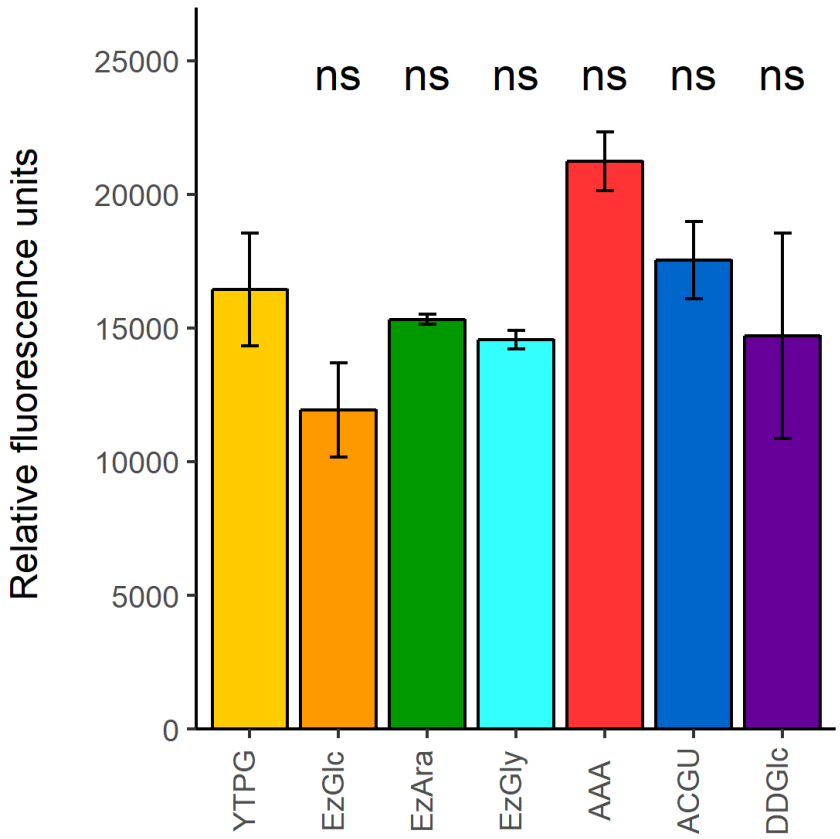


Figure 4.9: CFPS of sfGFP in cell-free systems from YTPG and EzRich variant growth media measured at t=8 hours. Glucose (40 mM), creatine kinase (2.5 U/mL) and creatine phosphate (67.7 mM) were provided to energize the reaction. Values represent averages of technical replicates (at least n=2). Significance was determined by a two-tailed Student's t-test compared to the YTPG cell-free system:*, $p<0.05$; ns, $p\geq0.05$.

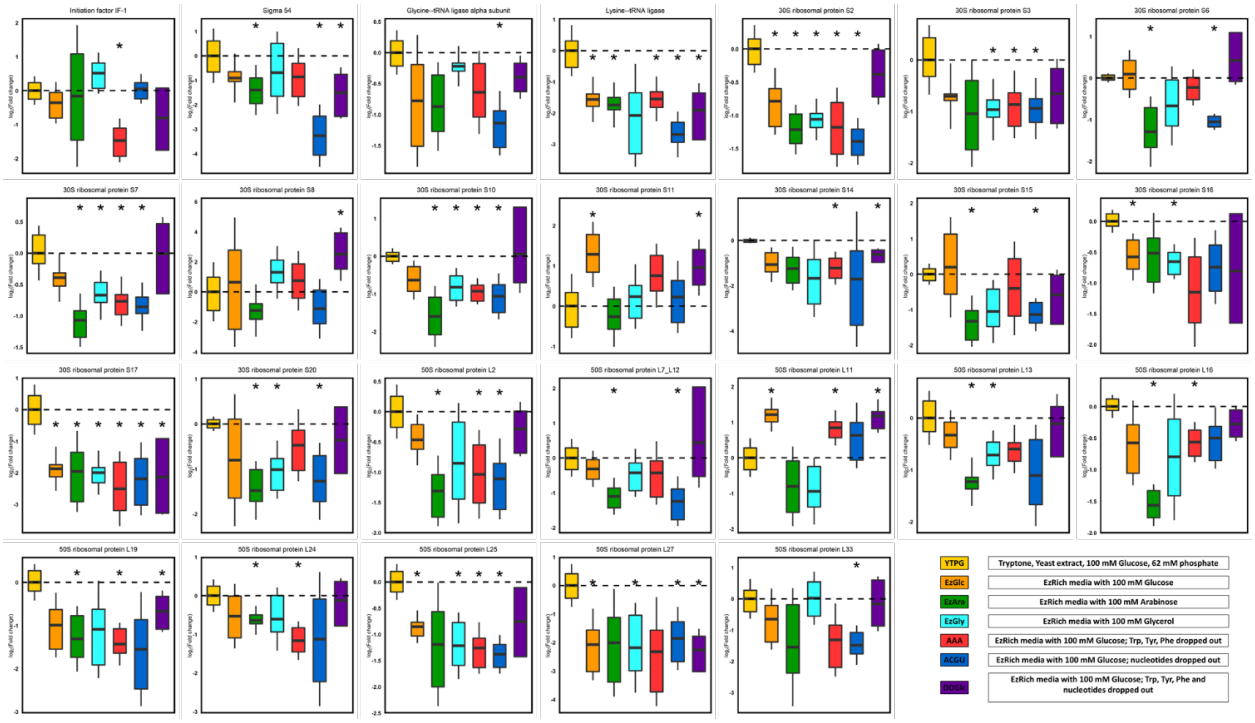
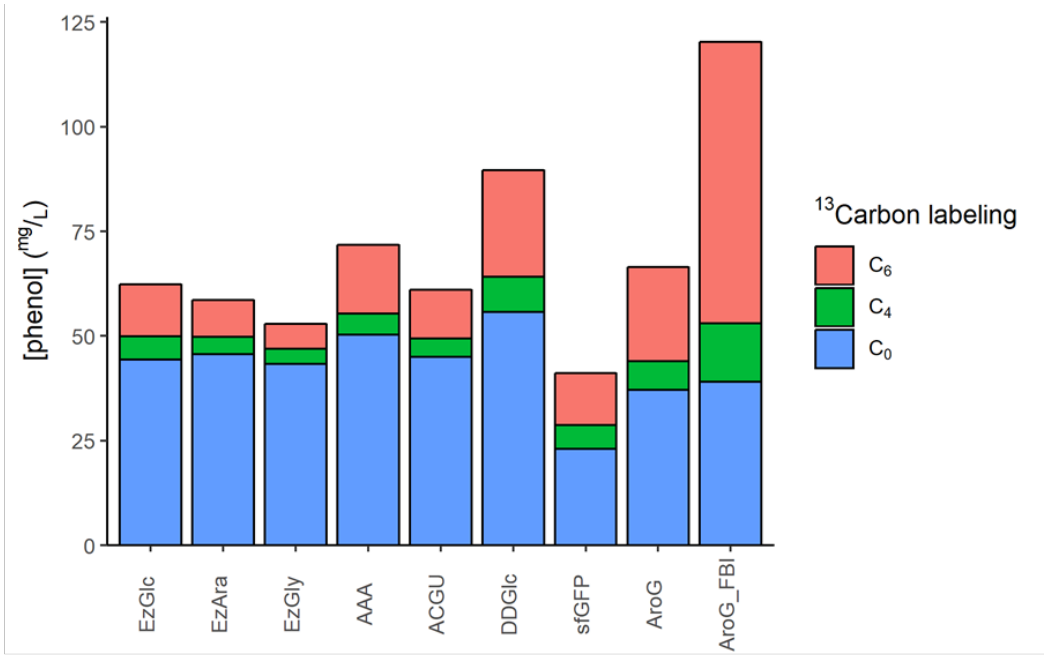


Figure 4.10: Comparison of protein abundance in transcriptional and translation machinery across variant growth conditions. Each bar represents the mean log2(fold change) in protein abundance in a variant growth media compared to mean protein abundance in the YTPG cell-free system. Significance was determined by a two-tailed Student's t-test compared to the YTPG cell-free system:*, $p<0.05$.



PTL	10	10	10	10	10	10	5	5	5
sfGFP	-	-	-	-	-	-	5	-	-
AroG	-	-	-	-	-	-	-	5	-
AroG-FBI	-	-	-	-	-	-	-	-	5

CFPS template conc. (ng/uL)

Figure 4.11: In vitro phenol biosynthesis from $^{13}C_6$ glucose in a one-pot CFPS-ME reaction in EzGlc, each of the variant growth media cell-free systems, and GlcDD with variant CPFS reactions were measured at 48 hours. PTL from *P. multocida* is expressed by CFPS. Bar chart coloration depicts $^{13}C_0$, $^{13}C_4$, and $^{13}C_6$ phenol is depicted (m/z=94.1, 98.1, 100.1). Values represent averages of technical replicates (n=3).

Table 4.2: Amino acid Composition of each EzRich derived media.

Amino acid	**2xYT conc (mM)**	**1x Ezrich conc. (mM)**	**% of 2xYT conc**	**XtrRich conc. (2.5x Ezrich) (mM)**	**% of 2xYT conc**
Ala	9.76	0.80	0.08	2.00	0.21
Arg	4.65	5.20	1.12	13.00	2.80
Asp	12.35	0.40	0.03	1.00	0.08
Cys	1.02	0.10	0.10	0.25	0.24
Glu	27.81	0.60	0.02	1.50	0.05
Gly	7.15	0.80	0.11	2.00	0.28
His	3.07	0.20	0.07	0.50	0.16
Ile	7.45	0.40	0.05	1.00	0.13
Leu	12.37	0.80	0.06	2.00	0.16
Lys	10.00	0.40	0.04	1.00	0.10
Met	3.06	0.20	0.07	0.50	0.16
Phe	5.47	0.40	0.07	1.00	0.18
Pro	13.97	0.40	0.03	1.00	0.07
Ser	10.30	10.00	0.97	25.00	2.43
Thr	7.43	0.40	0.05	1.00	0.13
Trp	1.21	0.10	0.08	0.25	0.21
Tyr	2.08	0.20	0.10	0.50	0.24
Val	10.17	0.60	0.06	1.50	0.15

Chapter 5

Describing the chemical diversity of lasso peptides in the root microbiome of Populus deltoides with cell-free peptide synthesis and untargeted peptidomics

A version of this chapter by Benjamin P. Mohr, Patricia M. Blair St. Vincent, Richard J. Giannone, Dale A. Pelletier, Robert L. Hettich, and Mitchel J. Doktycz is submitted to Microbial Biotechnology

PMBSV carried out initial discovery of lasso peptide BCGs and performed MALDI experiments. BPM performed bioinformatics and cell-free protein synthesis experiments. BPM and RJG performed mass spectroscopy. BPM analyzed mass spectroscopy data with assistance from RJG and RLH. BPM and MJD wrote the article with assistance from DAP,PMBSV, and RJG.

5.1 Abstract

The interactions with and within the microbiomes of plant roots can promote desirable host traits including drought tolerance, disease resistance and increased biomass yield. These interactions can be mediated by natural products expressed by the host plant and the microbes that populate its rhizosphere. Engineering the production of natural products from the plant root microbiome to enhance crop productivity and robustness would be a great boon to agriculture. However, connecting molecules to specific functions is often limited by both microbial culturability as well as identification of silent biosynthetic gene clusters encoding small but functional polypeptides that can exhibit non-canonical secondary structures. Here we develop a high-throughput workflow incorporating bioinformatics, cell-free protein synthesis, and untargeted peptidomics for the rapid characterization of lasso peptides – unique cyclical natural products formed by the post-translational processing of ribosomally synthesized peptides by clusters of tailoring enzymes present within the root microbiome of *Populus deltiodes*. Eighty putative lasso peptide gene clusters were identified; cluster diversity was driven by both precursor peptide sequence and presence of putative tailoring enzymes identified within the genetic neighborhood encoding the minimal biosynthetic machinery. Cell-free protein and precursor peptide synthesis was investigated as a tool for combinatorial expression and characterization of lasso peptide biosynthetic enzymes, including a novel sulfotransferase identified in twelve Paenibacillaceae. Assessment

of cell-free production through untargeted peptidomics revealed that identification of lasso peptides and their modifications was limited by *in vitro* peptide synthesis fidelity. *In vitro* expression enables the high-throughput discovery of natural product structural and chemical diversity from genome sequences; however, cell-free synthesis of short peptides and their characterization by associated analytic methods remains challenging, but is necessary to further the *in vitro* characterization of these ribosomally synthesized and post-translationally modified peptide natural products

5.2 Introduction

The plant root microbiome is flush with natural products (NPs) excreted by both fungal and bacterial members to influence their environment and the community around them. The structure and function of some of these compounds, such as indoles, are well described, but a larger number are less understood [141] al., 2016). Automated genome sequence annotation, based on homology modeling or sequence similarity, has been used to predict the molecules that members of a microbiome could produce [142]). This is a robust method for identifying well-described pathways in new organisms, but the functions of many predicted open reading frames (ORFs) remain unknown [143]. Structural prediction of NPs is particularly challenging for classes that have few described members or harbor uncommon modifications, including ribosomally synthesized and post translationally modified peptides (RiPPs) [2]).

RiPPs are DNA-templated NPs. To form an active product, a ribosomally produced precursor peptide is sequentially modified by various enzymes. Canonically, modification of a C-terminal core region is guided by the N-terminal leader peptide, which is cleaved upon maturation [144, 145]. This separation between recognition sequence and modification site allows for enzyme promiscuity, but also complicates prediction of product structure. Fortunately, development of new genome mining tools which use biosynthetic gene cluster (BGC) architecture to identify regions of interest have allowed identification of numerous RiPP BGCs [146, 147].

Lasso peptides constitute a rapidly growing subclass of RiPPs, that have numerous newly discovered but BGCs [148].The interaction of three protein domains thread the C-terminal tail of the precursor peptide through an internal macrolactam ring, forming the lasso-like structure which gives the subclass its name (**Figure 5.1**). These three functional domains are split across two to three proteins including a recognition element, called the RiPP recognition element (RRE) or E; a transglutaminase, called B; and a cyclase, called C [149].

The unique three-dimensional structure of this peptide makes it an attractive scaffold for bioengineering efforts; to this end, several lasso peptides have been isolated and structurally characterized with mass spectrometry (MS) and NMR [150, 151, 152]; however functional characterization remains elusive. Several lasso peptides have antimicrobial or antiviral properties, but most have no described activity and their natural function remains unknown [153, 154, 155]. As new lasso peptides are discovered, additional peptide modifications have been observed beyond the lariat knot. These modifications include acetylation, phosphorylation, citrulenation, methylation and glycosylation; most of these modifications are performed by enzymes within the lasso peptide BGC, but some are distally encoded [156, 157, 145]. Additional as-yet uncharacterized ORFs have also been observed within the gene neighborhood of many lasso peptides [147]. The cryptic nature of these additional modifications may contribute to both the lack of understanding regarding the natural function of lasso peptides as well as their sub-par identification by contemporary analytical methods.

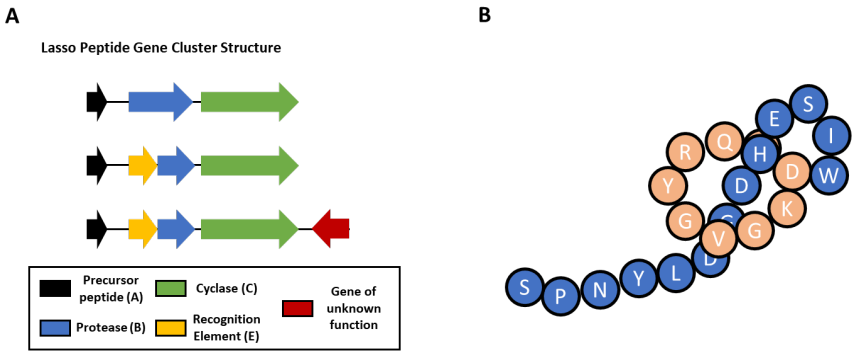


Figure 5.1: A. General structure of biosynthetic gene clusters for lasso peptides. B. Cartoon of the expected structure the mature lasso peptide from PDC88. The predicted macrolactam ring is colored orange and the tail is colored blue.

The development of a workflow to characterize the sequence, structure and modifications of lasso peptides would support genome mining efforts and further elucidation of their natural function. Lasso peptides are commonly isolated from cultured natural producers or expressed in a heterologous host [158], but both methods suffer from lengthy purification processes and time-consuming cell culture or genetic engineering steps. Conversely, coupling cell-free protein synthesis (CFPS) of lasso peptide precursors, maturation enzymes and putative tailoring enzymes with mass spectroscopy would enable rapid determination of peptide structure and chemical modification. This method would also facilitate quick production and validation of variant precursor peptides, which is particularly of interest for interrogation of the substrate specificity of maturation and tailoring enzymes.

Towards the characterization of the NP diversity of the root microbiome of *Populus deltoides*, genomes of 407 root bacterial isolates were mined for the presence of lasso peptide BGCs. The resulting clusters were grouped based upon sequence similarity and BGC architecture. Enzyme and peptide sequence motifs were identified and used to predict the activity and specificity of lasso peptide modifying enzymes, including a novel putative sulfotransferase. CFPS was investigated as a tool for characterization of peptide NP diversity within this collection. Untargeted peptide sequencing by LC MS/MS was explored as a method to confirm peptide production and modification [98].

5.3 Materials and methods

5.3.1 Identification of lasso peptide gene clusters.

Genomes were downloaded from the Joint Genome Institute Integrated Microbial Genomes and Microbiomes database (IMG, https://img.jgi.doe.gov/) in FASTA format (November 2018) and were submitted for analysis using antiSMASH 3.0 [142]. Lasso peptide BGCs were identified by analyzing each genome for the presence of an asparagine synthase (Pfam 00733). The asparagine synthase proteins were identified as lasso cyclase proteins if the gene neighborhood contained a transglutaminase (Pfam 13471). Sequence alignments and neighbor joining trees were generated from protein FASTA files with R (version

3.5.3) packages ape, phangorn and seqinR and visualized in iTOL (http://itol.embl.de/) [159, 160, 161, 162]. Neighbor joining trees were resampled 100 times and branch lengths with bootstrap supports less than 50 were set to 0.

5.3.2 Co-occurrence analysis.

Gene neighborhoods were analyzed in IMG. Co-occurrence based on protein families was tabulated for genes +/- 10 ORFs from the lasso cyclase (Pfam 00733). Protein families were identified as follows: PEP carboxykinase, SSF53795; NTP_transf_5, Pfam 14907, sulfotransfer_3, Pfam 13469. Frequency of occurrence was calculated for each protein family by dividing the number of counts in the collection by the total number of BGCs.

5.3.3 Matrix-assisted laser ion desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) analysis.

Cells were grown to stationary phase in Luria Broth or R2A media [163] and small molecules were extracted with methanol. Extracts were prepared for MALDI-TOF-MS by drying aliquots and resuspending in 50% acetonitrile. Samples were co-spotted on a MALDI-TOF-MS stainless steel target plate with saturated α-cyano-4-hydroxycinnamic acid in 75% (v/v) acetonitrile, 0.1% formic acid. Spectra were acquired on a Bruker Autoflex Speed II equipped with flexControl software (Bruker Daltonics) in positive ion reflector mode with a mass range of 800-6000 Da. Data were analyzed in flexAnalysis (Bruker Daltonics).

5.3.4 Sequence similarity networks (SSNs).

EFI-EST (http://efi.igb.illinois.edu/efi-est/) [164] was used to generate SSNs for lasso peptide precursors and transferases, which were visualized with Cytoscape v. 3.5.1 [165]. E-value for all-by-all BLAST was 5 for transferases and 10 for precursor peptides. Edge lengths are proportional to shared sequence identity. Nodes were colored based on phylogeny.

5.3.5 MEME suite motif identification.

MEME version 5.0.5 (http://meme-suite.org/tools/meme) [166] was run in default mode from protein FASTA files, identifying 3 motifs for both precursor peptides and transferases. E-values, representing the product of all p-values for identified motifs on each transferase protein sequence and p-values for each motif in the precursor peptides are listed in **Supplemental Table C-2** and **Supplemental Table C-3**, respectively.

5.3.6 Cloning.

PDC88 genes B, C, RRE and the putative sulfotranferase were codon-optimized and synthesized by Integrated DNA Technologies (IDT, Coralville, IA). Pjl1 was amplified with primers Pjl1_gib_fwd and Pjl1_gib_rev. Genes were assembled into Pjl1 with Gibson assembly master mix (NEB) according to protocol and transformed into TOP10 competent *E. coli* (Thermo Fischer). Plasmid was isolated from overnight culture with a Qiagen mini-prep kit according to protocol.

5.3.7 Cell-free protein synthesis.

Protein synthesis in the PURE systems (New England Biolabs, Ipswich, Massachusetts, USA) was carried out from PCR products at 37 °C for 2 hours with 250 ng template DNA per 25 µL reaction. Microcin J25 and klebsidin clusters were amplified with primers PURE_T7_fwd and PURE_T11_rev, PDC88 cluster was amplified with PURE_T7_fwd and PURE_pjl1_rev. *E. coli* S30 lysate was prepared as described previously [98]. CFPS from S30 lysate was carried out with plasmid DNA at 37 °C for 2 h with 150 ng DNA per 15 µL reaction. Synthesis of microcin J25 and klebsidin was carried out in one-pot reactions. PDC88 cluster enzymes were expressed individually, 5 mM 3'-phosphoadenosine-5'-phosphosulfate was added to reactions containing a sulfotransferase, and then combined with 10 ng of synthetic PDC88_A1 peptide (Genscript, Piscataway, New Jersey, USA) and reacted at 37 °C for an additional h. CFPS reactions were cleaned-up with a 10 kDa molecular weight cutoff size exclusion filter (Microcon YM-10, Millipore, Billerica MA) or a C18 desalting spin column (Thermo Scientific, Waltham, Massachusetts, USA), the solvent

was then evaporated to dryness and peptides were resuspended in 10 µL of solvent A (5% (v/v) acetonitrile, 0.1% formic acid) for subsequent analysis.

5.3.8 Untargeted peptide measurements.

Reactions were analyzed by reverse-phase (RP) liquid chromatography in line with nanospray tandem mass spectrometry (MS/MS) using an Q Exactive Orbitrap Plus MS (Thermo Scientific, Waltham, Massachusetts, USA). 2 µL of each peptide sample was directly loaded to a RP column packed with 5 cm of C18 resin (5 micron Kinetex C18, Phenomenex, Torrance, CA) over 30 minutes as previously detailed [167]. Bound and washed peptides were then separated by an organic gradient to 50% solvent B (70% (v/v) acetonitrile, 0.1% formic acid) over 105 minutes on an in-house pulled nanospray emitted pack with 30 cm of Kinetex C18 resin flowing at 300 nL/ min. The MS was operated in positive ion mode with data-dependent acquisition (one full scan at 70K resolution followed by 5 MS^2 scans, 1.8 m/z isolation window, 0.3 m/z isolation offset, normalized collision energy of 27, unassigned charged states and charge states 1, 7 and 8 were excluded from analysis). Extracted ion chromatograms (XICs) were generated from MS^1 spectra with Xcalibur (ThermoScientific).

5.3.9 Database dependent peptide identification.

MyriMatch (version 2.1.138, nonspecific digestion, unlimited missed cleavage sites, dynamic modifications - methionine + 15.9949 Da, tyrosine + 79.9568 Da) was employed to obtain peptide identifications by matching observed MS/MS spectra to predicted spectra from a database of either 4156 *Escherichia coli* BL21 (DE3) protein sequences (proteome UP000002032) or the 99 protein components of the PURE system appended with the heterologous protein sequences, common contaminants and decoy sequences [69]. Peptide spectrum matches (PSMs) were filtered and assigned to proteins using IDPicker, version 3.1.599 (PSM maximum false discover rate 5%) [100]. Identified peptide sequences were exported and processed in R (version 3.5.3).

5.3.10 *De novo* peptide sequencing.

PEAKSX (Bioinformation Solutions Inc, Ontario, Canada) was used to perform *de novo* peptide sequencing on each LC MS/MS run. Unassigned high confidence de novo peptides were assigned potential post-translation modifications with PEAKS PTM.

5.4 Results and Discussion

5.4.1 Analysis of lasso peptide gene cluster architecture within the Populus root microbiome.

Previously, 339 bacterial isolates from the Populus rhizosphere and endosphere were sequenced and annotated with the Integrated Microbial Genomes (IMG) pipeline and mined for BGCs using antiSMASH 3.0 [168, 169]. An additional sixty-eight newly sequenced isolates have been added to this collection and searched as described previously (**Supplemental Table C-1**). AntiSMASH identified eighty-four (84) lasso peptide cyclases from sixty-three (63) unique organisms. Clusters were selected that met the minimal requirements to synthesize a lasso peptide; specifically, the lasso cyclase gene neighborhood must contain a transglutaminase domain (B), a protein with a PQQ domain (E) and at least one potential lasso precursor peptide. In some cases, the transglutaminase and RRE are fused, but in many cases, such as in *Streptomyces*, the RRE is a separate ORF [147]. Eighty (80) intact lasso peptide BGCs were identified and a phylogenetic tree of the 80 cyclases from these clusters was constructed (**Figure 5.2**). The lasso peptide BGCs are distributed among 3 clades: clade 1 consists of cyclases from Actinobacteria, clade 2 consists of cyclases from α- and γ-Proteobacteria, and clade 3 includes cyclases from α-Proteobacteria, Bacteriodetes and Firmicutes. Lasso peptides have been isolated from β-Proteobacteria [151]; however, no lasso cyclases were detected in the sequenced β-Proteobacteria in this collection

Of the lasso peptide BGCs identified, most fall within the diversity outlined by Tietz *et al.* [147]. Precursor peptide sequences ranged in length from 13-32 aa, with an average of 21($\pm$4) aa. While the dataset of sequenced organisms is biased for Proteobacteria based on the prevalence of organisms within the the Populus rhizosphere and thus within the sequenced

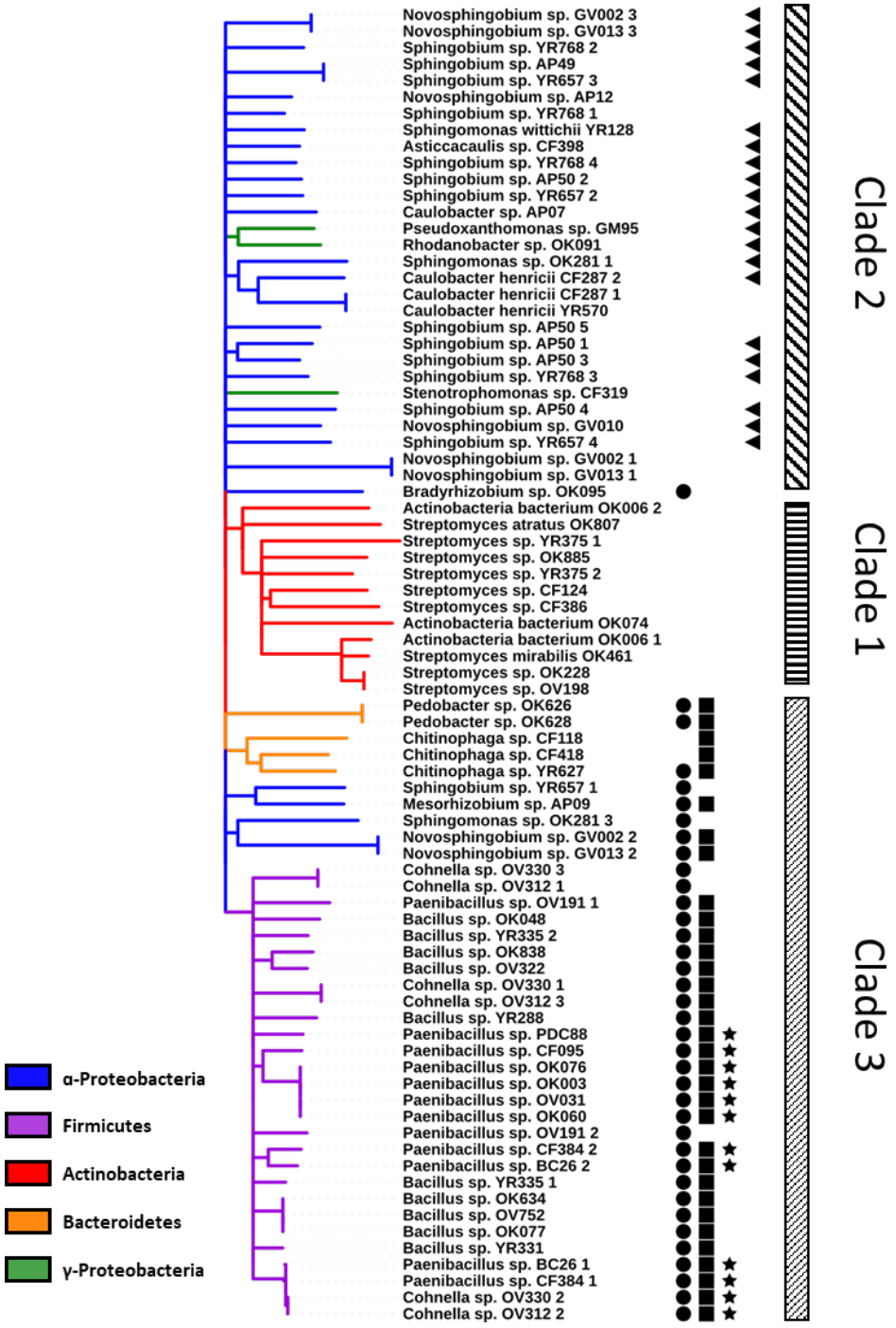


Figure 5.2: Neighbor joining tree of 80 lasso cyclases. Branches are color coded according to class. Tailoring enzymes found in the gene neighborhood of each lasso cyclase are mapped following the node name as follows: kinase, SSF53795, circle; nucleotidyltransferase, pfam 14907, square; sulfotransferase, pfam 13469, star; isopetidase, COG1506 , triangle.

organism collection (**Supplemental Figure 5.8**). Streptomyces make up only 3.9% of the isolates and harbor a disproportionate number (69%) of lasso peptide BGCs [168]. Given the large number of BGCs within the genomes of these organisms and the historical precedence for discovering complex NPs from Streptomyces, this result is not surprising (Newman and Cragg, 2016). Clade 1 is made up of the lasso peptides from Streptomyces species, many of which are predicted to contain disulfide bonds. As previously hypothesized, class I and III lasso peptides occurr extensively in Actinobacteria (**Table 5.1**) [158]. Interestingly, a peptide in clade 3 from *Bacillus* sp. YR335 is predicted to contain a disulfide bond as well.

Table 5.1

In phyla other than Actinobacteria, lasso peptide diversity is generated through tailoring enzymes; conserved enzymes present in the gene neighborhood of the minimal machinery that post-translationally modify the core peptide (**Figure 5.2**). Presence of tailoring enzymes and their putative activities are often shared among related BGCs. However, the predicted functions of most tailoring enzymes have not been validated and the extent of lasso peptide modification beyond the lariat knot is unclear [147, 157]. Furthermore, the presence of BGCs with different conserved sets of tailoring enzymes implies the potential of different functional classes of lasso peptides. Exploring this diversity is critical to understanding lasso peptide structure and function.

Clade 2 is comprised of Proteobacteria, Sphingomonadaceae, Caulobacteracea and Xanthomonadaceae species, which possess a conserved isopeptidase (COG1506) [170]. A similar enzyme from the lasso peptide BGC in *Asticcaucalis benevestitus* was found to catalyze macrolactam ring opening of the mature lasso peptide [171]. Isopeptidase activity has been described towards several lasso peptides and recognition is based upon topology

Table 5.1: Overview of 3 clades of lasso peptide biosynthetic gene clusters identified.

Clade	Lasso Cyclases	Disulfide	Kinases	Nucleotide trans-ferases	Sulfo trans-ferases	Isopeptidases
Clade 1	12	5	0%	0%	0%	0%
Clade 2	27	0	0%	0%	0%	78%
Clade 3	41	1	90%	80%	29%	0%
Total	80	6	46%	41%	15%	26%

and not sequence; however, the biological function of this interaction remains enigmatic [172, 173].

The third clade encompasses most of the lasso peptide BGCs identified in this collection including 28 clusters from Firmicutes, 5 clusters from Bacteriodetes and 8 clusters from α-Proteobacteria. Most of these clusters contain a kinase enzyme (SSF53795) [174] and frequently an uncharacterized nucleotide transferase (pfam 14907) [175]. Recently, a similar enzyme was described with kinase activity towards the lasso precursor peptide C-terminal serine [157]. This lasso peptide, paeninodin, is phosphorylated by the putative kinase prior to cyclization and leader peptide cleavage. The C-terminal serine can be phosphorylated several times *in vitro*; however, this peptide modification state has not been observed *in vitro* [176]. While the function of this modification is unclear, the presence of other conserved transferases in the gene neighborhood of paeninodin-like peptides suggests that the observed phosphorylation is part of a larger tailoring process that provides function. Intriguingly, 12 clusters from Paenibacillaceae contained a conserved sulfotransferase domain (pfam 13469) that has not previously been observed in lasso peptide BGCs. .

5.4.2 MALDI-TOF-MS screen for lasso peptides.

The presence of a lasso peptide BGC in a genome does not indicate that the corresponding metabolite is produced; often, NP BGCs are cryptic and must be activated or expressed in a heterologous host. In the case of lasso peptides, many systems require heterologous expression [177]. However, unlike predicted structures from genomic data, natively expressed NPs provide information about the structure of a fully mature product. To survey the extent of lasso peptide expression and modification without the presence of community members, individual isolates were screened for constitutively expressed lasso peptides.

In brief, bacteria were grown in liquid media to stationary phase, and the cells were extracted with methanol. Samples were analyzed by MALDI-TOF-MS for the presence of lasso peptide masses as predicted by the lasso peptide prediction program RODEO [147]. In a preliminary screen, masses +/- 25 m/z different from the predicted mass were identified in 8 bacterial isolates; the associated spectra are cataloged in (**Supplemental Figure 5.9**).

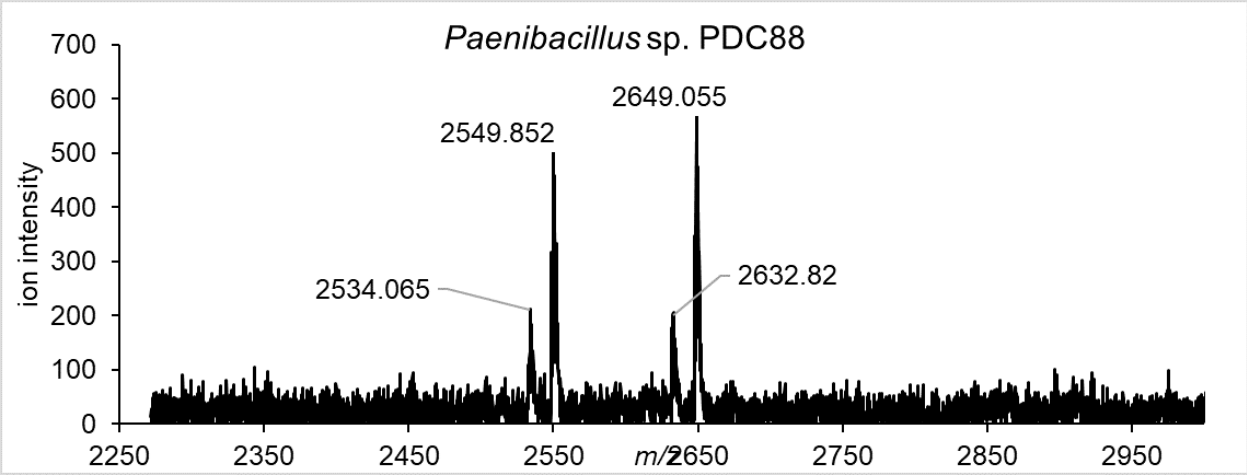


Figure 5.3: MALDI-TOF mass spectra of product from PDC88. Masses of interest are labeled.

Table 5.2: Comparison of the identified masses and the predicted masses for modified lasso peptide products.

PDC88	Monoisotopic mass (M)	($M+H^+$) (m/z)	($M+Na^+$) (m/z)	Measured mass (m/z)	Error (m/z)
Linear Core	2548.198	2549.206	-	2549.852	-0.646
Cyclized	2530.188	2531.195	-	2534.065	-2.870
Sulfated PDC88					
Linear Core	2628.155	-	2651.145	2649.055	2.090
Cyclized	2610.140	-	2633.130	2632.820	0.310
Phosphorylated PDC88					
Linear Core	2628.165	-	2651.154	2649.055	2.099
Cyclized	2610.154	-	2633.144	2632.820	0.324

Amongst the nine identified peptides; six were from clade 1, two from clade 2 and one from clade 3. The single peptide from clade 3 was produced by *Paenibacillus sp.* PDC88. In addition to the two masses which approximate the linear (2549.85 Da) and cyclized PDC88 peptide (2534.07 Da), a pair of masses approximately 100 Da greater than the predicted linear and cyclized peptide were detected (**Figure 5.3**). Based upon the presence of a kinase and putative sulfotransferase enzyme in the gene neighborhood of the PDC88 lasso cyclase, these additional peaks were hypothesized to be associated with the linear and cyclized sodiated mass of a tailored PDC88 lasso peptide. Phosphorylation and sulfation of hydroxyl groups cause similar changes in molecular mass, resulting in an increase in the monoisotopic mass of 79.966 m/z and 79.956 m/z, respectively (**Table 5.2**). Due to the high background in the methanolic extracts and low peak resolution, the identity of this modification could not be determined. To determine the plausibility of a sulfation occurring on the PDC88 lasso peptide, the peptides and gene neighborhoods of clade 3 were further analyzed.

5.4.3 Sequence analysis of precursor peptides in clade 3.

As the discovered sulfotransferase has not previously been associated with lasso peptide biosynthesis, we investigated the similarity between the 12 sulfotransferases and the putative lasso peptide kinases and nucleotide transferases found in the collection to search for shared recognition or substrate binding pockets. A sequence similarity network (SSN) was generated comparing all of the transferases to each other (**Figure 5.4**) [164]. At a low alignment score threshold (10^{-20}), nucleotidyltransferases from Firmicutes and Proteobacteria and kinases from Firmicutes, Proteobacteria and a member of Bacteriodetes cluster together, potentially due to their shared functional domain, but at a higher threshold (10^{-60}) the across-class clustering falls apart. Amino acid sequence motifs were identified with MEME [166]; several sequence motifs were identified within each group of transferase, but no motifs were found across all transferases (**Supplemental Table C-2**). No shared peptide recognition or binding domains were identified on the transferases; however, the lack of a shared recognition domain on the essential proteins for lasso peptide maturation suggests this is not required for activity [149].

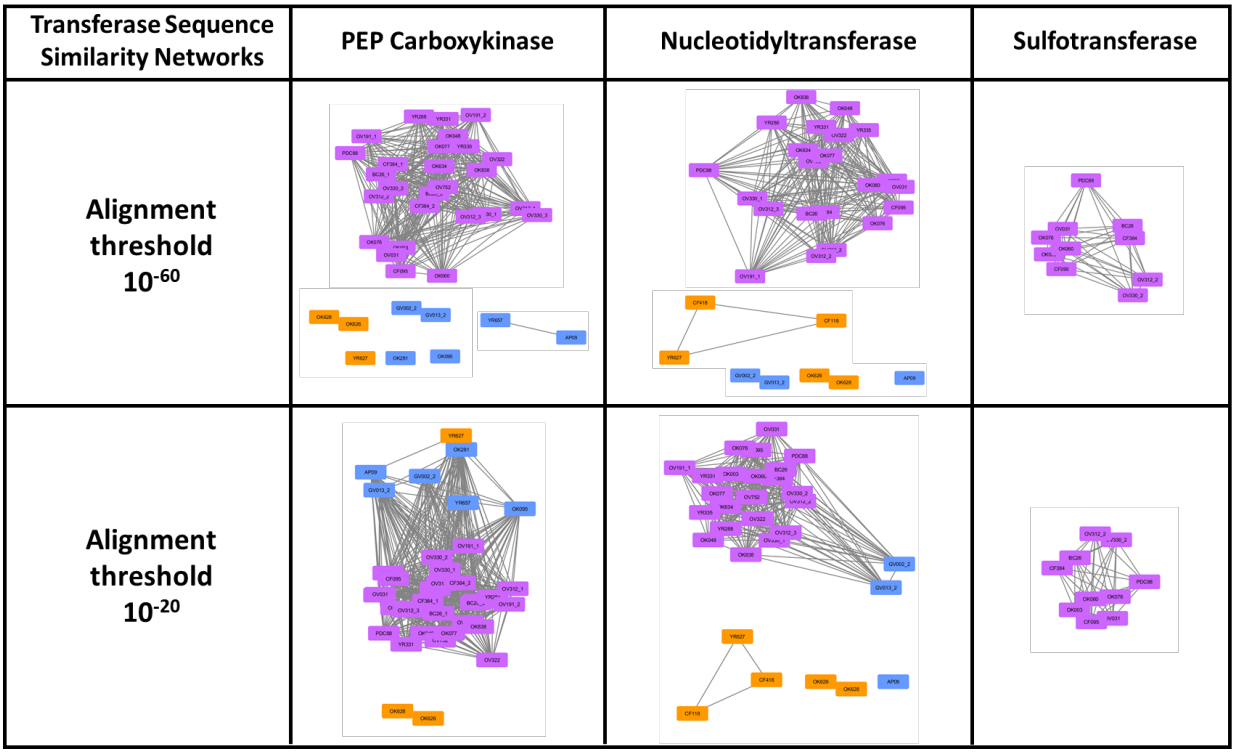


Figure 5.4: Sequence similarity network (SSN) for lasso peptide tailoring transferases in clade 3. Edge lengths are proportional to the percent sequence identity shared between each node. Networks are shown at two alignment thresholds (e=10^{-60} and e=10^{-20}). Nodes are labeled with a unique isolate identifier and a cluster number. Nodes are labeled with a unique isolate identifier and a cluster number if more than one cluster is present in genome. Nodes are colored according to class: Firmicutes, purple; Bacteriodetes, orange; Pproteobacteria, blue.

Lasso peptide cyclization is mediated by a RRE sequence found on the B or E protein (**Figure 5.1**). However, C-terminal phosphorylation of paeninodin proceeds in the absence of the leader peptide, albeit with decreased efficiency [157]. This suggests sequence motifs that guide post-translational modification could exist on either the leader or core peptide [178]. Peptide sequence motif analysis was carried out on all lasso peptides within clade 3 to identify potential shared or novel recognition sequences for the tailoring enzymes. Sequence motifs were found to be shared among peptides from different classes (**Figure 5.5A**). Specifically, a conserved N-terminal motif is shared with nearly every member of clade 3 (**Figure 5.5B**); this is likely the recognition sequence for the RRE [149]. Motifs 2 and 3 do not appear together on any peptide; the conserved Asp residues in the 5 position of motif 2 and the 7 position in motif 3 suggests there may be at least two different lasso ring sizes formed within this clade. Additionally, the acidic residues -2 and -4 from a conserved tyrosine in motif 2 resemble the site for tyrosine o-sulfation found in other organisms [179]. Motif 2 is present in 83% of the peptides with a sulfotransferase in their gene neighborhood, including PDC88. Given the possibility that PDC88 could be phosphorylated or sulfonated, and the precursor peptide contains a probable sulfation motif, a synthetic method based upon CFPS was developed to investigate the potential sulfation of a lasso peptide.

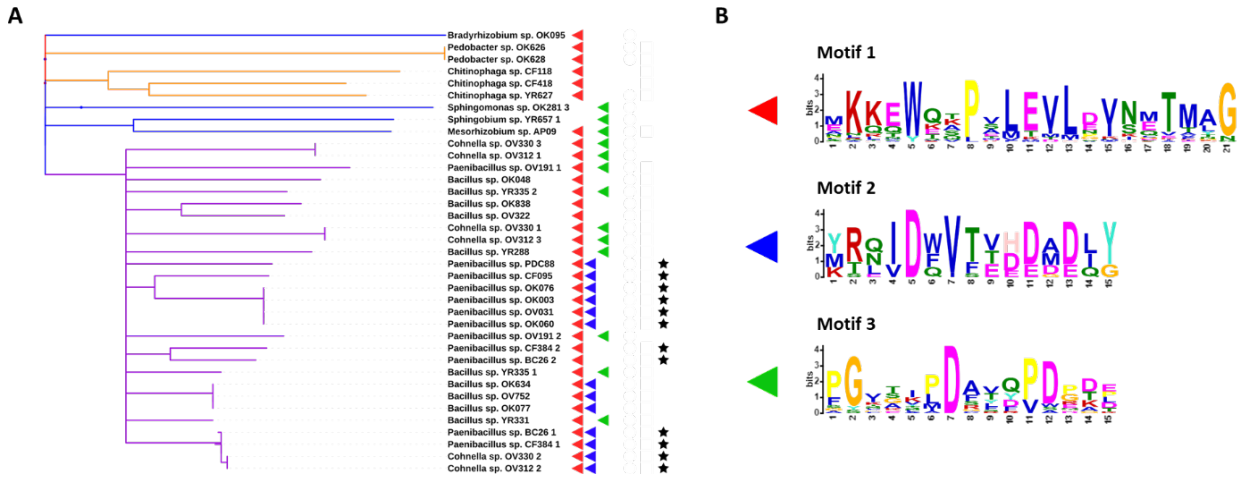


Figure 5.5: A. Neighbor joining tree of the lasso cyclases of clade 3. Sequence motifs identified in the precursor peptides found in the associated gene neighborhood are represented as three colored arrows. Clusters containing a sulfotransferase are marked with a star. B. Key with the LOGOs for the three lasso peptide precursor amino acid sequence motifs identified by MEME and shown in panel A.

5.4.4 Cell-free protein synthesis of functionalized lasso peptides.

Often RiPP biosynthetic enzymes and precursor peptides are expressed in heterologous hosts and purified for *in vitro* enzyme assays to determine their activities and end products. *in vitro* transcription and translation of proteins and peptides is an alternative that minimizes cloning and purification, is amenable to production of cytotoxic products, and allows modification of expression conditions [127]. While CFPS of functional enzymes is routine, small peptides are not often synthesized *in vitro*. Two well-described lasso peptides, microcin J25 and klebsidin [155], two cell-free systems, the *E. coli* crude S30 lysate [120] and the PURE system [180], were investigated for their ability to produce mature lasso peptides. In brief, the precursor peptide and the minimal maturation enzymes were expressed together in each cell-free system and the products were analyzed by LC-MS/MS with untargeted peptide sequencing.

Lasso peptide precursors go through several intermediates on their way to maturation, although the intermediates have not been observed from native hosts or in heterologous expression systems. Untargeted peptidomics enables monitoring of all modifications to the precursor prior to cyclization and is a high-throughput method to assess production of novel RiPPs. Neither full length precursor peptide was observed in either expression system, but smaller peptides were detected and sequenced for microcin J25 in both cell-free expression systems and klebsidin in the PURE system (**Figure 5.6A**). Peptide coverage was highest in the PURE system; each amino acid of the klebsidin peptide detected, including masses for the full length klebsidin core peptide (**Supplemental Figure 5.10**). The S30 system performed less well; only amino acids N8-G39 were detected for microcin J25 and no klebsidin peptides were detected. Neither of the mature lasso peptides were observed using either expression system.

Prompted by the low precursor peptide production in both systems, a fully synthetic precursor peptide was used instead of one produced *in situ* to investigate the tailoring enzymes of PDC88. Peptide coverage was much improved with a synthetic peptide (**Figure 5.6B**) however, reaction clean up using size exclusion filters reduced peptide coverage and the length of peptides detected. The molecular weight cutoff of size exclusion filters (10 kDa) assumed a globular protein, but linear peptides, even those much smaller than the cutoff, can

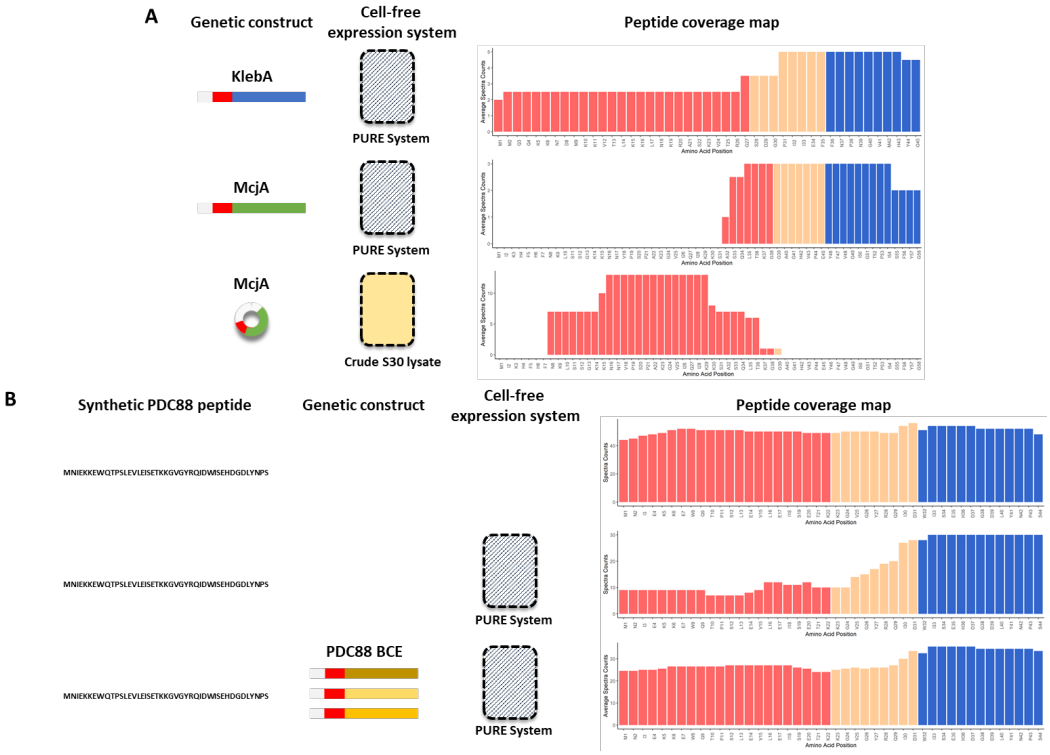


Figure 5.6: Schematic of CFPS of lasso peptides and average spectral counts for each amino acid position as determined by untargeted LC-MS/MS. Leader peptides are colored red, predicted macrolactam ring is colored orange and the peptide tail region is colored blue. A. Expression of control precursor peptides (n≥2) klebsidin (KlebA) in PURE system (top panel) and microcin J25 (McjA) in the PURE system (middle panel) and the crude S30 lysate (bottom panel). No klebsidin peptides were produced by CFPS in crude S30 lysate. Linear DNA templates were used for PURE system CFPS and plasmid DNA templates were used for crude S30 lysate CFPS. B. Changes in synthetic PDC88 precursor peptide as a result of incubation with PURE system without (middle panel) and with (bottom panel) DNA template for maturation enzymes. Linear DNA templates for PDC88 B, C, and E proteins were used for maturation study (n≥2) in PURE system.

be retained by the filter. Use of reverse-phase C18 columns led to increased peptide coverage and observed peptide sequence length, although the column functionalization can irreversibly bind some mature lasso peptides (unpublished data). More importantly, modifications that occur prior to cyclization are detectable with this method.

Incubation of synthetic PDC88 precursor peptide with the PURE expression system and without DNA template for the maturation enzymes resulted in decreased detection of peptide, particularly at the N-terminus (**Figure 5.6B**). Reduction in total peptide yield is likely decreased due to background proteins in the PURE system interfering with C18 column binding or ion suppression [181]. However, the decreased coverage of the N-terminus suggests that peptide degradation may be a factor even in a minimal expression system with no known proteases. Spontaneous cleavage and peptide hydrolysis has been observed in relatively mild conditions [182]. The increased prevalence of the lasso peptide tail (W32-S44) suggests that the tail region may be resistant to proteolysis even prior to modification. However, this complicates assessment of leader peptide cleavage when post-translational enzymes are present, as no additional enrichment of the lasso core region or total peptide coverage was observed when the DNA template for maturation enzymes was added to the reaction.

Computational tools exist that compare MS/MS spectra to characterized compounds, such as Dereplicator+, but application of this tool did not result in identification of lasso peptides in any experiment in this study [183]. Current untargeted MS/MS methods struggle to identity cyclic peptides due to their generally non-predictable fragmentation patterns [184]. However, targeted MS identified masses (mass error $\leq$ 5 ppm) corresponding to the core and cyclic PDC88 lasso peptides when synthetic precursor was incubated with all three maturation enzymes (**Supplemental Figure 5.11, Figure 5.7A**). XICs were prepared and the resulting MS^2 fragments were sequenced with PEAKSX [185]. The sequence of the linear peptide was confirmed (**Supplemental Table 5.4**); however, the ion corresponding to the cyclized lasso peptide mass was a false positive. Peptide sequencing revealed a dehydration to Asp39 instead of ring formation (**Table 5.3, Figure 5.7B**). It is possible the observed dehydration is the result of a reversible ring formation with Asp39. However, a 17 member macrolactam ring is significantly larger than any previously observed lasso peptide ring;

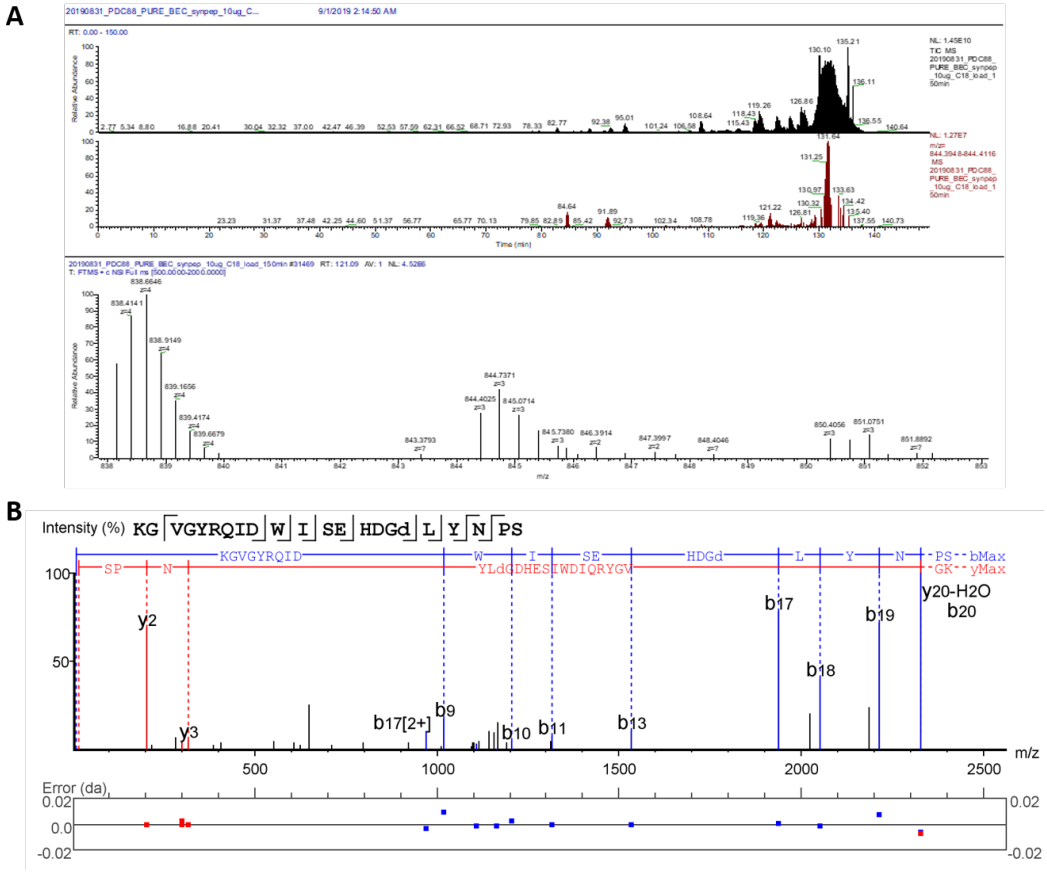


Figure 5.7: Analysis of potential cyclized PDC88 synthetic peptide when incubated with maturation enzymes. A. MS spectra of PDC88 synthetic peptide; top panel, total ion chromatogram of PDC88 synthetic peptide incubated with maturation enzymes; middle panel, extracted ion chromatogram of dehydrated PDC88 core ($[M+3H-18.0152]^{3+}$, 844.4032 Da, 10 ppm window); lower panel, full MS scan at peptide elution (retention time = 121.09 min). B. MS^2 spectrum of 844.4025 Da parent ion annotated with PEAKSX. Error map displayed under each annotated b and y ion.

Table 5.3: Identified b and y ions identified in MS/MS spectra using PEAKSX for synthetic peptide PDC88 parent mass (844.4025 Da).

b ion #	b	b (2+)	Amino acid seq	y	y-H20	y-NH3	y ion #
1	-	-	K	-	-	-	22
2	-	-	G	-	-	-	21
3	-	-	V	-	2328.08	-	20
4	-	-	G	-	-	-	19
5	-	-	Y	-	-	-	18
6	-	-	R	-	-	-	17
7	-	-	Q	-	-	-	16
8	-	-	I	-	-	-	15
9	1017.54	-	D	-	-	-	14
10	1203.62	-	W	-	-	-	13
11	1316.71	-	I	-	-	-	12
12	-	-	S	-	-	-	11
13	1532.79	-	E	-	-	-	10
14	-	-	H	-	-	-	9
15	-	-	D	-	-	-	8
16	-	-	G	-	-	-	7
17	1938.91	969.96	D(-18.01)	-	-	-	6
18	2052.00	-	L	-	-	-	5
19	2215.05	1108.03	Y	-	-	-	4
20	2329.11	1165.05	N	317.15	299.14	300.11	3
21	-	-	P	203.10	-	-	2
22	-	-	S	-	-	-	1

the average lasso peptide ring is seven to nine residues [147]. Incubation of the synthetic precursor peptide with the PURE system and DNA template for the sulfotransferase did not result in identification of a sulfation. Inclusion of the full set of maturation enzymes in the reaction also did not yield a sulfation. Furthermore, sulfation of the C-terminal Ser44 or Tyr41 was not detected in targeted methods, suggesting that sulfation does not occur on the unmodified C-terminus of PDC88.

Untargeted peptide sequencing enabled measurement of peptides synthesized by CFPS despite suboptimal production levels and interfering degradation processes. Targeted MS methods alone failed to identify any peptides produced by CFPS due to unanticipated truncations of the precursor. Additionally, differences in peptide coverage observed by this method revealed that implementation of purification techniques will be critical to characterization of lasso peptides. Nevertheless, a fully mature lasso peptide with or without sulfation was not detected by any method. This indicates that further development of cell-free production of small peptides in combination with *in vitro* lasso peptide cyclization methodologies and tandem MS/MS of peptide NPs is required for exploration of lasso peptide diversity and eventually functionality.

5.5 Conclusion

Newly developed genome mining tools have rapidly increased the number of gene sequences of predicted NPs, particularly for previously elusive classes with unusual BGC architectures like RiPPs. Although the lasso peptide cluster appears to be prevalent in the Populus root microbiome, with over eighty clusters from 407 organisms and spread across three bacterial clades, the function of lasso peptides in plant-microbe interactions remains elusive.

Analysis of lasso peptide BGCs revealed 12 clusters containing an enzyme with predicted sulfotransferase activity previously unknown to be associated with neither predicted nor characterized lasso peptides. Through constitutive expression screening, sequence similarity analysis, and amino acid motif identification, this enzyme was predicted to perform a novel sulfation on the C-terminus of 10 lasso peptides in this collection. To investigate this function and explore the chemical and structural diversity of lasso peptides, a synthetic methodology

combining CFPS and untargeted peptide sequencing by LC MS/MS was developed. Though a novel sulfation was not identified with this method, its development identified critical research areas for advancing high-throughput *in vitro* functional genomics.

Enzymatic generation of small molecule diversity will be greatly served by the development of a high-throughput method for screening the many unknown tailoring enzymes in the BGCs of lasso peptides and other RiPPs. The combination of CFPS of tailoring enzymes and precursor peptides with untargeted peptide sequencing enables rapid determination of peptide modifications and reaction order. Challenges remain in robust production of small peptides, particularly in crude extracts. Peptide production was adequate in the PURE system but was still limited by non-specific degradation. However, recent progress in the development of crude CFPS expression systems from *Vibrio natriegens* suggest high yield *in vitro* peptide synthesis is possible [186]. Untargeted peptide sequencing will be a necessary analytic tool for engineering improved peptide synthesis using CFPS.

Database-independent MS methods for discovery of post-translational modification has improved alongside mass accuracy and computational power, but cyclic peptide structures remain difficult to decipher due to the multiple possible peptide sequences that result from fragmentation. Most current discovery methods rely on spectra matching to databases of curated peptides, which limits rediscovery [187], but is less robust towards finding novel peptidic NPs whose post-translational modifications and final structures are not known. However, as more cyclic peptides are characterized, spectral libraries will become more broadly useful and untargeted approaches will more easily deconvolute new data [188, 184].

Existing heterologous expression workflows cannot keep up with the boom in high-quality genome sequence data. Predictive and computation tools can provide partial information about many ORFs but are not a substitute for empirical data. High-throughput cell-free expression studies are becoming a promising tool for functional genomics efforts and NP discovery, but often need to be engineered for specific products [117]. Incorporation of mass spectroscopy-based omics techniques into the engineering process for cell-free systems will enable high-yield and high-throughput production of complex small molecule and peptide NPs.

5.6 Appendix

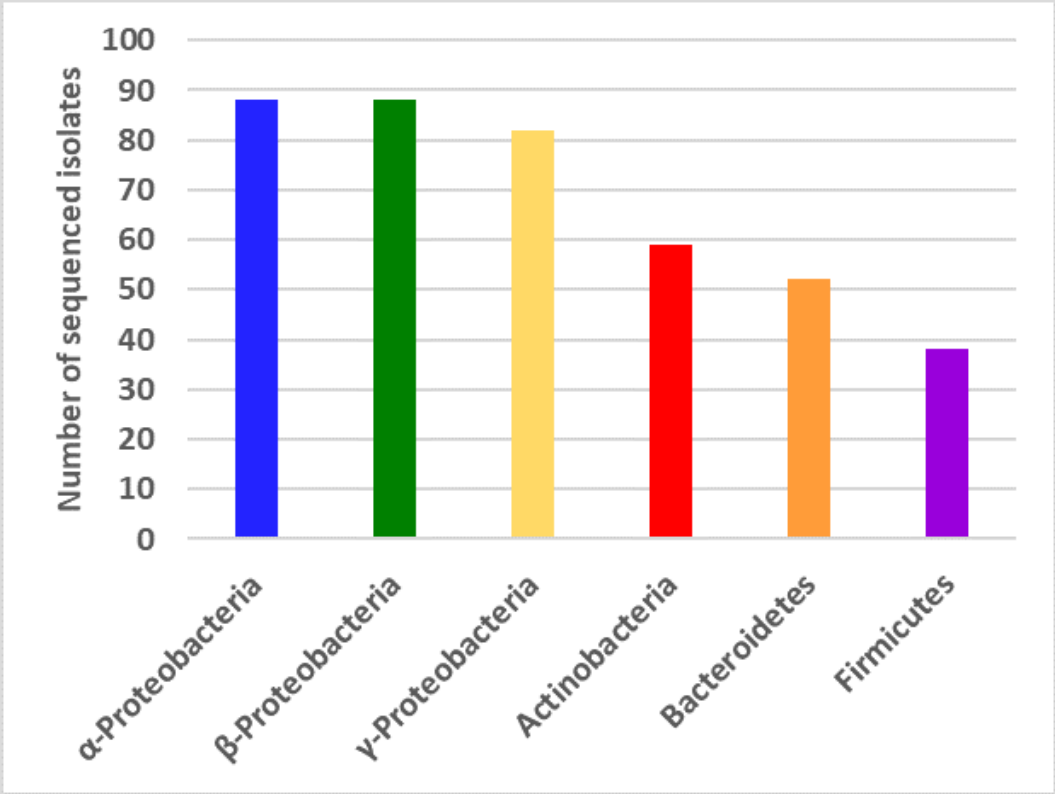


Figure 5.8: Phyla of 407 sequenced isolates. Isolate identities are described in Supplemental Table C-1.

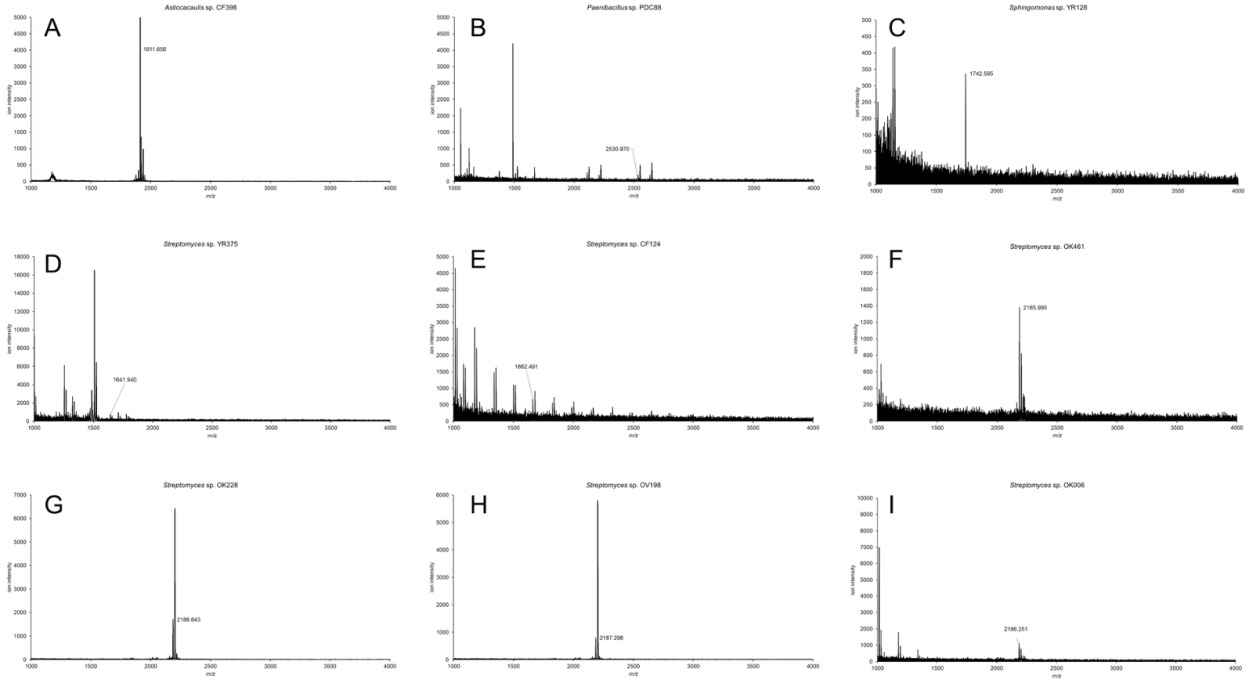


Figure 5.9: MALDI-TOF mass spectra of each potential product identified from an isolate. Masses of interest are labeled. A. PDC88 predicted mass 2531.2 Da, observed mass 2539.970 Da. B. YR128 predicted mass 1742.9 Da; observed mass 1742.595 Da. C. YR375 predicted mass 1619. Da7; observed mass 1641.940 Da. D. CF124 predicted mass 1639.9 Da, observed mass 1662.491 Da. E. OK461 predicted mass 2166.9 D, observed mass 2185.990 Da. F. OK228 predicted mass 2165.9 Da, observed mass 2186.643 Da. G. OV198 predicted mass 2166.9 Da, observed mass 2187.296 Da. H. OK006 predicted mass 2166.9 Da, observed mass 2186.251 Da.

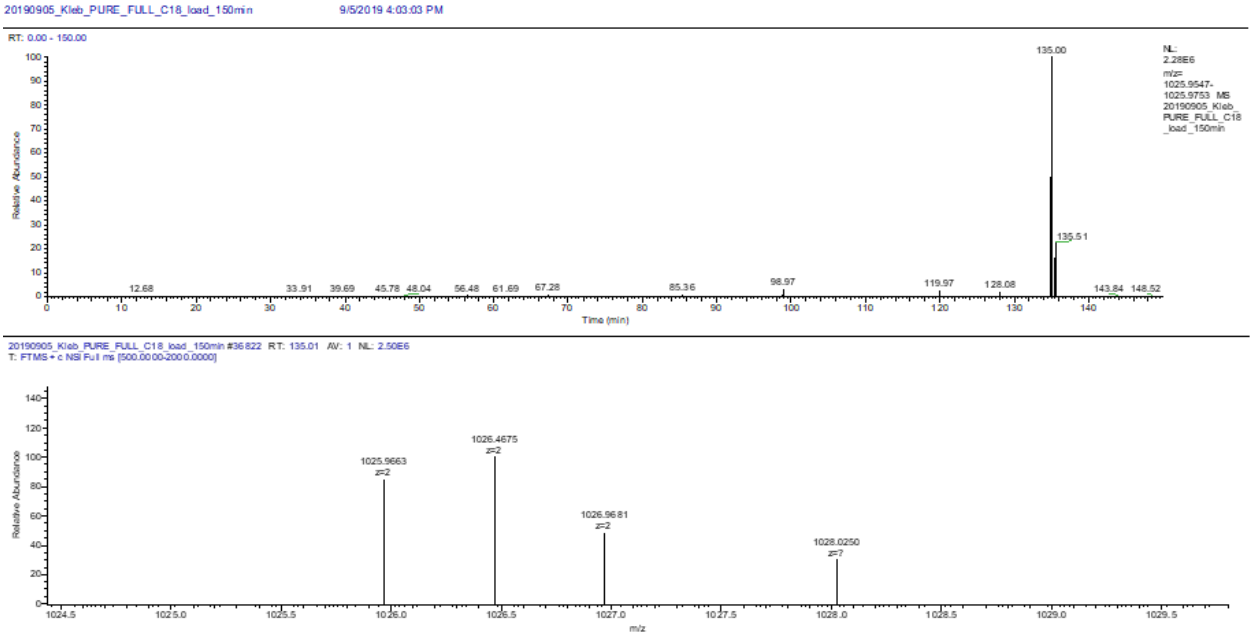


Figure 5.10: MS spectra for klebsidin. Top panel, extracted ion chromatogram of klebsidin core ($[M+2H]^{2+}$, m/z = 1025.9650, 10 ppm window); lower panel, full MS scan at peptide elution (retention time = 118.61 min).

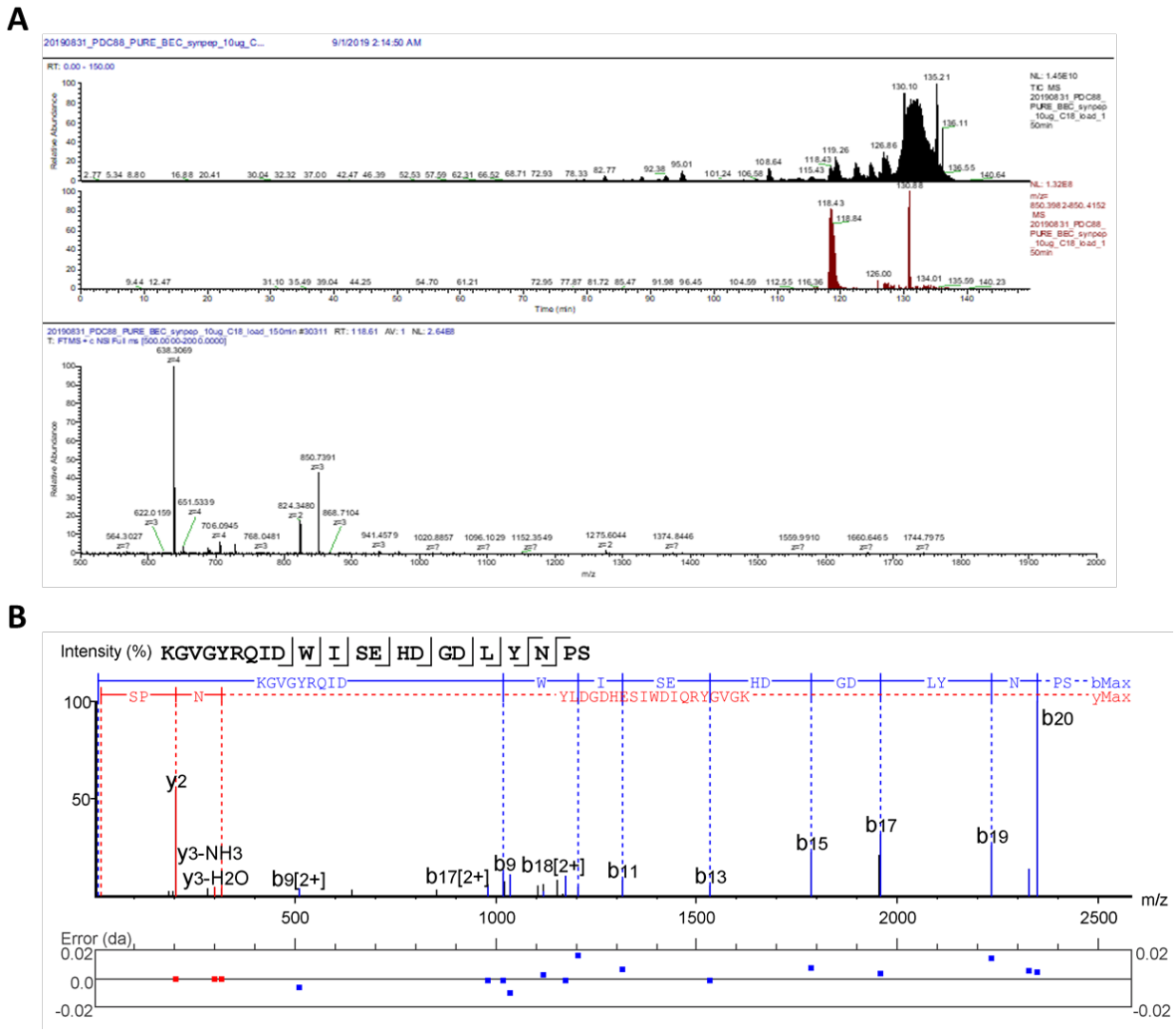


Figure 5.11: Analysis linear core PDC88 synthetic peptide when incubated with maturation enzymes. A. Total ion chromatogram (top panel), extracted ion chromatogram (middle panel), and full MS scan at peptide elution time (rt = 118.61 min) (lower panel) of the PDC88 synthetic peptide incubated with maturation enzymes ($[M+3H]^{3+}$, m/z = 850.4061, 10 ppm window). B. MS^2 spectrum of the 850.4055 Da parent ion annotated with PEAKSX. Error map displayed under each annotated b and y ion.

Table 5.4: Identified b and y ions for PDC88 synthetic peptide parent mass (850.4055 Da), identified in MS/MS spectra using PEAKSX.

b ion #	b	b-H20	b (2+)	Amino acid seq	y	y-H20	y-NH3	y ion #
1	-	-	-	K	-	-	-	22
2	-	-	-	G	-	-	-	21
3	-	-	-	V	-	-	-	20
4	-	-	-	G	-	-	-	19
5	-	-	-	Y	-	-	-	18
6	-	-	-	R	-	-	-	17
7	-	-	-	Q	-	-	-	16
8	-	-	-	I	-	-	-	15
9	1017.55	-	509.28	D	-	-	-	14
10	1203.61	-	-	W	-	-	-	13
11	1316.7	-	-	I	-	-	-	12
12	-	-	-	S	-	-	-	11
13	1532.79	-	-	E	-	-	-	10
14	-	-	-	H	-	-	-	9
15	1784.86	-	-	D	-	-	-	8
16	-	-	-	G	-	-	-	7
17	1956.92		978.96	D	-	-	-	6
18	-	-	1035.51	L	-	-	-	5
19	2233.05	-	1117.03	Y	-	-	-	4
20	2347.11	2329.09	1174.06	N	317.15	299.14	300.12	3
21	-	-	-	P	203.1	-	-	2
22	-	-	-	S	-	-	-	1

Chapter 6

Conclusions

Genome mining and functional prediction are powerful tools for the identification of novel NP pathways but must be paired with empirical validation. Engineering heterologous expression of NP gene clusters is low-throughput and production can be limited by cytotoxicity and the metabolic demands of growth-related processes. Alternatively, cell-free systems are a high-throughput method for the expression and characterization of peptides, enzymes and metabolic pathways.

Both the engineering and application of cell-free systems are aided by global analytic measures. Particularly, the utility of proteomics as an analytic method for cell-free protein synthesis in crude lysates cannot be overstated, as it can provide assessment of both *in vitro* expression levels and quantitative measures of the abundance of transcriptional and translational machinery [68]. In this thesis, the benefits of untargeted peptidomics were apparent across three cell-free systems, demonstrated by its value in evaluating enzyme activity and observing systemic impacts of growth media composition on lysate metabolic potential.

The number of newly discovered gene clusters expected to produce peptide NPs greatly outnumbers the described members of this class. Cell-free protein and peptide synthesis is particularly well-suited to the characterization of new RiPP BGCs. Through the investigation of a novel putative sulfotransferase, untargeted peptidomics was critical in measuring not only full-length peptide products, but also intermediates, partially translated peptides, and degraded products. The degradation of peptides, even in a purified system with

no known peptidases, was observed through these methods. The use of this tool is paramount in both the discovery of new peptide NPs and the improvement of *in vitro* peptide synthesis. Systematic engineering of cell-free systems will broaden the organismal diversity from which lysates are derived and will provide access to new functions *in vitro*. Already, early work on cell-free protein synthesis in Streptomyces and Vibrio species show increased yields of high G-C content proteins and short peptides, respectively [189, 190, 186].

Demonstrating its utility beyond measurement of individual peptide abundance, detection of global protein abundance by assembled peptidomics revealed the presence of protein families in the S30 lysate that were expected to be absent after cell lysis. Particularly, the identification of components of the phosphotransferase system (PTS) within crude *E. coli* cell extracts raises questions about the fate of sugar substrates and the depletion of phosphenolpyruvate as a result of the activity of PTS. Identification of these functions serves not only exploration of the metabolic malleability of organisms, but also facilitates the engineering of cell-free systems.

Illumination of the black box of crude cell-free systems with proteomics enables their engineering. The impacts of modifications to media and preparation conditions can have large impacts on protein content; however, guided by proteomics, these modifications can be leveraged to improve cell-free biosynthetic processes. To improve *in vitro* tyrosine biosynthesis, the effects of four changes to growth media for a crude S30 *E. coli* lysate were assessed with shotgun proteomics. Combining the modifications which upregulated proteins of interest more than doubled yield. Expression of a feedback insensitive variant of a rate-limiting enzyme by CFPS further grew yield, demonstrating the strengths of a combined cell-free protein synthesis and metabolic engineering platform.

Removed from the cellular milieu, biochemical processes are both simplified and more versatile in their applications than their cellular counterparts. However, wide-spread adoption of crude cell-free systems for metabolic engineering and functional genomics is hampered by the inherent complexity of a system derived from living cells. In addition to the large biological variability between and within populations of cells, numerous variables in extract preparation including lysis efficiency and cellular life stage make standardization of crude cell-free systems challenging. Post-lysis benchmarking and adjustment of the

components of cell-free systems may be a more tractable method to increase reproducibility. Identification of critical metabolites or proteins that are predictive of *in vitro* activity will be necessary to reduce the variability between cell-free systems. Furthermore, the post-translational modification of enzymes within crude lysates and the lability of these modifications *in vitro* remains an unexplored source of variability in cell-free systems that may significantly influence metabolic activity and the efficiency of protein synthesis [191, 192]. Controlling this complexity through engineering guided by global analytic tools will increase the reproducibility and the reliability of cell-free systems for both protein synthesis and metabolite production.

Reliability and reproducibility will be central to future development of CFPS platforms, but the current performance of these systems has been sufficient to allow wide-spread adoption of CFPS as a tool for investigating enzyme function and diversity. However, the use of CFPS to produce consumable biomolecules has been limited to pilot studies on the micro-scale, due to a combination of inadequate protein yield and the need for expensive substrates. Implementation of promising genetic and computational tools will aid improvement of *in vitro* protein yield through post-lysis modification of lysates and enable the engineering of phenotypes derived from interactions of tens of enzymes [47, 140, 139, 130]. The cost of substrates, including amino acids, tRNAs, nucleosides, and DNA templates currently limit the scalabiltiy of CFPS. Reducing reliance on exogenous substrates by capturing the metabolites present during lysis and activation of endogenous salvage and recycling pathways could greatly reduce the cost of *in vitro* protein synthesis. Additionally, further development of CFPS systems capable of using a linear DNA template are of great interest as PCR is a scalable and inexpensive source of template DNA. A high-yield and low-cost CFPS system would enable the on-demand production of customizable antibodies, enzyme therapeutics and peptide drugs.

In contrast to CFPS systems, crude lysates and other multi-enzyme reaction cascades which have undergone varying degrees of purification are common tools for carrying out large-scale industrial processes in the fine and commodity chemical industries [193, 194, 195]. Often individual enzymes are rationally-engineered or evolved to improve productivity. As the length of metabolic pathways of interest has increased, CFPS has been introduced as a

tool in these workflows to accelerate design-build-test cycles. Combined CFME-PS systems are primarily a prototyping platform for eventual *in vivo* bioproduction, due to substrate costs; however, use-cases exist particularly in scenarios where biocontainment and production of numerous diverse compounds are of value. Unlike living cells, genomic DNA is unnecessary for cell-free processes, which eases the requirements for their release into the environment. Building from the well-described use of CFPS for sensing of environmental contaminants [196, 197, 198], a cell-free system could be employed to detect and degrade pollutants including microplastics and pharmaceuticals in water supplies, with limited concerns about release of genetically-modified organisms. Additionally, the flexibility granted by CFPS would allow for the development of a lyophilized bioreactor, which could, with the addition of dried DNA and water, process low-cost sugar substrates to common intermediates that are then converted to active valuable compounds by enzymes expressed by CFPS.

The value of cell-free systems is predicated on the ability to isolate, investigate and control complex biochemistries in an environment not limited by organismal constraints. Free from membranes, cell-free processes like protein synthesis are both efficient and versatile, enabling rapid synthesis of enzymes from DNA templates. Further, removal of endogenous genetic material lessens dynamic shifts in protein content and metabolic activity. The combination of these factors makes cell-free systems a great enabling tool for functional genomics and metabolic engineering both *in vitro* and *in vivo*. Continued development and integration of cell-free systems and enabling tools will empower biologists to rapidly explore new BGCs and engineer novel NP pathways to improve worldwide health and sustainability.

Bibliography

[1] David J. Newman and Gordon M. Cragg. Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products*, 79(3):629–661, March 2016. 1

[2] Nadine Ziemert, Mohammad Alanjary, and Tilmann Weber. The evolution of genome mining in microbes – a review. *Natural Product Reports*, 33(8):988–1005, 2016. 1, 64

[3] Monica I. Abrudan, Fokko Smakman, Ard Jan Grimbergen, Sanne Westhoff, Eric L. Miller, Gilles P. van Wezel, and Daniel E. Rozen. Socially mediated induction and suppression of antibiosis during bacterial coexistence. *Proceedings of the National Academy of Sciences*, 112(35):11054–11059, September 2015. 1

[4] Tamás L. Czárán, Rolf F. Hoekstra, and Ludo Pagie. Chemical warfare between microbes promotes biodiversity. *Proceedings of the National Academy of Sciences*, 99(2):786–790, January 2002. 1

[5] Jane M. Moore, Elizabeth Bradshaw, Ryan F. Seipke, Matthew I. Hutchings, and Michael McArthur. Use and Discovery of Chemical Elicitors That Stimulate Biosynthetic Gene Clusters in Streptomyces Bacteria. *Methods in Enzymology*, 517:367–385, January 2012. 1

[6] Robin K. Pettit. Small-molecule elicitation of microbial secondary metabolites. *Microbial Biotechnology*, 4(4):471–478, July 2011. 1

[7] Mohammad R. Seyedsayamdost. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proceedings of the National Academy of Sciences*, 111(20):7266–7271, May 2014. 1

[8] Martin Tulp and Lars Bohlin. Rediscovery of known natural compounds: Nuisance or goldmine? *Bioorganic & Medicinal Chemistry*, 13(17):5274–5282, September 2005. 1

[9] Cameron R. Pye, Matthew J. Bertin, R. Scott Lokey, William H. Gerwick, and Roger G. Linington. Retrospective analysis of natural products provides insights for future discovery trends. *Proceedings of the National Academy of Sciences*, 114(22):5601–5606, May 2017. 1

[10] Alex Bateman, Ewan Birney, Lorenzo Cerruti, Richard Durbin, Laurence Etwiller, Sean R. Eddy, Sam Griffiths-Jones, Kevin L. Howe, Mhairi Marshall, and Erik L. L. Sonnhammer. The Pfam Protein Families Database. *Nucleic Acids Research*, 30(1):276–280, January 2002. 2

[11] Robert D. Finn, Jody Clements, and Sean R. Eddy. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Research*, 39(Web Server issue):W29–W37, July 2011. 2

[12] Randall A. Hughes and Andrew D. Ellington. Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. *Cold Spring Harbor Perspectives in Biology*, 9(1):a023812, January 2017. 2

[13] Shlomi Dvir, Lars Velten, Eilon Sharon, Danny Zeevi, Lucas B. Carey, Adina Weinberger, and Eran Segal. Deciphering the rules by which 5-UTR sequences affect protein expression in yeast. *Proceedings of the National Academy of Sciences*, 110(30):E2792–E2801, July 2013. 2

[14] Daniel G Gibson, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature methods*, 6(5):343–5, May 2009. 2

[15] Stephan Kirchmaier, Katharina Lust, and Joachim Wittbrodt. Golden GATEway cloning–a combinatorial approach to generate fusion and recombination constructs. *PloS one*, 8(10):e76117, January 2013. 2

[16] 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. 2

[17] Tessa E. F. Quax, Nico J. Claassens, Dieter Söll, and John van der Oost. Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular Cell*, 59(2):149–161, July 2015. 2

[18] Mamoru Komatsu, Takuma Uchiyama, Satoshi Ōmura, David E. Cane, and Haruo Ikeda. Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. *Proceedings of the National Academy of Sciences*, 107(6):2646–2651, February 2010. 2

[19] David Lubertozzi and Jay D. Keasling. Developing Aspergillus as a host for heterologous expression. *Biotechnology Advances*, 27(1):53–75, January 2009. 2

[20] Kozo Ochi and Takeshi Hosaka. New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Applied Microbiology and Biotechnology*, 97(1):87–98, January 2013. 2

[21] Bin Wang, Fang Guo, Shi-Hui Dong, and Huimin Zhao. Activation of silent biosynthetic gene clusters using transcription factor decoys. *Nature Chemical Biology*, 15(2):111–114, February 2019. 2

[22] Mingzi M. Zhang, Fong Tian Wong, Yajie Wang, Shangwen Luo, Yee Hwee Lim, Elena Heng, Wan Lin Yeo, Ryan E. Cobb, Behnam Enghiad, Ee Lui Ang, and Huimin Zhao. CRISPR-Cas9 strategy for activation of silent Streptomyces biosynthetic gene clusters. *Nature Chemical Biology*, 13(6):607–609, June 2017. 2

[23] P. Gupta and S.C. Phulara. Metabolic engineering for isoprenoid-based biofuel production. *Journal of Applied Microbiology*, 119(3):605–619, September 2015. 2

[24] Vincent J. J. Martin, Douglas J. Pitera, Sydnor T. Withers, Jack D. Newman, and Jay D. Keasling. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature Biotechnology*, 21(7):796–802, July 2003. 2

[25] Gionata Scalcinati, Siavash Partow, Verena Siewers, Michel Schalk, Laurent Daviet, and Jens Nielsen. Combined metabolic engineering of precursor and co-factor supply to increase -santalene production by Saccharomyces cerevisiae. *Microbial Cell Factories*, 11(1):117, August 2012. 2

[26] Naama Tepper and Tomer Shlomi. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics*, 26(4):536–543, February 2010. 2

[27] Mary J Dunlop, Zain Y Dossani, Heather L Szmidt, Hou Cheng Chu, Taek Soon Lee, Jay D Keasling, Masood Z Hadi, and Aindrila Mukhopadhyay. Engineering microbial biofuel tolerance and export using efflux pumps. *Molecular Systems Biology*, 7(1):487, January 2011. 2

[28] Ranjan Patnaik, Susan Louie, Vesna Gavrilovic, Kim Perry, Willem P. C. Stemmer, Chris M. Ryan, and Stephen del Cardayré. Genome shuffling of Lactobacillus for improved acid tolerance. *Nature Biotechnology*, 20(7):707–712, July 2002. 2

[29] Uwe Sauer. Evolutionary Engineering of Industrially Important Microbial Phenotypes. In Jens Nielsen, L. Eggeling, J. Dynesen, M. Gárdonyi, R. T. Gill, A. A. de Graaf, B. Hahn-Hägerdal, L. J. Jönsson, C. Khosla, R. Licari, R. McDaniel, M. McIntyre, C. Miiller, J. Nielsen, R. R. Cordero Otero, H. Sahm, U. Sauer, D. E. Stafford, G. Stephanopoulos, C. E. Wahlbom, K. S. Yanagimachi, and W. H. van Zyl, editors, *Metabolic Engineering*, Advances in Biochemical Engineering/Biotechnology, pages 129–169. Springer Berlin Heidelberg, Berlin, Heidelberg, 2001. 2

[30] Akira Nakagawa, Hiromichi Minami, Ju-Sung Kim, Takashi Koyanagi, Takane Katayama, Fumihiko Sato, and Hidehiko Kumagai. A bacterial platform for fermentative production of plant alkaloids. *Nature Communications*, 2:326, May 2011. 2

[31] Dae-Kyun Ro, Eric M Paradise, Mario Ouellet, Karl J Fisher, Karyn L Newman, John M Ndungu, Kimberly A Ho, Rachel A Eachus, Timothy S Ham, James Kirby, Michelle C Y Chang, Sydnor T Withers, Yoichiro Shiba, Richmond Sarpong, and Jay D Keasling. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440:940–943, 2006. 2

[32] Stephen Sarria, Betty Wong, Hector García Martín, Jay D Keasling, and Pamela Peralta-Yahya. Microbial synthesis of pinene. *ACS synthetic biology*, 3(7):466–75, July 2014. 2

[33] Gang Wu, Qiang Yan, J. Andrew Jones, Yinjie J. Tang, Stephen S. Fong, and Mattheos A. G. Koffas. Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends in Biotechnology*, 34(8):652–664, August 2016. 3

[34] Richard H Baltz. Streptomyces and Saccharopolyspora hosts for heterologous expression of secondary metabolite gene clusters. *Journal of industrial microbiology & biotechnology*, 37(8):759–72, August 2010. 3

[35] P. F. Fox and P. L. H. McSweeney. Rennets: their role in milk coagulation and cheese ripening. In B. A. Law, editor, *Microbiology and Biochemistry of Cheese and Fermented Milk*, pages 1–49. Springer US, Boston, MA, 1997. 3

[36] Harold McGee. *On Food and Cooking: The Science and Lore of the Kitchen*. Simon and Schuster, March 2007. Google-Books-ID: bKVCtH4AjwgC. 3

[37] Shaorong Chong. Overview of Cell-Free Protein Synthesis: Historic Landmarks, Commercial Systems, and Expanding Applications: Overview of Cell-Free Protein Synthesis. In Frederick M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J.G. Seidman, John A. Smith, and Kevin Struhl, editors, *Current Protocols in Molecular Biology*, pages 16.30.1–16.30.11. John Wiley & Sons, Inc., Hoboken, NJ, USA, October 2014. 3, 36

[38] Jerard Hurwitz. The Discovery of RNA Polymerase. *Journal of Biological Chemistry*, 280(52):42477–42485, December 2005. 3, 36

[39] Bryan A. Wilson, Jonathan C. Schisler, and Monte S. Willis. Sir Hans Adolf Krebs: Architect of Metabolic Cycles. *Laboratory Medicine*, 41(6):377–380, June 2010. 3, 36

[40] Alexander S. Spirin, Vladimir I. Baranov, Lubov A. Ryabova, Sergey Yu Ovodov, and Yuly B. Alakhov. A continuous cell-free translation system capable of producing polypeptides in high yield. *Science*, 242(4882):1162, 1988. 3

[41] H. Z. Chen and G. Zubay. Prokaryotic coupled transcription-translation. *Methods in Enzymology*, 101:674–690, 1983. 3

[42] Adam D Silverman, Nancy Kelley-Loughnane, Julius B Lucks, and Michael C Jewett. Deconstructing cell-free extract preparation for in vitro activation of transcriptional genetic circuitry. *bioRxiv*, September 2018. 4

[43] Melissa K. Takahashi, Clarmyra A. Hayes, James Chappell, Zachary Z. Sun, Richard M. Murray, Vincent Noireaux, and Julius B. Lucks. Characterizing and prototyping genetic networks with cell-free transcription-translation reactions. *Methods (San Diego, Calif.)*, 86:60–72, September 2015. 4

[44] Yoshihiro Shimizu, Akio Inoue, Yukihide Tomari, Tsutomu Suzuki, Takashi Yokogawa, Kazuya Nishikawa, and Takuya Ueda. Cell-free translation reconstituted with purified components. *Nature Biotechnology*, 19(8):751–755, August 2001. 4, 12, 17, 46

[45] Barbora Lavickova and Sebastian J. Maerkl. A Simple, Robust, and Low-Cost Method To Produce the PURE Cell-Free System. *ACS Synthetic Biology*, 8(2):455–462, February 2019. 4

[46] Fernando Villarreal, Luis E. Contreras-Llano, Michael Chavez, Yunfeng Ding, Jinzhen Fan, Tingrui Pan, and Cheemeng Tan. Synthetic microbial consortia enable rapid assembly of pure translation machinery. *Nature Chemical Biology*, 14(1):29–35, January 2018. 4

[47] Harris H. Wang, Po-Yi Huang, George Xu, Wilhelm Haas, Adam Marblestone, Jun Li, Steven P. Gygi, Anthony C. Forster, Michael C. Jewett, and George M. Church. Multiplexed in Vivo His-Tagging of Enzyme Pathways for in Vitro Single-Pot Multienzyme Catalysis. *ACS Synthetic Biology*, 1(2):43–52, February 2012. 4, 13, 17, 92

[48] Matthias Harbers. Wheat germ systems for cell-free protein expression. *FEBS Letters*, 588(17):2762–2773, August 2014. 4

[49] Matthias Buntru, Simon Vogel, Katrin Stoff, Holger Spiegel, and Stefan Schillberg. A versatile coupled cell-free transcription-translation system based on tobacco BY-2 cell lysates. *Biotechnology and Bioengineering*, 112(5):867–878, May 2015. 4

[50] Andreas K. Brödel, Andrei Sonnabend, and Stefan Kubick. Cell-free protein expression based on extracts from CHO cells. *Biotechnology and Bioengineering*, 111(1):25–36, January 2014. 4

[51] Toru Ezure, Takashi Suzuki, Shoken Higashide, Eiichi Shintani, Kohki Endo, Shinichiro Kobayashi, Masamitsu Shikata, Masaaki Ito, Koji Tanimizu, and Osamu Nishimura. Cell-free protein synthesis system prepared from insect cells by freeze-thawing. *Biotechnology Progress*, 22(6):1570–1577, December 2006. 4

[52] Takashi Suzuki, Masaaki Ito, Toru Ezure, Masamitsu Shikata, Eiji Ando, Toshihiko Utsumi, Susumu Tsunasawa, and Osamu Nishimura. Protein prenylation in an insect cell-free protein synthesis system and identification of products by mass spectrometry. *Proteomics*, 7(12):1942–50, June 2007. 4

[53] Anne Zemella, Lena Thoring, Christian Hoffmeister, and Stefan Kubick. Cell-Free Protein Synthesis: Pros and Cons of Prokaryotic and Eukaryotic Systems. *ChemBioChem*, 16(17):2420–2431, 2015. 4

[54] Filippo Caschera and Vincent Noireaux. Synthesis of 2.3 mg/ml of protein with an all Escherichia coli cell-free transcription–translation system. *Biochimie*, 99:162–168, April 2014. 4, 13, 36

[55] Dong-Myung Kim and James R. Swartz. Prolonging cell-free protein synthesis with a novel ATP regeneration system. *Biotechnology and Bioengineering*, 66(3):180–188, January 1999. 4

[56] Kalavathy Sitaraman, Dominic Esposito, George Klarmann, Stuart F Le Grice, James L Hartley, and Deb K Chatterjee. A novel cell-free protein synthesis system. *Journal of Biotechnology*, 110(3):257–263, June 2004. 4

[57] Dong-Myung Kim and James R. Swartz. Regeneration of adenosine triphosphate from glycolytic intermediates for cell-free protein synthesis. *Biotechnology and Bioengineering*, 74(4):309–316, August 2001. 4, 18, 31, 37

[58] Jennifer E. Kay and Michael C. Jewett. Lysate of engineered Escherichia coli supports high-level conversion of glucose to 2,3-butanediol. *Metabolic Engineering*, 32:133–142, November 2015. 4, 37

[59] Quentin M Dudley, Connor J Nash, and Michael C Jewett. Cell-free biosynthesis of limonene using enzyme-enriched *Escherichia coli* lysates. *Synthetic Biology*, 4(1), January 2019. 4

[60] Richard Kelwick, Luca Ricci, Soo Mei Chee, David Bell, Alexander J Webb, and Paul S Freemont. Cell-free prototyping strategies for enhancing the sustainable production of polyhydroxyalkanoates bioplastics. *Synthetic Biology*, 3(1), January 2018. 4

[61] Tyler P. Korman, Paul H. Opgenorth, and James U. Bowie. A synthetic biochemistry platform for cell free production of monoterpenes from glucose. *Nature Communications*, 8:15526, May 2017. 5, 18, 36

[62] Joseph A. Rollin, Julia Martin del Campo, Suwan Myung, Fangfang Sun, Chun You, Allison Bakovic, Roberto Castro, Sanjeev K. Chandrayan, Chang-Hao Wu, Michael W. W. Adams, Ryan S. Senger, and Y.-H. Percival Zhang. High-yield hydrogen production from biomass by in vitro metabolic engineering: Mixed sugars coutilization and kinetic modeling. *Proceedings of the National Academy of Sciences*, 112(16):4964–4969, April 2015. 5, 14

[63] Paul H. Opgenorth, Tyler P. Korman, and James U. Bowie. A synthetic biochemistry molecular purge valve module that maintains redox balance. *Nature Communications*, 5, June 2014. 5

[64] Joseph A. Rollin, Yannick J. Bomble, Peter C. St. John, and Addison K. Stark. Biochemical Production with Purified Cell-Free Systems. *Biochemical Engineering Journal*, July 2018. 5, 36

[65] Abel C. Chiao, Richard M. Murray, and Zachary Z. Sun. Development of prokaryotic cell-free systems for synthetic biology. *bioRxiv*, page 048710, April 2016. 5

[66] Elizabeth Strychalski. NIST Cell-Free Workshop, June 2019. 5

[67] R.G. Kim and C.Y. Choi. Expression-independent consumption of substrates in cell-free expression system from Escherichia coli. *Journal of Biotechnology*, 84(1):27–32, November 2000. 5

[68] Yaoyang Zhang, Bryan R. Fonslow, Bing Shan, Moon-Chang Baek, and John R. Yates. Protein Analysis by Shotgun/Bottom-up Proteomics. *Chemical Reviews*, 113(4):2343–2394, April 2013. 5, 90

[69] David L. Tabb, Christopher G. Fernando, and Matthew C. Chambers. MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. *Journal of Proteome Research*, 6(2):654–661, February 2007. 6, 21, 41, 69

[70] Linfeng Wu and David K. Han. Overcoming the dynamic range problem in mass spectrometry-based shotgun proteomics. *Expert Review of Proteomics*, 3(6):611–619, December 2006. 6

[71] James F. Zawada, Gang Yin, Alexander R. Steiner, Junhao Yang, Alpana Naresh, Sushmita M. Roy, Daniel S. Gold, Henry G. Heinsohn, and Christopher J. Murray. Microscale to manufacturing scale-up of cell-free cytokine production—a new approach for shortening protein production development timelines. *Biotechnology and Bioengineering*, 108(7):1570–1578, 2011. 7, 36

[72] Saurabh (Rob) Aggarwal. What's fueling the biotech engine—2012 to 2013. *Nature Biotechnology*, 32(1):32–39, January 2014. 11

[73] David J. Craik, David P. Fairlie, Spiros Liras, and David Price. The Future of Peptide-based Drugs. *Chemical Biology & Drug Design*, 81(1):136–147, January 2013. 11

[74] Nabih A. Baeshen, Mohammed N. Baeshen, Abdullah Sheikh, Roop S. Bora, Mohamed Morsi M. Ahmed, Hassan A. I. Ramadan, Kulvinder Singh Saini, and Elrashdy M. Redwan. Cell factories for insulin production. *Microbial Cell Factories*, 13:141, 2014. 11

[75] RONALD A. RADER and ERIC S. LANGER. Biopharmaceutical Manufacturing: Historical and Future Trends in Titers, Yields, and Efficiency in Commercial-Scale Bioprocessing. *BioProcessing*, 13(4):47–54, 2014. 11

[76] Keith Pardee, Alexander A. Green, Tom Ferrante, D. Ewen Cameron, Ajay DaleyKeyser, Peng Yin, and James J. Collins. Paper-based Synthetic Gene Networks. *Cell*, 159(4):940–954, November 2014. 12

[77] Sean X. Hu, Murray L. Aitken, Arnold M. Epstein, Mark R. Trusheim, and Ernst R. Berndt. Market watch: Defining and quantifying the use of personalized medicines. *Nature Reviews Drug Discovery*, 12(12):896–897, December 2013. 12

[78] Y.-H. Percival Zhang. Production of biocommodities and bioelectricity by cell-free synthetic enzymatic pathway biotransformations: Challenges and opportunities. *Biotechnology and Bioengineering*, 105(4):663–677, March 2010. 12

[79] Rita Sachse, Doreen Wüstenhagen, Mária Šamalíková, Michael Gerrits, Frank F. Bier, and Stefan Kubick. Synthesis of membrane proteins in eukaryotic cell-free systems. *Engineering in Life Sciences*, 13(1):39–48, January 2013. 12

[80] Matthias Bujara, Michael Schümperli, René Pellaux, Matthias Heinemann, and Sven Panke. Optimization of a blueprint for in vitro glycolysis by metabolic real-time analysis. *Nature Chemical Biology*, 7(5):271–277, May 2011. 12, 17

[81] Quentin M. Dudley, Ashty S. Karim, and Michael C. Jewett. Cell-free metabolic engineering: Biomanufacturing beyond the cell. *Biotechnology Journal*, 10(1):69–82, January 2015. 12

[82] Seok Hoon Hong, Yong-Chan Kwon, Rey W. Martin, Benjamin J. Des Soye, Alexandra M. de Paz, Kirsten N. Swonger, Ioanna Ntai, Neil L. Kelleher, and

Michael C. Jewett. Improving Cell-Free Protein Synthesis through Genome Engineering of Escherichia coli Lacking Release Factor 1. *ChemBioChem*, 16(5):844–853, March 2015. 12

[83] Michael C. Jewett and James R. Swartz. Mimicking the Escherichia coli cytoplasmic environment activates long-lived and efficient cell-free protein synthesis. *Biotechnology and Bioengineering*, 86(1):19–26, April 2004. 12

[84] Andrea C. Timm, Peter G. Shankles, Carmen M. Foster, Mitchel J. Doktycz, and Scott T. Retterer. Toward Microfluidic Reactors for Cell-Free Protein Synthesis at the Point-of-Care. *Small*, 12(6):810–817, February 2016. 13

[85] L. J. Millet, J. D. Lucheon, R. F. Standaert, S. T. Retterer, and M. J. Doktycz. Modular microfluidics for point-of-care protein purifications. *Lab on a Chip*, 15(8):1799–1811, March 2015. 13

[86] Vladimir Gubala, Leanne F. Harris, Antonio J. Ricco, Ming X. Tan, and David E. Williams. Point of Care Diagnostics: Status and Future. *Analytical Chemistry*, 84(2):487–515, January 2012. 13

[87] Federico Katzen, Geoffrey Chang, and Wieslaw Kudlicki. The past, present and future of cell-free protein synthesis. *Trends in Biotechnology*, 23(3):150–156, March 2005. 14

[88] Naoki Goshima, Yoshifumi Kawamura, Akiko Fukumoto, Aya Miura, Reiko Honma, Ryohei Satoh, Ai Wakamatsu, Jun-ichi Yamamoto, Kouichi Kimura, Tetsuo Nishikawa, Taichi Andoh, Yuki Iida, Kumiko Ishikawa, Emi Ito, Naoko Kagawa, Chie Kaminaga, Kei-ichi Kanehori, Bunsei Kawakami, Kiyokazu Kenmochi, Rie Kimura, Miki Kobayashi, Toshihiro Kuroita, Hisashi Kuwayama, Yukio Maruyama, Kiyoshi Matsuo, Kazuyoshi Minami, Mariko Mitsubori, Masatoshi Mori, Riyo Morishita, Atsushi Murase, Akira Nishikawa, Shigemichi Nishikawa, Toshihiko Okamoto, Noriko Sakagami, Yutaka Sakamoto, Yukari Sasaki, Tomoe Seki, Saki Sono, Akio Sugiyama, Tsuyoshi Sumiya, Tomoko Takayama, Yukiko Takayama, Hiroyuki Takeda, Takushi Togashi, Kazuhide Yahata, Hiroko Yamada, Yuka Yanagisawa, Yaeta Endo, Fumio

Imamoto, Yasutomo Kisu, Shigeo Tanaka, Takao Isogai, Jun-ichi Imai, Shinya Watanabe, and Nobuo Nomura. Human protein factory for converting the transcriptome into an in vitro-expressed proteome,. *Nature Methods*, 5(12):1011–1017, December 2008. 14

[89] S. Harrer, S. C. Kim, C. Schieber, S. Kannam, N. Gunn, S. Moore, D. Scott, R. Bathgate, S. Skafidas, and J. M. Wagner. Label-free screening of single biomolecules through resistive pulse sensing technology for precision medicine applications. *Nanotechnology*, 26(18):182502, 2015. 14

[90] Chariz Peñalber-Johnstone, Xudong Ge, Kevin Tran, Nicholas Selock, Neha Sardesai, Chandrasekhar Gurramkonda, Manohar Pilli, Michael Tolosa, Leah Tolosa, Yordan Kostov, Douglas D. Frey, and Govind Rao. Optimizing cell-free protein expression in CHO: Assessing small molecule mass transfer effects in various reactor configurations. *Biotechnology and Bioengineering*, 114(7):1478–1486, July 2017. 17

[91] Yoshihiro Toya and Hiroshi Shimizu. Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. *Biotechnology Advances*, 31(6):818–826, November 2013. 17

[92] Joseph A. Rollin, Tsz Kin Tam, and Y.-H. Percival Zhang. New biotechnology paradigm: cell-free biosystems for biomanufacturing. *Green Chemistry*, 15(7):1708–1719, 2013. 17

[93] Y.-H. Percival Zhang, Barbara R. Evans, Jonathan R. Mielenz, Robert C. Hopkins, and Michael W. W. Adams. High-Yield Hydrogen Production from Starch and Water by a Synthetic Enzymatic Pathway. *PLOS ONE*, 2(5):e456, May 2007. 17

[94] Fernando Villarreal and Cheemeng Tan. Cell-free systems in the new age of synthetic biology. *Frontiers of Chemical Science and Engineering*, pages 1–8, January 2017. 17

[95] Ashty S. Karim and Michael C. Jewett. A cell-free framework for rapid biosynthetic pathway prototyping and enzyme discovery. *Metabolic Engineering*, 36:116–126, July 2016. 18, 37

[96] Yong Y Wu, Stephanie Culler, Julia Khandurina, Stephen Van Dien, and Richard M Murray. Prototyping 1,4-butanediol (BDO) biosynthesis pathway in a cell-free transcription-translation (TX-TL) system. *bioRxiv*, 017814, January 2015. 18

[97] Tae-Wan Kim, Jung-Won Keum, In-Seok Oh, Cha-Yong Choi, Ho-Cheol Kim, and Dong-Myung Kim. An economical and highly productive cell-free protein synthesis system utilizing fructose-1,6-bisphosphate as an energy source. *Journal of Biotechnology*, 130(4):389–393, July 2007. 18

[98] Gregory B. Hurst, Keiji G. Asano, Charles J. Doktycz, Elliot J. Consoli, William L. Doktycz, Carmen M. Foster, Jennifer L. Morrell-Falvey, Robert F. Standaert, and Mitchel J. Doktycz. Proteomics-Based Tools for Evaluation of Cell-Free Protein Synthesis. *Analytical Chemistry*, 89:11443–11451, October 2017. 18, 21, 37, 46, 66, 68

[99] Michael P. Washburn, Dirk Wolters, and John R. Yates. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology*, 19(3):242, March 2001. 20, 22, 40

[100] Ze-Qiang Ma, Surendra Dasari, Matthew C. Chambers, Michael D. Litton, Scott M. Sobecki, Lisa J. Zimmerman, Patrick J. Halvey, Birgit Schilling, Penelope M. Drake, Bradford W. Gibson, and David L. Tabb. IDPicker 2.0: Improved Protein Assembly with High Discrimination Peptide Identification Filtering. *Journal of proteome research*, 8(8):3872–3881, August 2009. 21, 41, 69

[101] Minoru Kanehisa, Yoko Sato, and Kanae Morishima. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *Journal of Molecular Biology*, 428(4):726–731, February 2016. 21, 22, 41

[102] Aravind Subramanian, Pablo Tamayo, Vamsi K. Mootha, Sayan Mukherjee, Benjamin L. Ebert, Michael A. Gillette, Amanda Paulovich, Scott L. Pomeroy, Todd R. Golub, and Eric S. Lander. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, 102(43):15545–15550, 2005. 21, 26

[103] Xiaolin Zhang, Christopher J. Tervo, and Jennifer L. Reed. Metabolic assessment of E. coli as a Biofactory for commercial products. *Metabolic Engineering*, 35:64–74, May 2016. 22, 29, 37

[104] Zachary Z. Sun, Clarmyra A. Hayes, Jonghyeon Shin, Filippo Caschera, Richard M. Murray, and Vincent Noireaux. Protocols for Implementing an Escherichia coli Based TX-TL Cell-Free Expression System for Synthetic Biology. *Journal of Visualized Experiments : JoVE*, (79), September 2013. 22

[105] Jurek Failmezger, Michael Rauter, Robert Nitschel, Michael Kraml, and Martin Siemann-Herzberg. Cell-free protein synthesis from non-growing, stressed Escherichia coli. *Scientific Reports*, 7(1):16524, November 2017. 22, 32, 46

[106] Thomas Nyström. Stationary-Phase Physiology. *Annual Review of Microbiology*, 58(1):161–181, 2004. 22

[107] Alexander Schmidt, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer, Luciano Callipo, Kèvin Knoops, Manuel Bauer, Ruedi Aebersold, and Matthias Heinemann. The quantitative and condition-dependent *Escherichia coli* proteome. *Nature Biotechnology*, 34(1):nbt.3418, December 2015. 22

[108] Chaodong Wu, Salmaan A. Khan, Li-Jen Peng, and Alex J. Lange. Roles for fructose-2,6-bisphosphate in the control of fuel metabolism: Beyond its allosteric effects on glycolytic and gluconeogenic enzymes. *Advances in Enzyme Regulation*, 46(1):72–88, January 2006. 25

[109] Johann M. Rohwer, Pieter W. Postma, Boris N. Kholodenko, and Hans V. Westerhoff. Implications of macromolecular crowding for signal transduction and metabolite channeling. *Proceedings of the National Academy of Sciences*, 95(18):10547–10552, September 1998. 25

[110] Drew S. Cunningham, Zhu Liu, Nathan Domagalski, Richard R. Koepsel, Mohammad M. Ataai, and Michael M. Domach. Pyruvate kinase-deficient Escherichia

coli exhibits increased plasmid copy number and cyclic AMP levels. *Journal of Bacteriology*, 191(9):3041–3049, May 2009. 31

[111] Matthew D. Rolfe, Christopher J. Rice, Sacha Lucchini, Carmen Pin, Arthur Thompson, Andrew D. S. Cameron, Mark Alston, Michael F. Stringer, Roy P. Betts, József Baranyi, Michael W. Peck, and Jay C. D. Hinton. Lag Phase Is a Distinct Growth Phase That Prepares Bacteria for Exponential Growth and Involves Transient Metal Accumulation. *Journal of Bacteriology*, 194(3):686–701, February 2012. 32

[112] Elad Noor, Eran Eden, Ron Milo, and Uri Alon. Central Carbon Metabolism as a Minimal Biochemical Walk between Precursors for Biomass and Energy. *Molecular Cell*, 39(5):809–820, September 2010. 32

[113] Aaron R Goerke and James R Swartz. Development of cell-free protein synthesis platforms for disulfide bonded proteins. *Biotechnology and bioengineering*, 99(2):351–67, February 2008. 32

[114] Y. Li, J. Chen, and S.-Y. Lun. Biotechnological production of pyruvic acid. *Applied Microbiology and Biotechnology*, 57(4):451–459, November 2001. 32

[115] Neda Maleki and Mark Eiteman. Recent Progress in the Microbial Production of Pyruvic Acid. *Fermentation*, 3(1):8, February 2017. 32

[116] C. Eric Hodgman and Michael C. Jewett. Cell-Free Synthetic Biology: Thinking Outside the Cell. *Metabolic Engineering*, 14(3):261–269, May 2012. 36

[117] Quentin M. Dudley, Kim C. Anderson, and Michael C. Jewett. Cell-Free Mixing of Escherichia coli Crude Extracts to Prototype and Rationally Engineer High-Titer Mevalonate Synthesis. *ACS Synthetic Biology*, 5(12):1578–1588, December 2016. 36, 84

[118] Dong-Myung Kim, Takanori Kigawa, Cha-Yong Choi, and Shigeyuki Yokoyama. A highly efficient cell-free protein synthesis system from Escherichia coli. *European Journal of Biochemistry*, 239(3):881–886, 1996. 37

[119] Filippo Caschera and Vincent Noireaux. A cost-effective polyphosphate-based metabolism fuels an all E. coli cell-free expression system. *Metabolic Engineering*, 27:29–37, January 2015. 37

[120] Daniel Foshag, Erik Henrich, Ekkehard Hiller, Miriam Schäfer, Christian Kerger, Anke Burger-Kentischer, Irene Diaz-Moreno, Sofía M. García-Mauriño, Volker Dötsch, Steffen Rupp, and Frank Bernhard. The E. coli S30 lysate proteome: A prototype for cell-free protein production. *New Biotechnology*, 40:245–260, January 2018. 37, 78

[121] David C. Garcia, Benjamin P. Mohr, Jakob T. Dovgan, Gregory B. Hurst, Robert F. Standaert, and Mitchel J. Doktycz. Elucidating the potential of crude cell extracts for producing pyruvate from glucose. *Synthetic Biology*, 3(1), January 2018. 37, 40, 43, 50

[122] Tom Taverner, Yuliya V. Karpievitch, Ashoka D. Polpitiya, Joseph N. Brown, Alan R. Dabney, Gordon A. Anderson, and Richard D. Smith. DanteR: an extensible R-based tool for quantitative analysis of -omics data. *Bioinformatics*, 28(18):2404–2406, September 2012. 41

[123] Stefka Tyanova, Tikira Temu, Pavel Sinitcyn, Arthur Carlson, Marco Y Hein, Tamar Geiger, Matthias Mann, and Jürgen Cox. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nature Methods*, 13(9):731–740, September 2016. 41

[124] Wickham Hadley. *Ggplot2*. Springer Science+Business Media, LLC, New York, NY, 2016. 41

[125] Patrick T. O'Kane, Quentin M. Dudley, Aislinn K. McMillan, Michael C. Jewett, and Milan Mrksich. High-throughput mapping of CoA metabolites by SAMDI-MS to optimize the cell-free biosynthesis of HMG-CoA. *Science Advances*, 5(6):eaaw9180, June 2019. 42

[126] Byoungjin Kim, Hyegwon Park, Dokyun Na, and Sang Yup Lee. Metabolic engineering of Escherichia coli for the production of phenol from glucose. *Biotechnology Journal*, 9(5):621–629, May 2014. 42

[127] Peng Zhang, Junqian Wang, Xuanwei Ding, Jun Lin, Hao Jiang, Hongjun Zhou, and Yuan Lu. Exploration of the Tolerance Ability of a Cell-Free Biosynthesis System to Toxic Substances. *Applied Biochemistry and Biotechnology*, June 2019. 42, 78

[128] Michael C. Jewett, Kara A. Calhoun, Alexei Voloshin, Jessica J. Wuu, and James R. Swartz. An integrated cell-free metabolic platform for protein production and synthetic biology. *Molecular Systems Biology*, 4(1):220, January 2008. 43

[129] F C Neidhardt, P L Bloch, and D F Smith. Culture medium for enterobacteria. *Journal of bacteriology*, 119(3):736–747, September 1974. 45

[130] Jorge Alonso-Gutierrez, Eun-Mi Kim, Tanveer S. Batth, Nathan Cho, Qijun Hu, Leanne Jade G. Chan, Christopher J. Petzold, Nathan J. Hillson, Paul D. Adams, Jay D. Keasling, Hector Garcia Martin, and Taek Soon Lee. Principal component analysis of proteomics (PCAP) as a tool to direct metabolic engineering. *Metabolic Engineering*, 28:123–133, March 2015. 45, 92

[131] James Pittard, Helen Camakaris, and Ji Yang. The TyrR regulon. *Molecular Microbiology*, 55(1):16–26, 2005. 46

[132] M. I. Chavez-Bejar, A. R. Lara, H. Lopez, G. Hernandez-Chavez, A. Martinez, O. T. Ramirez, F. Bolivar, and G. Gosset. Metabolic Engineering of Escherichia coli for L-Tyrosine Production by Expression of Genes Coding for the Chorismate Mutase Domain of the Native Chorismate Mutase-Prephenate Dehydratase and a Cyclohexadienyl Dehydrogenase from Zymomonas mobilis. *Applied and Environmental Microbiology*, 74(10):3284–3290, May 2008. 48

[133] Sunil S. Chandran, Jian Yi, K. M. Draths, Ralph von Daeniken, Wolfgang Weber, and J. W. Frost. Phosphoenolpyruvate Availability and the Biosynthesis of Shikimic Acid. *Biotechnology Progress*, 19(3):808–814, 2003. 52

[134] Nick Wierckx, Harald J. Ruijssenaars, Johannes H. de Winde, Andreas Schmid, and Lars M. Blank. Metabolic flux analysis of a phenol producing mutant of Pseudomonas putida S12: Verification and complementation of hypotheses derived from transcriptomics. *Journal of Biotechnology*, 143(2):124–129, August 2009. 53

[135] Rui Ding, Lifei Liu, Xuhui Chen, Zhenhai Cui, Ao Zhang, Daming Ren, and Lijun Zhang. Introduction of two mutations into AroG increases phenylalanine production in Escherichia coli. *Biotechnology Letters*, 36(10):2103–2108, October 2014. 53

[136] E D Carlson, R Gan, C E Hodgman, and M C Jewett. Cell-free protein synthesis: applications come of age. *Biotechnol Adv*, 30:1185–1194, 2012. 53

[137] Benjamin P. Mohr, Scott T. Retterer, and Mitchel J. Doktycz. While-you-wait proteins? Producing biomolecules at the point of need. *Expert Review of Proteomics*, 13(8):707–709, August 2016. 55

[138] Stephanie D. Cole, Kathryn Beabout, Kendrick B. Turner, Zachary K. Smith, Vanessa L. Funk, Svetlana V Harbaugh, Alvin T. Liem, Pierce A. Roth, Brian A. Geier, Peter A. Emanuel, Scott A. Walper, Jorge Luis Chávez, and Matthew W. Lux. Quantification of Interlaboratory Cell-free Protein Synthesis Variability. *ACS Synthetic Biology*, August 2019. 55

[139] Zak Costello and Hector Garcia Martin. A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *npj Systems Biology and Applications*, 4(1):1–14, May 2018. 55, 92

[140] Tobias Fuhrer, Dominik Heer, Boris Begemann, and Nicola Zamboni. High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection–Time-of-Flight Mass Spectrometry. *Analytical Chemistry*, 83(18):7074–7080, September 2011. 55, 92

[141] Christina A. Müller, Melanie M. Obermeier, and Gabriele Berg. Bioprospecting plant-associated microbiomes. *Journal of Biotechnology*, 235:171–180, October 2016. 64

[142] Victor M. Markowitz, Konstantinos Mavromatis, Natalia N. Ivanova, I.-Min A. Chen, Ken Chu, and Nikos C. Kyrpides. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics*, 25(17):2271–2278, September 2009. 64, 66

[143] Steven L. Salzberg. Next-generation genome annotation: we still struggle to get it right. *Genome Biology*, 20(1):92, May 2019. 64

[144] Julian D. Hegemann, Marcel Zimmermann, Xiulan Xie, and Mohamed A. Marahiel. Lasso Peptides: An Intriguing Class of Bacterial Natural Products. *Accounts of Chemical Research*, 48(7):1909–1919, July 2015. 64

[145] Tatyana Zyubko, Marina Serebryakova, Julia Andreeva, Mikhail Metelev, Guy Lippens, Svetlana Dubiley, and Konstantin Severinov. Efficient in vivo synthesis of lasso peptide pseudomycoidin proceeds in the absence of both the leader and the leader peptidase. *Chemical Science*, August 2019. 64, 65

[146] Anne de Jong, Sacha A. F. T. van Hijum, Jetta J. E. Bijlsma, Jan Kok, and Oscar P. Kuipers. BAGEL: a web-based bacteriocin genome mining tool. *Nucleic Acids Research*, 34(Web Server issue):W273–W279, July 2006. 64

[147] Jonathan I. Tietz, Christopher J. Schwalen, Parth S. Patel, Tucker Maxson, Patricia M. Blair, Hua-Chia Tai, Uzma I. Zakai, and Douglas A. Mitchell. A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nature chemical biology*, 13(5):470–478, May 2017. 64, 65, 70, 72, 73, 83

[148] Sophie Duquesne, Delphine Destoumieux-Garzón, Séverine Zirah, Christophe Goulard, Jean Peduzzi, and Sylvie Rebuffat. Two Enzymes Catalyze the Maturation of a Lasso Peptide in Escherichia coli. *Chemistry & Biology*, 14(7):793–803, July 2007. 65

[149] Brandon J. Burkhart, Graham A. Hudson, Kyle L. Dunbar, and Douglas A. Mitchell. A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nature Chemical Biology*, 11(8):564–570, August 2015. 65, 75, 77

[150] Issara Kaweewan, Hikaru Hemmi, Hisayuki Komaki, Shigeyoshi Harada, and Shinya Kodani. Isolation and structure determination of a new lasso peptide specialicin based on genome mining. *Bioorganic & Medicinal Chemistry*, 26(23):6050–6055, December 2018. 65

[151] Thomas A. Knappe, Uwe Linne, Séverine Zirah, Sylvie Rebuffat, Xiulan Xie, and Mohamed A. Marahiel. Isolation and Structural Characterization of Capistruin, a Lasso Peptide Predicted from the Genome Sequence of Burkholderia thailandensis E264. *Journal of the American Chemical Society*, 130(34):11446–11454, August 2008. 65, 70

[152] Marcel Zimmermann, Julian D. Hegemann, Xiulan Xie, and Mohamed A. Marahiel. The Astexin-1 Lasso Peptides: Biosynthesis, Stability, and Structural Studies. *Chemistry & Biology*, 20(4):558–569, April 2013. 65

[153] Pikyee Ma, Kenzo Nishiguchi, Hayley M. Yuille, Lianne M. Davis, Jiro Nakayama, and Mary K. Phillips-Jones. Anti-HIV siamycin I directly inhibits autophosphorylation activity of the bacterial FsrC quorum sensor and other ATP-dependent enzyme activities. *FEBS Letters*, 585(17):2660–2664, September 2011. 65

[154] Mikhail O. Maksimov, Si Jia Pan, and A. James Link. Lasso peptides: structure, function, biosynthesis, and engineering. *Natural Product Reports*, 29(9):996, 2012. 65

[155] Kelly-Anne Wilson, Markus Kalkum, Jennifer Ottesen, Julia Yuzenkova, Brian T. Chait, Robert Landick, Tom Muir, Konstantin Severinov, and Seth A. Darst. Structure of microcin J25, a peptide inhibitor of bacterial RNA polymerase, is a lassoed tail. *Journal of the American Chemical Society*, 125(41):12475–12483, October 2003. 65, 78

[156] Yu Su, Meng Han, Xianbin Meng, Yue Feng, Shizhong Luo, Changyuan Yu, Guojun Zheng, and Shaozhou Zhu. Discovery and characterization of a novel C-terminal peptide carboxyl methyltransferase in a lassomycin-like lasso peptide biosynthetic pathway. *Applied Microbiology and Biotechnology*, 103(6):2649–2664, March 2019. 65

[157] Shaozhou Zhu, Julian D. Hegemann, Christopher D. Fage, Marcel Zimmermann, Xiulan Xie, Uwe Linne, and Mohamed A. Marahiel. Insights into the Unique Phosphorylation of the Lasso Peptide Paeninodin. *Journal of Biological Chemistry*, 291(26):13662–13678, June 2016. 65, 72, 73, 77

[158] Julian D. Hegemann, Marcel Zimmermann, Shaozhou Zhu, Dennis Klug, and Mohamed A. Marahiel. Lasso peptides from proteobacteria: Genome mining employing heterologous expression and mass spectrometry. *Peptide Science*, 100(5):527–542, 2013. 66, 72

[159] Delphine Charif and Jean R. Lobry. SeqinR 1.0-2: A Contributed Package to the R Project for Statistical Computing Devoted to Biological Sequences Retrieval and Analysis. In Ugo Bastolla, Markus Porto, H. Eduardo Roman, and Michele Vendruscolo, editors, *Structural Approaches to Sequence Evolution: Molecules, Networks, Populations*, Biological and Medical Physics, Biomedical Engineering, pages 207–232. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007. 67

[160] Ivica Letunic and Peer Bork. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Research*, 44(W1):W242–W245, July 2016. 67

[161] Emmanuel Paradis and Klaus Schliep. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics (Oxford, England)*, 35(3):526–528, February 2019. 67

[162] Klaus Peter Schliep. phangorn: phylogenetic analysis in R. *Bioinformatics (Oxford, England)*, 27(4):592–593, February 2011. 67

[163] D J Reasoner and E E Geldreich. A new medium for the enumeration and subculture of bacteria from potable water. *Applied and Environmental Microbiology*, 49(1):1–7, January 1985. 67

[164] John A. Gerlt, Jason T. Bouvier, Daniel B. Davidson, Heidi J. Imker, Boris Sadkhin, David R. Slater, and Katie L. Whalen. Enzyme Function Initiative-Enzyme Similarity

Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica Et Biophysica Acta*, 1854(8):1019–1037, August 2015. 67, 75

[165] Paul Shannon, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, 13(11):2498–2504, November 2003. 67

[166] Timothy L. Bailey, Mikael Boden, Fabian A. Buske, Martin Frith, Charles E. Grant, Luca Clementi, Jingyuan Ren, Wilfred W. Li, and William S. Noble. MEME Suite: tools for motif discovery and searching. *Nucleic Acids Research*, 37(suppl_2):W202–W208, July 2009. 68, 75

[167] Sonya M. Clarkson, Richard J. Giannone, Donna M. Kridelbaugh, James G. Elkins, Adam M. Guss, and Joshua K. Michener. Construction and Optimization of a Heterologous Pathway for Protocatechuate Catabolism in Escherichia coli Enables Bioconversion of Model Aromatic Compounds. *Applied and Environmental Microbiology*, 83(18), August 2017. 69

[168] Patricia M. Blair, Miriam L. Land, Marek J. Piatek, Daniel A. Jacobson, Tse-Yuan S. Lu, Mitchel J. Doktycz, and Dale A. Pelletier. Exploration of the Biosynthetic Potential of the Populus Microbiome. *mSystems*, 3(5):e00045–18, October 2018. 70, 72

[169] Tilmann Weber, Kai Blin, Srikanth Duddela, Daniel Krug, Hyun Uk Kim, Robert Bruccoleri, Sang Yup Lee, Michael A. Fischbach, Rolf Müller, Wolfgang Wohlleben, Rainer Breitling, Eriko Takano, and Marnix H. Medema. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, 43(W1):W237–W243, July 2015. 70

[170] Aron Marchler-Bauer, Yu Bo, Lianyi Han, Jane He, Christopher J. Lanczycki, Shennan Lu, Farideh Chitsaz, Myra K. Derbyshire, Renata C. Geer, Noreen R. Gonzales, Marc Gwadz, David I. Hurwitz, Fu Lu, Gabriele H. Marchler, James S. Song, Narmada Thanki, Zhouxi Wang, Roxanne A. Yamashita, Dachuan Zhang, Chanjuan Zheng,

Lewis Y. Geer, and Stephen H. Bryant. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research*, 45(D1):D200–D203, 2017. 72

[171] Mikhail O. Maksimov and A. James Link. Discovery and Characterization of an Isopeptidase That Linearizes Lasso Peptides. *Journal of the American Chemical Society*, 135(32):12038–12047, August 2013. 72

[172] Jonathan R. Chekan, Joseph D. Koos, Chuhan Zong, Mikhail O. Maksimov, A. James Link, and Satish K. Nair. Structure of the Lasso Peptide Isopeptidase Identifies a Topology for Processing Threaded Substrates. *Journal of the American Chemical Society*, 138(50):16452–16458, December 2016. 73

[173] Christopher D. Fage, Julian D. Hegemann, Annika J. Nebel, Roman M. Steinbach, Shaozhou Zhu, Uwe Linne, Klaus Harms, Gert Bange, and Mohamed A. Marahiel. Structure and Mechanism of the Sphingopyxin I Lasso Peptide Isopeptidase. *Angewandte Chemie International Edition*, 55(41):12717–12721, 2016. 73

[174] J. Gough, K. Karplus, R. Hughey, and C. Chothia. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *Journal of Molecular Biology*, 313(4):903–919, November 2001. 73

[175] Sara El-Gebali, Jaina Mistry, Alex Bateman, Sean R. Eddy, Aurélien Luciani, Simon C. Potter, Matloob Qureshi, Lorna J. Richardson, Gustavo A. Salazar, Alfredo Smart, Erik L. L. Sonnhammer, Layla Hirsh, Lisanna Paladin, Damiano Piovesan, Silvio C. E. Tosatto, and Robert D. Finn. The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1):D427–D432, January 2019. 73

[176] Julian D. Hegemann, Christopher J. Schwalen, Douglas A. Mitchell, and Wilfred A. van der Donk. Elucidation of the roles of conserved residues in the biosynthesis of the lasso peptide paeninodin. *Chemical Communications*, 54(65):9007–9010, 2018. 73

[177] Mikhail Metelev, Anatolii Arseniev, Leah B. Bushin, Konstantin Kuznedelov, Tatiana O. Artamonova, Ruslan Kondratenko, Mikhail Khodorkovskii, Mohammad R.

Seyedsayamdost, and Konstantin Severinov. Acinetodin and Klebsidin, RNA Polymerase Targeting Lasso Peptides Produced by Human Isolates of Acinetobacter gyllenbergii and Klebsiella pneumoniae. *ACS chemical biology*, 12(3):814–824, 2017. 73

[178] Marcel Zimmermann, Julian D. Hegemann, Xiulan Xie, and Mohamed A. Marahiel. Characterization of caulonodin lasso peptides revealed unprecedented N-terminal residues and a precursor motif essential for peptide maturation. *Chemical Science*, 5(10):4032–4043, 2014. 77

[179] Yonghao Yu, Adam J Hoffhines, Kevin L Moore, and Julie A Leary. Determination of the sites of tyrosine O-sulfation in peptides and proteins. *Nature Methods*, 4(7):583–588, July 2007. 77

[180] Yoshihiro Shimizu, Takashi Kanamori, and Takuya Ueda. Protein synthesis by pure translation systems. *Methods (San Diego, Calif.)*, 36(3):299–304, July 2005. 78

[181] Thomas M. Annesley. Ion Suppression in Mass Spectrometry. *Clinical Chemistry*, 49(7):1041–1044, July 2003. 80

[182] Claire I. Butré, Sofie Buhler, Stefano Sforza, Harry Gruppen, and Peter A. Wierenga. Spontaneous, non-enzymatic breakdown of peptides during enzymatic protein hydrolysis. *Biochimica Et Biophysica Acta*, 1854(8):987–994, August 2015. 80

[183] Hosein Mohimani, Alexey Gurevich, Alexander Shlemov, Alla Mikheenko, Anton Korobeynikov, Liu Cao, Egor Shcherbin, Louis-Felix Nothias, Pieter C. Dorrestein, and Pavel A. Pevzner. Dereplication of microbial metabolites through database search of mass spectra. *Nature Communications*, 9(1):1–12, October 2018. 80

[184] Timo H. J. Niedermeyer and Martin Strohalm. mMass as a Software Tool for the Annotation of Cyclic Peptide Tandem Mass Spectra. *PLOS ONE*, 7(9):e44913, September 2012. 80, 84

[185] Ngoc Hieu Tran, Rui Qiao, Lei Xin, Xin Chen, Chuyi Liu, Xianglilan Zhang, Baozhen Shan, Ali Ghodsi, and Ming Li. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nature Methods*, 16(1):63–66, 2019. 80

[186] Benjamin J. Des Soye, Samuel R. Davidson, Matthew T. Weinstock, Daniel G. Gibson, and Michael C. Jewett. Establishing a High-Yielding Cell-Free Protein Synthesis Platform Derived from Vibrio natriegens. *ACS Synthetic Biology*, 7(9):2245–2255, September 2018. 84, 91

[187] Mingxun Wang, Jeremy J Carver, Vanessa V Phelan, Laura M Sanchez, Neha Garg, Yao Peng, Don Duy Nguyen, Jeramie Watrous, Clifford A Kapono, Tal Luzzatto-Knaan, Carla Porto, Amina Bouslimani, Alexey V Melnik, Michael J Meehan, Wei-Ting Liu, Max Crüsemann, Paul D Boudreau, Eduardo Esquenazi, Mario Sandoval-Calderón, Roland D Kersten, Laura A Pace, Robert A Quinn, Katherine R Duncan, Cheng-Chih Hsu, Dimitrios J Floros, Ronnie G Gavilan, Karin Kleigrewe, Trent Northen, Rachel J Dutton, Delphine Parrot, Erin E Carlson, Bertrand Aigle, Charlotte F Michelsen, Lars Jelsbak, Christian Sohlenkamp, Pavel Pevzner, Anna Edlund, Jeffrey McLean, Jörn Piel, Brian T Murphy, Lena Gerwick, Chih-Chuang Liaw, Yu-Liang Yang, Hans-Ulrich Humpf, Maria Maansson, Robert A Keyzers, Amy C Sims, Andrew R. Johnson, Ashley M Sidebottom, Brian E Sedio, Andreas Klitgaard, Charles B Larson, Cristopher A Boya P., Daniel Torres-Mendoza, David J Gonzalez, Denise B Silva, Lucas M Marques, Daniel P Demarque, Egle Pociute, Ellis C O'Neill, Enora Briand, Eric J. N. Helfrich, Eve A Granatosky, Evgenia Glukhov, Florian Ryffel, Hailey Houson, Hosein Mohimani, Jenan J Kharbush, Yi Zeng, Julia A Vorholt, Kenji L Kurita, Pep Charusanti, Kerry L McPhail, Kristian Fog Nielsen, Lisa Vuong, Maryam Elfeki, Matthew F Traxler, Niclas Engene, Nobuhiro Koyama, Oliver B Vining, Ralph Baric, Ricardo R Silva, Samantha J Mascuch, Sophie Tomasi, Stefan Jenkins, Venkat Macherla, Thomas Hoffman, Vinayak Agarwal, Philip G Williams, Jingqui Dai, Ram Neupane, Joshua Gurr, Andrés M. C. Rodríguez, Anne Lamsa, Chen Zhang, Kathleen Dorrestein, Brendan M Duggan, Jehad Almaliti, Pierre-Marie Allard,

Prasad Phapale, Louis-Felix Nothias, Theodore Alexandrov, Marc Litaudon, Jean-Luc Wolfender, Jennifer E Kyle, Thomas O Metz, Tyler Peryea, Dac-Trung Nguyen, Danielle VanLeer, Paul Shinn, Ajit Jadhav, Rolf Müller, Katrina M Waters, Wenyuan Shi, Xueting Liu, Lixin Zhang, Rob Knight, Paul R Jensen, Bernhard O Palsson, Kit Pogliano, Roger G Linington, Marcelino Gutiérrez, Norberto P Lopes, William H Gerwick, Bradley S Moore, Pieter C Dorrestein, and Nuno Bandeira. Sharing and community curation of mass spectrometry data with GNPS. *Nature biotechnology*, 34(8):828–837, August 2016. 84

[188] Marnix H. Medema, Yared Paalvast, Don D. Nguyen, Alexey Melnik, Pieter C. Dorrestein, Eriko Takano, and Rainer Breitling. Pep2path: Automated Mass Spectrometry-Guided Genome Mining of Peptidic Natural Products. *PLOS Computational Biology*, 10(9):e1003822, September 2014. 84

[189] Jian Li, He Wang, Yong-Chan Kwon, and Michael C. Jewett. Establishing a high yielding streptomyces-based cell-free protein synthesis system. *Biotechnology and Bioengineering*, 114(6):1343–1353, 2017. 91

[190] Simon J Moore, Hung-En Lai, Hannah Needham, Karen M Polizzi, and Paul S Freemont. *Streptomyces venezuelae* TX-TL - a next generation cell-free synthetic biology tool. *Biotechnology Journal*, page 1600678, January 2017. 91

[191] Colin W. Brown, Viswanadham Sridhara, Daniel R. Boutz, Maria D. Person, Edward M. Marcotte, Jeffrey E. Barrick, and Claus O. Wilke. Large-scale analysis of post-translational modifications in E. coli under glucose-limiting conditions. *BMC Genomics*, 18:301, April 2017. 92

[192] Deniz Simsek and Maria Barna. An emerging role for the ribosome as a nexus for post-translational modifications. *Current opinion in cell biology*, 45:92–101, April 2017. 92

[193] Kerstin Petroll, Dominik Kopp, Andrew Care, Peter L. Bergquist, and Anwar Sunna. Tools and strategies for constructing cell-free enzyme pathways. *Biotechnology Advances*, December 2018. 92

[194] Konrad B Otte and Bernhard Hauer. Enzyme engineering in the context of novel pathways and products. *Current Opinion in Biotechnology*, 35:16–22, December 2015. 92

[195] Fernando Lopez-Gallego and Claudia Schmidt-Dannert. Multi-enzymatic synthesis. *Current opinion in chemical biology*, 14(2):174–83, April 2010. 92

[196] Saurabh Gupta, Sounik Sarkar, Alexandros Katranidis, and Jaydeep Bhattacharya. Development of a Cell-Free Optical Biosensor for Detection of a Broad Range of Mercury Contaminants in Water: A Plasmid DNA-Based Approach. *ACS Omega*, 4(5):9480–9487, May 2019. 93

[197] David K Karig. Cell-free synthetic biology for environmental sensing and remediation. *Current Opinion in Biotechnology*, 45:69–75, June 2017. 93

[198] Mehran Soltani, Brady R. Davis, Hayley Ford, J. Andrew D. Nelson, and Bradley C. Bundy. Reengineering cell-free protein synthesis as a biosensor: Biosensing with transcription, translation, and protein-folding. *Biochemical Engineering Journal*, 138:165–171, October 2018. 93

Vita

Benjamin Pintz Mohr was born in Powell, Ohio to Dr. Evelyn Pintz and Dr. James Mohr. He attended The Ohio State University for his undergraduate studies. He received his Bachelors of Science with a double major in Biochemistry and Evolution and Ecology. At Ohio State he developed teaching labs in Organic Chemistry with Dr. Noel M. Paul and completed an undergraduate honors dissertation titled "Towards colorimetric anion-selective indicators: Solid-phase methods for azo dye synthesis and screening". He pursued his PhD in Energy Science and Engineering in the Bredesen Center for Interdisciplinary Research. His graduate research in developing cell-free systems for natural product synthesis was overseen by Dr. Mitchel J. Doktycz.